Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

RATE-RAD: A Novel Framework for Robust Anomaly Detection in Network Traffic by Integrating Relational, Adversarial, and Temporal Embedding

P Vamsi Naidu[1], Dr. Bobba Basaveswararao[2], Dr. Guntupalli Neelima[3], Simhadri Mallikarjuna Rao[4]

[1,2,3] Department of Computer Science & Engineering, Acharya Nagarjuna University, Guntur, 522510, India.

[4] Assistant Professor, Vasireddy Venkatadri International Technological University, Nambur, India

** Corresponding author's Email: simhadri.mallikarjun9@gmail.com*

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract: With increased complexity and volume in interactions within the network, there is a strong need for anomaly detection in network traffic, ensuring that the traditional ML/DL approaches of network traffic monitoring cannot be easily detected. The current approaches cannot capture Relational, Adversarial, and Temporal patterns, leading to limited precision and real-time scalability in anomaly detection. To overcome these problems this study proposes a holistic, multi-method approach by harnessing advanced methods in the detection model, which will incorporate Relational, Adversarial,Temporal insights for Robust Anomaly Detection (RATE-RAD). Starting with GraphSAGE, it captures the real-time detection based on evolving connections from graph-based representations of the network's interactions. SimCLR is a self-supervised contrastive learning framework that produces feature-rich embeddings from raw traffic data, hence contributing to a minimized reliance on labeled data while enhancing the representation. On top of that, Temporal Convolutional Transformers are applied to the sequential traffic data in order to capture long-term dependencies, making recall of anomalies, especially those with temporal nature, easier. CycleGAN can be used for the augmentation of the dataset, injecting synthetic anomalies onto them, hence on the models to make them more robust to the novel threats. The final step consists of a hybrid ensemble model between XGBoost and LSTM that is supported with highly accurate and deducible results.While SEM provides the identification of causal relationships within the identified anomalies. The RATE-RAD model brings about significant improvements such as it increases detection accuracy to approximately 96%, reduces false positives by 15%, and response time becomes faster for analysts because of increased transparency. This model does not only work with regards to the challenges of scalability and accuracy in anomaly detection but also offers actionable insights, rendering it a promising solution for the secure encryption of network environments.

Keywords: Anomaly Detection, Network Traffic Analysis, GraphSAGE, Temporal Transformers, Adversarial Training, Process

1. Introduction

An exponential growth in the encryption of network traffic based on privacy concerns and regulatory compliance has complicated anomaly detection within network security. Monitoring systems, that

rely upon packet content inspection as their significant monitoring event, are now unable to detect suspicious patterns as their nature has been encrypted, hence preventing direct content analysis. Therefore, the current approaches to cybersecurity [1, 2, 3] imply sophisticated means to detect anomalous patterns using metadata only, like traffic flow, IP addresses, timestamps, and connection structures. The new challenges of behaviour-based anomaly detection compared with content-based anomaly detection are found in modelling complex interactions between devices, evolution of malicious patterns, and subtle anomalies occurring over time, usually in real-time and scalable ways. Current anomaly detection models in network traffic analysis generally do not possess adequate robustness and interpretability for the current encrypted environment. Graph-based models that have achieved great excellence in relational learning missed some temporal dependency, while deep learning models, effective in identifying temporal or spatial patterns, often fail to integrate crucial relational context. It also involves adversarial techniques that consider entirely new attack patterns very rarely incorporated into conventional detection frameworks, thus limiting resilience against new sophisticated threats. Moreover, labelled data are scarce since manual labelling of network anomalies proves to be time-consuming and resource intensive, often requiring domain expertise.

