Causal-IDS: Detecting Network Intrusions as Causal Mechanism Violations

1st Phuc Hao Do
Department of Software Engineering
Danang Architecture University
Da Nang, Viet Nam
haodp@dau.edu.vn

2nd Tran Duc Le
Department of Mathematics, Statistics & Computer Science
University of Wisconsin-Stout
Menomonie, WI, USA
let@uwstout.edu

3rd Truong Duy Dinh
Faculty of Information Security
Posts and Telecommunications Institute of Technology
Ha Noi, Viet Nam
duydt@ptit.edu.vn

4th Van Dai Pham
Metropolia Vietnam
FPT University
Ha Noi, Viet Nam
daipv11@fe.edu.vn

***Abstract*—Machine learning-based Intrusion Detection Systems (IDS) often suffer from high false positive rates, as they learn statistical correlations rather than the underlying cause-and-effect dynamics of network traffic. This paper introduces Causal-IDS, a novel framework that addresses this gap by leveraging causal inference. Causal-IDS first learns a Structural Causal Model representing the normal operational mechanisms of a network from benign data. Intrusions are then identified not as mere statistical outliers, but as significant violations of these learned causal laws, quantified by a Causal Anomaly Score. We demonstrate that accurately modeling complex, non-linear relationships is critical, as a naive causal model with linear assumptions fails. Our enhanced Causal-IDS, using Gradient Boosting, is evaluated on the CIC-IDS2017 dataset, where it significantly outperforms traditional methods like Isolation Forest and Autoencoders. Notably, it achieves a superior Area Under the ROC Curve (AUC) of 0.84, showcasing its ability to reliably distinguish attacks. By focusing on the disruption of causal mechanisms, our work paves the way for a new class of robust, interpretable, and more trustworthy intrusion detection systems.**

***Index Terms*—Network Intrusion Detection, Causal Inference, Anomaly Detection, Machine Learning, Network Security.**

I. Introduction

The proliferation of sophisticated cyber threats necessitates the development of advanced Intrusion Detection Systems (IDS) [1]. While machine learning (ML) has become a cornerstone for automating threat detection, many existing models are fundamentally limited by their reliance on statistical correlations. This approach renders them vulnerable to high false positive (FP) rates, which leads to alert fatigue and diminishes operator trust [2]. For instance, a standard ML model might learn a spurious correlation between high traffic volume and CPU load, consequently misclassifying a benign flash crowd event as a malicious Distributed Denial-of-Service (DDoS) attack because it fails to discern the underlying cause-and-effect relationship [3].

This paper argues that the key to building more robust and reliable IDSs lies in a paradigm shift from correlational to causal reasoning. We posit that an attack is not merely a statistical anomaly but an *external intervention* that disrupts the natural causal mechanisms of a network. A DDoS attack, for example, does not arise from the network's internal dynamics (e.g., a sudden increase in legitimate users); it is exogenously imposed by a botnet. This intervention breaks the established causal chain between legitimate user activity and network traffic, a fundamental distinction that correlation-based models cannot inherently make.

To address this gap, we introduce Causal-IDS, a framework grounded in structural causal inference [4]. The core principle is to first learn a model of a network's normal cause-and-effect relationships from benign traffic. Intrusions are then identified as significant violations of these learned causal laws. Our main contributions are threefold:

- We design a novel, two-phase IDS framework that separates causal graph discovery from the learning of its functional mechanisms, enhancing both model robustness and interpretability.
- We demonstrate the critical importance of accurate mechanism modeling by showing that a naive Causal-IDS with linear assumptions fails, whereas our enhanced version using non-linear models and data standardization achieves superior performance.
- We conduct a rigorous comparative evaluation on the CIC-IDS2017 dataset [5], showing that our en-

Corresponding author: duydt@ptit.edu.vn

hanced Causal-IDS significantly outperforms standard anomaly detection baselines with a superior Area Under the ROC Curve (AUC) of 0.84.

The remainder of this paper is organized as follows. Section II reviews related work. Section III details the Causal-IDS framework. Section IV presents the experimental evaluation. Section V discusses the implications and limitations of our findings. Finally, Section VI concludes the paper.