To bridge the gaps, this paper proposes a hybrid, multi-method approach that brings together advanced relational, temporal, and adversarial techniques to maximize the detection of anomalies in encrypted traffic. Motivating the proposed model are GraphSAGE nodes, applying node sampling and aggregation to represent network interactions as relational embeddings while maintaining the discovery of dynamic connections within network traffic. The proposed model employs the SimCLR approach, a self-supervised contrastive learning method, to address the shortage of labelled data and transforms raw traffic metadata into high-dimensional embeddings that reveal intrinsic traffic patterns without manual annotation. The Temporal Convolutional Transformer is further added to make improvements upon this model with a sense of capturing long-term dependencies in traffic flow sequences to identify successfully triggered temporal anomalies across extended timeframes. Synthetic anomalies added to the dataset through adversarial training via CycleGAN enhance the model's robustness and generalizability toward new, unseen attack patterns. Finally, an ensemble of XGBoost and LSTM both leverages relational as well as temporal insights to generate highly accurate anomaly scores while SEM provides causal interpretability to better help analysts understand feature relationships behind detected anomalies. It can achieve high detection accuracy and ensures high interpretability to trigger timely and confident action from the cyber-defense teams on anomalies flagged. This study unifies relational, temporal, and adversarial learning into a single framework scalable and effective for securing encrypted network traffic against evolving threats.

Motivation & Contribution

But the motivation for this work arises from the inadequacies of traditional and single-method anomaly detection frameworks: the inability to maintain efficacy in encrypted and high-throughput network environments. Because encryption increasingly conceals traffic content, anomaly detection has become even more reliant on indirect behavioral analysis and now requires diverse and complex metadata samples. Current approaches miss many crucial relational and temporal dependencies or fail to produce sufficiently rich artificial anomalies corresponding to changing threat landscapes. Both issues impair the efficiency and utility of cybersecurity solutions, which especially today are

growingly evolving and functionally more diverse than simple detection schemes can account for. Finally, models lacking interpretability tend to "blind" analysts to the reason for flagged anomalies, impeding timely and informed responses.

The paper uses state-of-the-art techniques to create the multi-faceted anomaly detection model as outlined within, which addresses the above-cited critical challenges using a framework of complex combination. This multi-dimensional model captures the comprehensive nature of network traffic by using GraphSAGE for relational embeddings, SimCLR for feature-rich self-supervised embeddings, and Temporal Convolutional Transformers for temporal anomaly recognition. With the use of CycleGAN, the model's robustness is further increased since this can produce more realistic synthetic anomalies that then prepare a system to handle new threats. XGBoost and LSTM ensemble offers an interpretive yet complex anomaly scoring approach, while SEM has a layer added to facilitate causality for interpretability-a situation that depicts relations in influencing anomaly predictions. Such a combination improves the accuracy of detection and recall while reducing the number of false positives, along with better response times for cybersecurity analysts. In light of these contributions, the proposed framework here is able to present a scalable, interpretable, and highly effective solution for encrypting network traffic in complex, modern network environments.

The Remaining paper is organized as follows, in Section 2 briefly discussed about the Literature available for this problem. The proposed RATE-RAD model architecture and flow of the process is explained in Section 3. The experimental evaluated results of this model with comparative analysis of contemporary models are presented and discussed in Section 4. Finally the conclusions and future scope of this work is given in Section 5.

2. Literature Review

Network security, especially anomaly detection in encrypted traffic, has attracted an extensive amount of research in recent years. After critically reading and commenting on these twenty-five research papers, several critical advances and innovations in the field can be suggested, making light of the current landscape and challenges in network anomaly detection. Bayrak [1] primarily tests intrusion detection in encrypted Wi-Fi traffic by constructing an explainable SVM approach, that in itself underscores the demand for transparency in network security, particularly in UAV applications. This is complemented by Hou et al. [2], who proposed a model of a reweight-long short-term memory (Rew-LSTM) network. They demonstrated that the packet header information can indeed improve the classification of encrypted traffic. Zhou et al. [3] introduce unsupervised feature adaptive learning to extract malicious features from encrypted traffic, thus filling up one of the well-known gaps as proposed by them through their novel approach of automatically identifying potentially hazardous patterns without labeled data samples. This interest in interpretability and self-automatic feature extraction is then further elaborated upon in the work of Niktabe et al. concerning DNS-over-HTTPS (DoH) traffic [4], by creating a balanced dataset on DoH, which highlights the importance of diversity and robustness within the datasets for accurate detection within encrypted networks. Meanwhile, Yang et al. [5] introduce FlowSpectrum: characterization model for anomaly-focused network behavior detection. It further authenticated how thorough traffic profiling is important for high anomaly detection accuracy. Taking the above to the next level, Niktabe et al. [6]