II. Related Work

The application of machine learning to IDS has been extensively studied, primarily following two paradigms: supervised classification and unsupervised anomaly detection. Supervised methods like Support Vector Machines (SVM) and Random Forests excel at detecting known threats but fail to identify novel, zero-day attacks [6]. To address this, unsupervised methods such as Isolation Forest and Autoencoders have gained traction by learning a model of normal behavior and flagging significant deviations as potential intrusions [7]–[9].

Despite their successes, a fundamental challenge persists across both paradigms: they are inherently correlational. These models learn statistical regularities but lack a deeper understanding of the system's underlying data-generating process. This "correlation vs. causation" gap is a primary contributor to high false positive rates and limited model robustness, motivating a shift towards a more principled approach [10].

Causal inference offers a direct theoretical solution to this problem by focusing on discovering and modeling cause-and-effect relationships from data. In network analysis, its application has centered on management and troubleshooting, where causal models have been used for root cause analysis (RCA) of performance degradations and failures [11]. These studies typically employ causal discovery algorithms like the Greedy Equivalence Search (GES) to construct a causal graph from observational data [12].

However, applying these techniques to *intrusion detection* is a largely unexplored frontier, and it is crucial to distinguish our approach from prior causality-informed methods. Some work, for instance, has used Granger causality for time-series anomaly detection [13]. Such methods primarily assess predictive causality between pairs of time-series variables. In contrast, our framework learns a comprehensive Structural Causal Model (SCM) of the entire system [14]. This provides a more holistic view, allowing us to detect an intrusion not just as a disruption between two variables, but as a violation of the underlying functional mechanisms governing any feature given its direct causes.

While previous works focus on explaining system behavior, our work pioneers the use of these SCM violations as the primary signal for detecting external, malicious interventions. This paper aims to fill this critical gap by proposing a complete framework for a causality-aware IDS that moves beyond pairwise statistical prediction to a mechanistic understanding of network operations.

III. The Proposed Causal-IDS Framework

Our proposed Causal-IDS framework is designed to detect intrusions by identifying violations of a learned structural causal model. The framework operates in two distinct phases: an offline training phase for model cc
sc

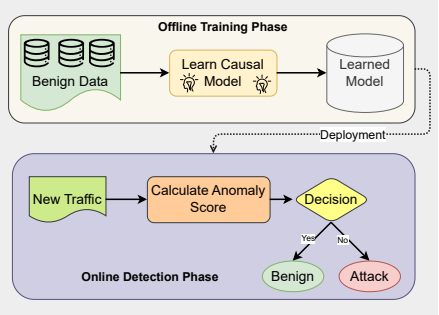

Fig. 1. The two-phase architecture of the Causal-IDS framework. The offline phase learns a causal model from benign data, which is then used by the online phase to score new traffic.

Let $X = \{X_1, X_2, \ldots, X_n\}$ be the set of n random variables representing the network features. We assume the data is generated from a Structural Causal Model [14], which consists of a directed acyclic graph (DAG) $\mathcal{G}$ and a set of functions $\{f_i\}$ describing the causal mechanisms. The entire process is formally summarized in Algorithm 1.

A. Phase 1: Causal Mechanism Training (Offline)

This phase leverages a dataset of purely benign traffic, $D_{benign} \in \mathbb{R}^{m \times n}$, where m is the number of samples, to learn the SCM of a normally operating network.

1) Causal Structure Discovery: The first step is to discover the underlying causal graph $\mathcal{G} = (V, E)$. An edge $X_i \to X_j$ in $\mathcal{G}$ signifies that X_i is a direct cause of X_j. We employ the GES algorithm [15], a prominent score-based method that heuristically searches the space of essential graphs to find the one that maximizes a scoring function. The objective is to find the graph $\mathcal{G}^*$ that best fits the data:

$$\mathcal{G}^* = \arg\max_{\mathcal{G}} \text{Score}(D_{benign}, \mathcal{G}) \quad (1)$$

In this work, we utilize the Bayesian Information Criterion (BIC) as our scoring function, which balances model fit with model complexity to prevent overfitting.

It is important to acknowledge that the validity of the discovered graph $\mathcal{G}^*$ hinges on several key

Algorithm 1 The Causal-IDS Framework Algorithm

Input: D_{benign} (Benign training data), D_{test} (Test data)
Output: Anomaly scores for D_{test}

Phase 1: Training Phase
$\mathcal{G}^* \leftarrow$ GES(D_{benign}) {Discover causal graph}
Initialize an empty model set $\mathcal{F} = \{\}$
for each feature X_i in D_{benign} **do**
  $Pa(X_i) \leftarrow$ Get parents of X_i from $\mathcal{G}^*$
  if $Pa(X_i)$ is empty **then**
    $\hat{f}_i \leftarrow$ mean(X_i) {Model is the mean}
  else
    $X_{parents} \leftarrow$ Select columns for $Pa(X_i)$
    $\hat{f}_i \leftarrow$ GradientBoostingRegressor()
    $\hat{f}_i$.fit($X_{parents}, X_i$) {Train mechanism model}
  end if
  Add $\hat{f}_i$ to $\mathcal{F}$
end for
Phase 2: Detection Phase
Initialize empty list $Scores$
for each instance x in D_{test} **do**
  $total\_error \leftarrow 0$
  for each feature x_i in x **do**
    Get model $\hat{f}_i$ from $\mathcal{F}$ corresponding to x_i
    $Pa(x_i) \leftarrow$ Get parent values of x_i from x
    $\hat{x}_i \leftarrow \hat{f}_i$.predict($Pa(x_i)$)
    $error \leftarrow (x_i - \hat{x}_i)^2$
    $total\_error \leftarrow total\_error + error$
  end for
  $CAS \leftarrow \sqrt{total\_error/n}$ {Calculate Causal Anomaly Score (RMSE)}
  Append CAS to $Scores$
end for
return $Scores$

assumptions inherent to GES. These include acyclicity (the absence of feedback loops), faithfulness (all conditional independencies in the data are reflected in the graph structure), and, most critically, causal sufficiency (no unmeasured or hidden common causes of the observed variables). While the acyclicity and faithfulness assumptions are standard in many causal discovery settings, causal sufficiency may be violated in complex network environments. We proceed under the working assumption that our selected features capture the most direct causal influences, while acknowledging the potential for hidden confounders as a limitation of this study. We chose GES over constraint-based methods like the PC algorithm due to its demonstrated statistical efficiency and strong performance in finding the correct equivalence class of graphs from finite data.

2) Causal Mechanism Learning: Once the optimal causal graph $\mathcal{G}^*$ is determined, we learn the functional relationships for each variable. For each node $X_i \in V$, the SCM assumes that its value is generated by a function of its parents, $Pa_{\mathcal{G}^*}(X_i)$, and an independent noise term ϵ_i:

$$X_i := f_i(Pa_{\mathcal{G}^*}(X_i), \epsilon_i) \tag{2}$$

We train a predictive model $\hat{f}_i$ for each variable X_i to approximate this mapping. We investigate two distinct approaches for this task:

- **Causal-IDS (Linear):** A baseline approach assuming linear relationships, where each $\hat{f}_i$ is a Linear Regression model.
- **Causal-IDS (GBoost):** Our enhanced approach capable of capturing complex, non-linear relationships. The data is first standardized, and each $\hat{f}_i$ is a Gradient Boosting Regressor model.

The complete learned model consists of the pair $(\mathcal{G}^*, \{\hat{f}_1, \ldots, \hat{f}_n\})$.