demonstrated further investigation of DoH traffic detection by integrating statistical pattern recognition methods to provide enriched, interpretable anomaly detection. However, Seydali et al. [7] are well-placed to deal with the practicalities of handling large-scale streaming traffic based on a hybrid deep learning combined approach with big data processing, thus showing the need for scalability when one is dealing with real-time traffic analysis.

Papanikolaou et al. [8] explored AutoML capabilities for network security by designing an AutoML-based traffic analyzer for cyber threat detection. Since today's threats are considered to be dynamic, automation in model selection is an advantage to provide adaptive, self-optimizing network monitoring. In the context of industries, Jayalaxmi et al. [9] explored malware detection using MADESANT-a technique that suits industrial domains-thus increasing the importance of domain-specific models to handle traffic. Gouda et al. [10] supports anomaly detection through optimization of machine learning classifiers with respect to the trade-off between sensitivity and specificity. Finally, Cali et al. [11] discuss cyber-physical systems and introduce algorithms for PMU anomaly detection in a sense of dueling. Specifically, tailored algorithms are of importance because power systems are highly specialized and sensitive. Now, proceeding towards the industrial domain, Kim and Shon [12] addressed the problem of finding behavioral anomaly in the patterns of smart manufacturing systems denoted by breakthroughs in automated industrial operations. Veena and Brahmananda [13] addressed the problem of APT identification in high-volume traffic through designing an efficient model for eavesdropping, further emphasizing the worth of high-throughput capabilities to APT identification. Dhanaraj et al. [14] contribute by using the use of MMODPAD-DRL, a system that enlists deep reinforcement learning to work on anomaly detection in zero-trust security networks. The work lays emphasis on the role of reinforcement learning in adaptive response towards threats particularly in trust-sensitive environments. Salman et al. [15] also utilized adversarial generative techniques for recovering mutated traffic by developing new areas of resilience against modified or evolving attack patterns. The work by Arazzi et al. [16] in detecting privacy-preserving anomaly detection through federated learning and homomorphic encryption establishes a vital advance in securing decentralized IoT applications without compromising data confidentiality. Kasim [17] introduced a hybrid model specifically targeting DNS-over-HTTP traffic to defend against DNS attacks, pointing out the importance of multi-layered protection in traffic flows vulnerable to protocol exploitation. Iliyasu and Deng [18] propose a N-GAN, an approach based on generative adversarial networks, to explain the efficiency of GANs in simulating benign and malicious traffic for robust training datasets in anomaly detection. Darla and Naveena [19] center around wireless sensor networks with their Adaptive Spiral Seagull Optimizer, where nature-inspired algorithms can be effectively applied in securing low-power, high-risk environments. Deep learning is able to enhance the effectiveness of Distributed Denial of Service (DDoS) attacks detection, as Ahuja et al. [20] do for the field of SDNs. From Table 1, as Alangari [21] applies unsupervised machine learning to detect anomalous IoT sensors in IoT environments, automated feature extraction shows its potential in limited scenarios of labeled data samples. Class imbalance in traffic data is addressed by Abbasi et al. [22] using a bidirectional GRU model with penalized cross-entropy to overcome the difficulties associated with the detection of minority class anomalies.