B. Phase 2: Causal Anomaly Detection (Online)

For a new, unseen network flow instance $x = (x_1, \ldots, x_n)$, we quantify its conformity to the learned causal model. An intrusion is hypothesized to be an external intervention that forces one or more variables to deviate from their natural causal mechanisms. For each feature x_i, we use its learned model $\hat{f}_i$ to generate a prediction based on the values of its parents, $Pa_{\mathcal{G}^*}(x)$:

$$\hat{x}_i = \hat{f}_i(Pa_{\mathcal{G}^*}(x)) \tag{3}$$

The final Causal Anomaly Score (CAS) [13] for the entire instance is the Root Mean Squared Error (RMSE) of these causal errors across all features:

$$\text{CAS}(x) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \tag{4}$$

A high CAS value indicates a significant violation of the learned causal laws, signaling that the instance is likely malicious. A threshold is then applied to the CAS for classification.

IV. Experimental Evaluation

In this section, we present a comprehensive evaluation of the proposed Causal-IDS framework. We aim to answer two key research questions: (1) How critical is non-linear mechanism modeling for a causal-based IDS? (2) How does our enhanced Causal-IDS perform against standard, non-causal anomaly detection methods?

A. Dataset and Preprocessing

We conduct our experiments on the CIC-IDS2017 dataset [5], a widely recognized benchmark for evaluating IDSs. To ensure focused analysis and manageable computation times, we curated a representative subset by combining traffic from three specific days:

- Monday (Working Hours): Contains exclusively benign traffic, ideal for training our models.
- Friday Afternoon (DDoS): Contains a large volume of DDoS attack traffic.
- Friday Afternoon (PortScan): Contains Port Scanning attack traffic.

After cleaning invalid rows (containing NaN or Infinity values), our final dataset consists of 1,041,288 flow instances. The benign portion, comprising 754,459 instances, is used for training all models, while the entire dataset is used for evaluation. We selected the 12 key numerical, flow-based features listed in Table I.

This feature subset was chosen based on established practices in network traffic analysis and intrusion detection research. The selected variables represent fundamental properties of a network flow, such as its duration, packet counts, byte rates, and inter-arrival times. This approach prioritizes core traffic characteristics while excluding highly correlated or redundant features that can compromise the stability and interpretability of causal discovery algorithms.

TABLE I
SELECTED FEATURES FOR EXPERIMENTS

Feature Name	Feature Name
Flow Duration	Flow Bytes/s
Total Fwd Packets	Flow Packets/s
Total Backward Packets	Flow IAT Mean
Total Length of Fwd Packets	Fwd IAT Mean
Fwd Packet Length Mean	Init_Win_bytes_forward
Bwd Packet Length Mean	Init_Win_bytes_backward

B. Baselines and Evaluation Metrics

We compare our Causal-IDS variants against two widely-used anomaly detection baselines:

- Isolation Forest (IF): A tree-based ensemble method that identifies anomalies by their susceptibility to isolation [8].
- Autoencoder (AE): A neural network that learns a compressed representation of normal data, using reconstruction error as the anomaly score [9].

Performance is assessed using Precision, Recall, F1-Score, False Positive Rate, and the Area Under the Receiver Operating Characteristic Curve. For threshold-based metrics, we select the threshold as the 95th percentile of anomaly scores on the benign training data, targeting a 5% FPR.

C. Implementation Details

All models were implemented in Python. For **Causal-IDS**, we used the ‘causal-learn‘ library for GES discovery. The Linear model used ‘scikit-learn‘’s ‘LinearRegression‘, and the GBoost model used ‘GradientBoostingRegressor‘ with 50 estimators and a max depth of 3.

For the baseline models, we ensured a robust comparison by providing sufficient complexity and performing basic hyperparameter tuning. The Autoencoder, built with PyTorch, featured a symmetric architecture of 12-8-4-8-12 neurons with ReLU activation functions, an encoding dimension of 4, and was trained for 10 epochs using the Adam optimizer and Mean Squared Error (MSE) loss. For both the Isolation Forest and the Autoencoder, key hyperparameters (e.g., the ‘contamination‘ parameter for IF; latent dimension and learning rate for AE) were selected based on common practices and performance on a validation split of the training data. All experiments were conducted on a server equipped with an NVIDIA RTX 4090 GPU.

D. Results and Analysis

The comprehensive performance comparison is presented in Table II, with ROC curves shown in Fig. 2. All reported AUC values are consistent across the text, table, and figures.

TABLE II
PERFORMANCE COMPARISON OF ALL METHODS. OUR ENHANCED CAUSAL-IDS (GBOOST) ACHIEVES THE HIGHEST AUC WHILE MAINTAINING THE TARGET LOW FPR. ALL AUC VALUES ARE CONSISTENT WITH FIG. 2.

Method	Precision	Recall	F1-Score	FPR	AUC
Isolation Forest	0.2879	0.2923	0.2901	0.2748	0.6039
Autoencoder	0.6837	0.2843	0.4017	0.0500	0.6698
Causal-IDS (Linear)	0.0002	0.0000	0.0000	0.0500	0.4011
Causal-IDS (GBoost)	**0.6872**	**0.2889**	**0.4068**	**0.0500**	**0.8400**

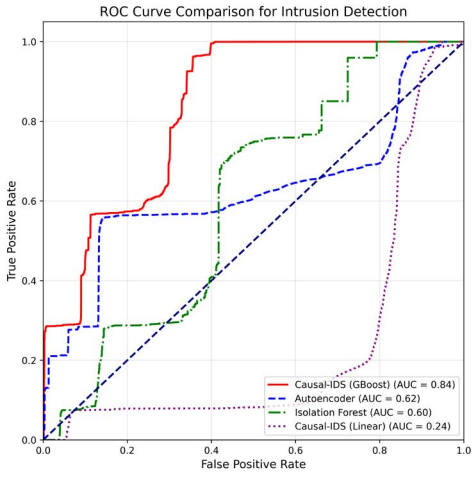

Fig. 2. ROC Curve comparison. Causal-IDS (GBoost) demonstrates superior discriminative power across all thresholds. AUC values in the legend should be updated to match Table II: GBoost (0.84), Autoencoder (0.67), Isolation Forest (0.60), and Linear (0.40).

The Importance of Non-Linear Modeling: The first key finding is the stark contrast between the two Causal-IDS variants. The Linear model completely fails (F1-

Score ≈ 0, AUC ≈ 0.40), indicating that its assumption of linear causal mechanisms is fundamentally flawed for complex network traffic. In contrast, the GBoost version achieves a strong F1-score of 0.4068. This result empirically validates our hypothesis that accurately modeling the complexity of causal mechanisms is as crucial as discovering the causal structure itself.

Comparison with Baselines: The Isolation Forest baseline suffers from an extremely high FPR (27.5%), rendering it impractical for real-world deployment. The Autoencoder is a much stronger baseline, achieving a competitive F1-score at the target 5% FPR. Impressively, our Causal-IDS (GBoost) model matches this F1 performance at the chosen threshold. However, the true superiority of our causal approach is revealed by the AUC metric. Our **Causal-IDS (GBoost) achieves an AUC of 0.84**, significantly outperforming the Autoencoder (0.67) and Isolation Forest (0.60). This indicates that the Causal Anomaly Score is a more robust and reliable indicator of malicious activity than reconstruction error or isolation scores across all decision thresholds.

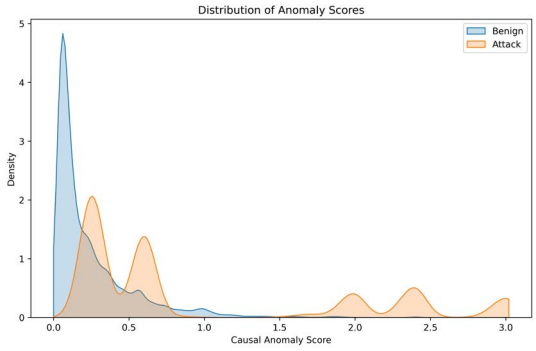

Fig. 3. Distribution of Causal Anomaly Scores for the enhanced Causal-IDS (GBoost) model, showing a clear separation between the scores of Benign and Attack traffic.

V. Discussion

Our experimental results give rise to several important discussion points regarding the application of causal inference to intrusion detection.