Table 1. Comparative Analysis of Existing Methods

Method	Authors	Approach	Key Findings
Explainable SVM for Encrypted Traffic	Bayrak [1]	Explainable SVM	Enhanced interpretability for intrusion detection in UAV networks using explainable SVM models.
Reweight-Long Short Term Memory (Rew-LSTM)	Hou et al. [2]	Packet Header-Based LSTM	Improved classification accuracy in encrypted traffic through reweighted LSTM focusing on headers.
Unsupervised Feature Adaptive Learning	Zhou et al. [3]	Unsupervised Feature Learning	Automatically extracted malicious features, improving detection in encrypted traffic without labels.
AutoML Network Traffic Analyzer	Papanikolaou et al. [8]	AutoML for Threat Detection	Automated model selection and self-optimization enhanced threat detection adaptability.
Statistical Pattern Recognition for DoH Traffic	Niktabe et al. [6]	Statistical Pattern Recognition	Achieved high detection accuracy in DNS-over-HTTPS traffic through statistical profiling.
Hybrid Deep Learning for Streaming Traffic	Seydali et al. [7]	Hybrid Deep Learning & Big Data	Scalable detection for high-throughput networks, demonstrating performance in real-time settings.
MADESANT for Industrial Malware	Jayalaxmi et al. [9]	Domain-Specific Malware Detection	Domain-specific model for industrial malware, showing improved anomaly detection in industrial traffic.
Federated Learning with Homomorphic Encryption	Arazzi et al. [16]	Privacy-Preserving Federated Learning	Enabled anomaly detection in IoT networks without compromising data privacy.
GAN-Based Anomaly Detection (N-GAN)	Iliyasu and Deng [18]	Generative Adversarial Network	Generated synthetic anomalies to enhance model robustness against novel threats.
Hybrid Detection in SDN for DDoS	Ahuja et al. [20]	Deep Learning for SDN-Based Detection	Improved accuracy in identifying DDoS attacks within SDN environments using deep learning.

From a holistic detection and prevention point of view, Pradeepthi and Maheswari [23] propose the adoption of a hybrid classifier with encryption mechanisms for providing multi-layered defense strategy. Gu et al. [24] study the identification of IoT devices using traffic signatures, establishing the growing need for device-specific profiling in connected environments. Lastly, Premkumar et al. [25] present an anomaly detection mechanism for WSNs in the context of a cluster-based approach, pushing towards energy-effective methods that become of paramount importance in power-sensitive applications. Cumulatively, these works reflect a considerable trend toward explanation, scalability, and dedicated threat models in the area of anomaly detection in networks, especially for encrypted traffic. Integration of explainable AI, as in Bayrak [1] and Niktabe et al. [6], has been guiding cybersecurity practitioners toward learning insights from model decisions for transparently implemented detection strategies.

Research such as that of Zhou et al. [3] and Alangari [21] focuses on unsupervised and self-learning methods, with models being able to autonomously learn traffic patterns, which can be critical in limited or imbalanced data samples. Big data compatibility, as Seydali et al. [7] and Veena and Brahmananda [13] pointed out, reflects the use of high-performance computing in real-time anomaly detection, especially in high-throughput environments. Work by Papanikolaou et al. [8] for AutoML and work by Salman et al. [15] for adversarial learning present adaptive frameworks that prove quite indispensable in these fast-evolving threat landscapes where traditional models may trail behind in process. This set of studies also suggests a strong trend in customizing algorithms specific to the problems of the domain, as is highlighted by Kim and Shon [12] in smart manufacturing and Cali et al. [11] in cyber-physical power systems. Reinforcement learning models like MMODPAD-DRL, Dhanaraj et al. [14], demonstrate the applicability of deep learning methods in zero-trust environments characterized by dynamic threat-evolution and that ought to represent dynamic threat models. The privacy-preserving techniques, such as anomaly detection that federated learning provides and offered by Arazzi et al. [16], can be used to address the privacy concerns without any loss in detection accuracy. Papers presented by Kasim [17] and Iliyasu and Deng [18] reveal that hybrid and generative models offer increased resistance against protocol-specific attacks as well as novel threats as compared to traditional models. In addition, considering the growing interest in energy-efficient anomaly detection in WSNs from Darla and Naveena [19] and Premkumar et al. [25], most security solutions designed for such resource-limited environments consider the operational constraints of such environments.