A. Why a Causal Approach Works

The superior performance of the Causal-IDS (GBoost) model, particularly its high AUC, suggests that a causality-based approach captures a more fundamental aspect of intrusions. While methods like Autoencoders are adept at learning a low-dimensional manifold for normal data, they are agnostic to the underlying data-generating process. An attack that is statistically novel but does not significantly deviate from this learned manifold might be missed.

In contrast, our Causal-IDS focuses on the *mechanisms* that produce the data. An attack, by its nature as an external intervention, often manipulates a specific feature (e.g., flooding a port, thus controlling packet rates) while leaving its normal causes unchanged. This breaks the learned functional relationship $X_i \approx \hat{f}_i(Pa_{\mathcal{G}^*}(X_i))$, resulting in a high Causal Anomaly Score for that feature. This focus on mechanism violation, rather than just distributional rarity, provides a more robust and direct detection signal.

B. Interpretability and Actionable Insights

A significant advantage of our framework is its inherent interpretability, which operates at two levels: the global causal graph and the local, instance-specific score. While the individual Gradient Boosting models are complex, the high-level causal graph $\mathcal{G}^*$ is human-readable, allowing an analyst to inspect the model's baseline assumptions about normal network behavior. More importantly, when an alert is triggered, the Causal Anomaly Score can be decomposed to provide actionable insights.

1) An Interpretability Walkthrough: To make this concrete, we provide a procedural walkthrough of how an analyst could diagnose an alert for a DDoS attack.

- **Inspect the Relevant Causal Sub-Graph:** First, the analyst examines the learned graph $\mathcal{G}^*$. Suppose it contains the plausible chain: `Flow Duration` $\rightarrow$ `Flow Packets/s` $\rightarrow$ `Flow IAT Mean`. This represents the normal mechanism where, for a given duration, the rate of packets causally influences their mean inter-arrival time.
- **Decompose the Anomaly Score:** A DDoS attack instance triggers an alert with a high CAS. The analyst's first diagnostic step is to decompose the total squared error, which defines the CAS, into its per-feature contributions: $(x_i - \hat{x}_i)^2$. They can then rank features by the magnitude of this causal error.
- **Pinpoint the Violated Mechanism:** In our DDoS scenario, the analyst would likely find that the feature `Flow Packets/s` has the largest causal error. The observed value, $x_{\text{packets/s}}$, is exceptionally high due to the attack flood. However, the model's prediction, $\hat{x}_{\text{packets/s}}$, which is based on the value of its cause `Flow Duration`, would be much lower, reflecting normal behavior for a flow of that duration.
- **Derive Actionable Insight:** The large discrepancy between the observed and predicted packet rate provides a clear diagnosis: the mechanism for `Flow Packets/s` has been violated. The feature is being manipulated by an external force, independent of its learned cause. This allows the analyst to conclude not just *that* the traffic is anomalous, but *why*: the packet rate is unnaturally

high for its context. This level of transparency is absent in purely black-box models and is crucial for building operator trust.

C. Limitations and Future Work

Despite the promising results, our work has several limitations that open avenues for future research.

- **Causal Discovery Accuracy:** As discussed in Section III, the performance of Causal-IDS depends on the correctness of the discovered graph, which hinges on assumptions like causal sufficiency. While we proceeded by selecting features known to be important, future work should explore advanced causal discovery algorithms that are robust to hidden confounders or can incorporate expert domain knowledge to constrain the graph search space.
- **Static Graph Assumption:** Our current model learns a single, static causal graph. However, network behavior can be dynamic, and the underlying causal relationships might change over time (concept drift). Developing methods for dynamically updating the causal model online is a crucial next step for long-term deployment.
- **Scalability:** While demonstrated on a substantial dataset, applying causal discovery to networks with hundreds or thousands of features remains computationally challenging. Exploring scalable causal discovery techniques or feature selection methods guided by causal principles would be a valuable research direction.

VI. Conclusion

In this paper, we addressed the critical challenge of high false positive rates in machine learning-based IDSs by proposing a paradigm shift from correlational to causal reasoning. We introduced Causal-IDS, a framework that models a network's normal behavior as a Structural Causal Model and identifies intrusions as violations of its learned cause-and-effect mechanisms.

Our extensive experiments on the CIC-IDS2017 dataset yielded two key findings. First, we demonstrated that accurately modeling the complex, non-linear relationships in network traffic is critical, as a naive causal model with linear assumptions failed entirely. Second, our enhanced Causal-IDS, combining causal discovery with robust Gradient Boosting models, proved highly effective. It not only matched a strong Autoencoder baseline at a fixed low false positive rate but also significantly surpassed all baselines in overall discriminative power, achieving a superior AUC of 0.84.

Acknowledgements

This work has been sponsored by the scientific research from Posts and Telecommunications Institute of Technology, Vietnam

References

[1] G. D. C. Bertoli, L. A. P. Júnior, O. Saotome, A. L. Dos Santos, F. A. N. Verri, C. A. C. Marcondes, S. Barbieri, M. S. Rodrigues, and J. M. P. De Oliveira, "An end-to-end framework for machine learning-based network intrusion detection system," *IEEE access*, vol. 9, pp. 106 790–106 805, 2021.

[2] M. N. Zaeem and M. Komeili, "Cause and effect: Concept-based explanation of neural networks," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, Oct. 2021, p. 2730–2736.

[3] K. N. Mallikarjunan, K. Muthupriya, and S. M. Shalinie, "A survey of distributed denial of service attack," in *2016 10th International Conference on Intelligent Systems and Control (ISCO)*. IEEE, Jan. 2016, p. 1–6.

[4] H. Sitter, W. Lorenz, H.-J. Klotter, and H. Lill, "Models for causality assessment," in *Handbook of Mediators in Septic Shock*. CRC press, 2019, pp. 500–522.

[5] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, vol. 1, no. 2018. SCITEPRESS - Science and Technology Publications, 2018, pp. 108–116.

[6] M. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, "Support vector machine and random forest modeling for intrusion detection system (ids)," *Journal of Intelligent Learning Systems and Applications*, vol. 06, no. 01, p. 45–52, 2014.

[7] P. H. Do, T. D. Le, V. Vishnevsky, A. Berezkin, and R. Kirichek, "A horizontal federated learning approach to iot malware traffic detection: An empirical evaluation with n-baiot dataset," in *2024 26th International Conference on Advanced Communications Technology (ICACT)*. IEEE, Feb. 2024, p. 1494–1506.

[8] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*. IEEE, Dec. 2008, p. 413–422.

[9] Y. Song, S. Hyun, and Y.-G. Cheong, "Analysis of autoencoders for network intrusion detection," *Sensors*, vol. 21, no. 13, p. 4294, Jun. 2021.

[10] V. Z. Mohale and I. C. Obagbuwa, "Evaluating machine learning-based intrusion detection systems with explainable ai: enhancing transparency and interpretability," *Frontiers in Computer Science*, vol. 7, May 2025.

[11] Y. Zhang, Z. Guan, H. Qian, L. Xu, H. Liu, Q. Wen, L. Sun, J. Jiang, L. Fan, and M. Ke, "Cloudrca: A root cause analysis framework for cloud computing platforms," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, ser. CIKM '21. ACM, Oct. 2021, p. 4373–4382.

[12] M. Chickering, "Statistically efficient greedy equivalence search," in *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, ser. Proceedings of Machine Learning Research, J. Peters and D. Sontag, Eds., vol. 124. PMLR, 03–06 Aug 2020, pp. 241–249. [Online]. Available: https://proceedings.mlr.press/v124/chickering20a.html

[13] Z. Liu, M. Gao, and P. Jiao, "Gcad: Anomaly detection in multivariate time series from the perspective of granger causality," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 18, 2025, pp. 19 041–19 049.

[14] S. Arif and M. A. MacNeil, "Applying the structural causal model framework for observational causal inference in ecology," *Ecological Monographs*, vol. 93, no. 1, Nov. 2022.

[15] A. Nazaret and D. Blei, "Extremely greedy equivalence search," *arXiv preprint arXiv:2502.19551*, 2025.