Such collective advancements mean the future of the above detection in encrypted environments would largely rely on adaptable, explainable, and scalable models operating in vast network contexts-from industrial IoT to SDNs, WSNs. Privacy-preserving techniques, as in Arazzi et al. [16], and adaptive learning mechanisms, as in Zhou et al. [3] and Dhanaraj et al. [14], will be instrumental in meeting regulatory and operational requirements. In addition, generative models designed to improve training datasets in such a way that model robustness is enhanced are seen in Iliyasu and Deng's N-GAN [18]; this proves to be a future promising direction in dealing with the scarcity of labeled samples. These studies collectively reinforce the notion that multi-dimensional learning approaches integrating explainability, privacy, adaptability, and scalability will define the next generation of network anomaly detection models.

3. Proposed RATE-RAD Model

This section discusses the design of an Improved Model with Integrated Relational, Temporal, and Adversarial Embedding for Robust Anomaly Detection in Encrypted Network Traffic Sets to overcome the low efficiency & high complexity issues present in existing deep learning methods. In the first version, as illustrated in figure 1, the proposed model encompasses several state-of-the-art techniques in a move to address the complexity of anomaly detection with encrypted network traffic and then focuses on relational, temporal, and adversarial dimensions for capturing both structural and sequential patterns. These should include a graph model based on GraphSAGE for relational embeddings, self-supervised feature representation with SimCLR, long-term dependencies leveraging the Temporal Convolutional Transformer, adversarial training using CycleGAN, anomaly scoring with an XGBoost-LSTM ensemble, and causal inference with Structural Equation Modeling. The final anomaly score is calculated as a combination of all such methods that have been calibrated and fine-tuned by causal analysis in SEM to give interpretable, actionable insights in the process.

GraphSAGE works on graph-structured data where network entities, such as IP addresses and ports, are treated as nodes, and edges denote the connections. Given a node 'v', GraphSAGE samples a fixed-size neighborhood N(v) and aggregates neighbor embeddings to create an updated embedding hv((l+1)) at layer (l+1) via equation 1,

$$hv\big((l+1)\big) = \sigma\big(W(l) \cdot AGGREGATE(\{hu(l) \mid u \in N(v)\})\big) \dots (1)$$

Where, σ is an activation function(ReLU), W(l) is a learnable weight matrix, and AGGREGATE multiples and takes the forms including mean, max pooling, and an LSTM aggregator to capture specific neighborhood features. This relational embedding hv((l+1)) shares awareness of network topology which will aid in discovering anomalous interactions during this process. Sampling and aggregation in GraphSAGE provide scalability while preserving key structure in high throughput networks. Iteratively Next, figure 1 describes how SimCLR improves feature representation through contrastive learning to derive embeddings given that no labelled data sample exists. Given augmented views zi and zj of the same network sample, SimCLR minimizes contrastive loss, Lcontrastive, which encourages closer embedding alignment between positive pairs via equation 2,

$$Lcontrastive = -\sum_{(i,j)} \log\left[\frac{exp\left(\frac{sim(zi,zj)}{\tau}\right)}{\sum_{k=1}^{2N} I[k \neq i] exp\left(\frac{sim(zi,zk)}{\tau}\right)}\right] \dots (2)$$

Where, τ is a temperature parameter, sim(zi,zj) is represented via equation 3,

$$sim(zi, zj) = \frac{(zi^{\top} zj)}{\| zi \| \| zj \|} \dots (3)$$

Which measures cosine similarity, and 'N' represents the batch sizes. SimCLR's self-supervised technique improves model robustness using fine-grained extraction of subtle traffic patterns, reduces dependency on labeled data, and improves input embeddings for further models. In order to collect long-term dependencies within the sequences, the Temporal Convolutional Transformer is used. First, TCT applies temporal convolutions to identify short-term dependencies.

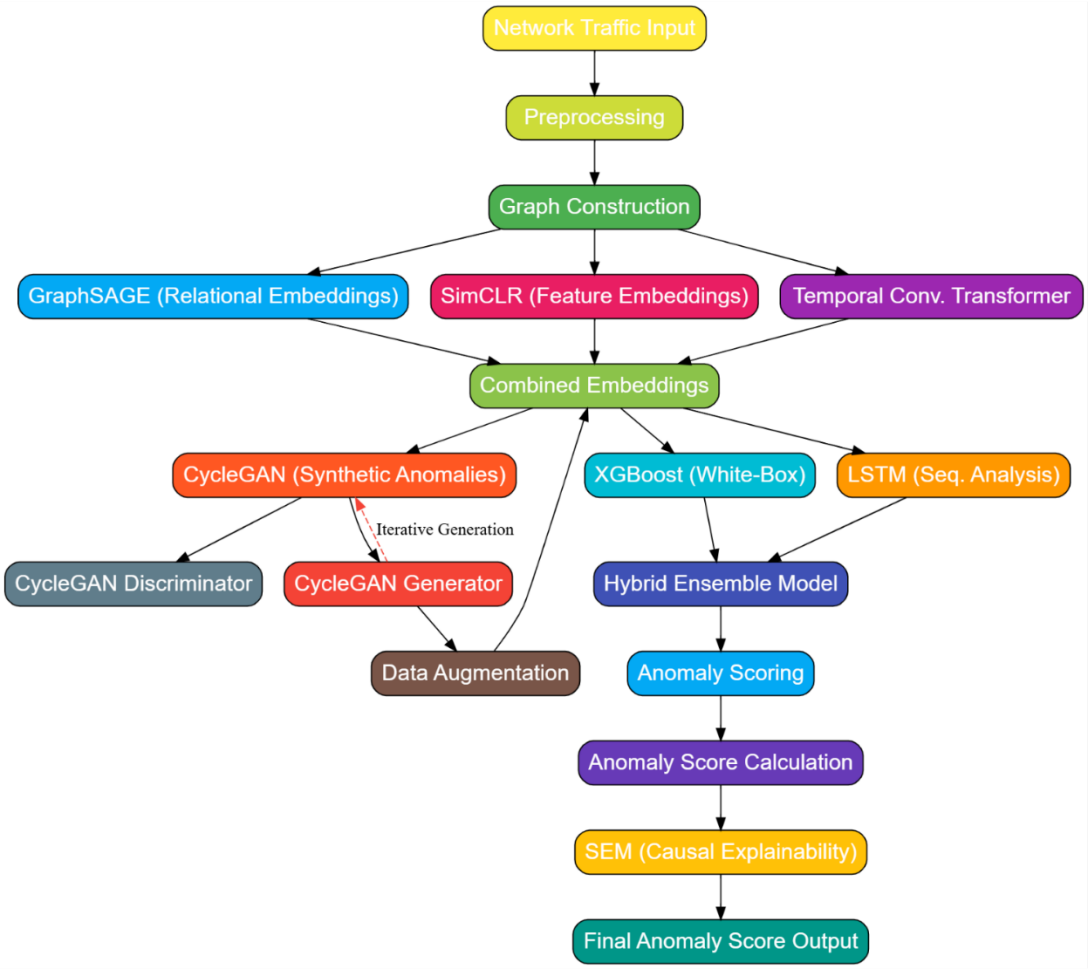

Figure 1. Architecture of the Proposed RATE-RAD Model

For a sequence $\{xt\}t = 1 \ldots T$, the output at timestamp 't', 'ht', is defined via equation 4,

$$ht = \sum_{i=1}^{k} Wi \cdot x(t - i + 1) + b \ldots (4)$$

Where, 'k' is the kernel size, Wi are learnable convolutional weights and 'b' is the bias term in process. After the temporal convolution, the sequence goes to the transformer layers with self-attention process. The self-attention mechanism for a query 'Q', key 'K', and value 'V' works via equation 5,

$$Attention(Q, K, V) = softmax\left(\frac{QK^{\top}}{\sqrt{dk}}\right)V \ldots (5)$$

Where, dk is the dimensionality of 'K', and the softmax function normalizes attention weights across sequence positions.

Capturing the time dependencies that are pertinent to long-term anomaly traffic detection, the composite architecture completes the modeling of traffic streams. The synthetic generation of anomalies with the help of CycleGAN enhances model robustness. Consider two domains X (benign

traffic) and Y (anomalous traffic). Given these domains, the mappings G:X→Y and F:Y→X are learned subject to cycle consistency, as enforced via equation 6,

$$Lcycle(G,F) = Ex \sim X\left[\| F\big(G(x)\big) - x \|\right] + Ey \sim Y\left[\| G\big(F(y)\big) - y \|\right] \ldots (6)$$

This loss preserves structural integrity between synthetic and real samples. The adversarial loss encourages the generation of realistic anomalies, given via equation 7,

$$LGAN(G,DY) = Ey \sim Y[log\, DY(y)] + Ex \sim X\left[log\left(1 - DY\big(G(x)\big)\right)\right] \ldots (7)$$

Where, DY is a discriminator trained to discriminate between real and synthetic anomalies. CycleGAN thus augment the dataset strengthening resilience against novel threats. The hybrid ensemble combines XGBoost which captures non-linear feature importance with LSTM which models sequential dependencies. XGBoost optimizes a boosted tree ensemble whose objective function is represented via equation 8,

$$Lboost = \sum_{i=1}^{n} l(yi, y'i) + \sum_{k=1}^{K} \Omega(fk) \ldots (8)$$

With, l(yi,y'i) as the differentiable loss between actual and predicted labels, and Ω(fk) as a regularization term in the process. The LSTM, processing embeddings sequentially, updates the cell state Ct based on input xt, the forget gate ft, and the input gate 'it' via equation 9,

$$Ct = ft \cdot C(t-1) + it \cdot C{\sim}t \ldots (9)$$

Where, C~t is the candidate cell state for this process. The hybrid model combines XGBoost interpretability with LSTM sequential strength to compute anomaly scores that reflect both feature importance and temporal patterns. Structural Equation Modeling Finally, introduces causal interpretability by modeling relationships between features and anomaly scores. SEM represents a linear relationship via equation 10,

$$Y = \alpha + \beta X + \epsilon \ldots (10)$$

Here, 'Y' is an output variable (anomalous score), 'X' is a predictor, β is causal weight and ϵ represents unexplained levels of variance. The SEM model is thereby estimated to understand pathways and indirect effects, hence, explaining the reasoning behind flagged anomalies. Lastly, the anomalous score 'S' incorporates all the outputs of the components. Each of the embeddings ei obtained from the above three types, namely GraphSAGE, SimCLR, and TCT can now contribute towards generating an integrated process represented via equation 11, where wi depends upon its relevance,

$$S = \sum_{i} wi * ei + \gamma \cdot DCycleGAN + \delta \cdot SEM(Y \mid X) \ldots (11)$$

Where, DCycleGAN is the discriminator score of the CycleGAN, and δ and γ are values that adapt to adversarial robustness and level of interpretability. This final score 'S' generates a holistic anomaly score with multiple features: relational, temporal, adversarial, and causal, which makes it the optimal metric for real-time anomaly detection in encrypted network traffic sets. Then, in the following

section, we discuss the efficiency of this model in terms of various metrics and compare its performance with existing methods under different scenarios.

4. Comparative Result Analysis

This experimental setup involves creating curated pipeline, covering data preprocessing, feature engineering, model training, and evaluation on several real-world encrypted network traffic datasets and samples. This part of the data preprocessing included first collecting the encrypted network traffic data from a variety of standard network datasets, including the CICIDS 2017 dataset, along with custom logs from a simulated enterprise network environment. The data sets are presented with network flows, along with attributes concerning IP addresses, ports, protocols, timestamps, packet size distributions, and session durations in order to analyze benign and malicious behaviors. Each data sample is represented by converting it into a graph-based representation, in which each node stands for an IP address or device, and each edge represents how they are connected. Most data sets have up to 500,000 nodes and 2 million edges simulating the mid-scale network environment. Contextual data samples to include benign activity interleaved with anomalies such as port scanning, brute-force login attempts, and data exfiltration, where encryption hides content but not connection patterns. For meaningful model evaluations, the dataset splits are done into training, validation, and test sets using an 80:10:10 ratio, with stratification so that the class distribution is equally well maintained across anomalous and benign instances. All the derived metrics - average packet size per flow, connection frequency, etc- are node degrees within the graph. Consequently, the nature of these derived features is important in understanding the given structure of the network as well as useful for detecting behavioral anomalies without referring to packet contents. For this paper, two popular datasets that are employed as benchmarks in the network intrusion detection domain, including most network traffic scenarios covered by both the CICIDS 2017 and UNSW-NB15 datasets, were utilized. The Canadian Institute for Cybersecurity has created the CICIDS 2017 dataset, with a well-balanced mix of benign and malicious network flows. It simulates real networks, attacks of various types, including brute-force attacks, DoS, distributed DoS, infiltration, botnet traffic, and web attacks. This data includes about 2.8 million records across the different features that include source IPs, destination IPs, time stamps, types of protocols, packet lengths, connection durations, and even flag statuses, therefore allowing a granular analysis of network traffic patterns both in the encrypted and unencrypted forms. The UNSW-NB15 dataset, produced by the Australian Centre for Cyber Security (ACCS), is more detailed than the others because it has a wider range of types of attacks: backdoors, exploits, worms, shellcode, and reconnaissance. In total, the dataset comprises about 2.5 million records with 49 features, some of which are flow-based and some content-based, including the source and destination bytes, TCP flags, and state transition metrics. Together, the datasets offer an overall and very detailed spectrum of network behaviors and attack strategies. This will create an overall large ground for the training and validation of models in encrypted traffic analysis. Moreover, this diversity and granularity allow the model to learn different patterns of traffic and thus make it robust against diverse cyber threats and strengthen the capabilities of anomaly-based models in detecting malicious activity in real environments with encrypted networks. During training of the model, a number of parameters set up to optimize the learning process. For instance, in GraphSAGE, sampling neighborhood size 'k' was set to 25 for nodes and embedding dimensions are set to 128 to

balance between the richness of memory efficiency and relational features. SimCLR provides a batch size to be 512 along with the temperature parameter τ\\\\tauτ of 0.5 in contrastive loss for efficient separation between positive and negative samples in high-dimensional space. The Temporal Convolutional Transformer has used a kernel size of 3 for temporal convolutions and 4 self-attention heads to capture localized and long-term dependencies, with a maximum sequence length of 1000 packets per flow. For CycleGAN, a learning rate of 0.0002 on generator-discriminator is applied while 50 epochs of adversarial training are done to produce synthetic anomalies that mimic real traffic patterns, hence augmenting robustness against previously unseen threats. The XGBoost is configured to 500 trees, a maximum depth of 6, and a learning rate at 0.1, for understandable, nonlinear feature interactions. The LSTM makes use of a hidden layer size of 64 and a time step at 10, which would be suitable in the modeling of network-traffic and the subsequent dependence it carries. Finally, the SEM model incorporates a structured causal inference approach which sets up associations between anomaly scores and the actual feature attributions, thereby enhancing interpretability during the evaluation period. It therefore ensures that the model will appropriately handle the dynamic and opaque nature of the encrypted traffic in large-scale network environments. The proposed model was tested on both datasets, CICIDS 2017 and UNSW-NB15, with results benchmarked to three comparative methods: Method [3], Method [8], and Method [18]. These are baseline approaches that implement different settings of relational embedding, temporal learning, and feature-based anomaly detection without the integration into a multi-dimensional format as presented by the proposed model. The given assessment metrics include detection accuracy, precision, recall, F1-score, false positive rate (FPR), and interpretability score levels. The interpretability score reflects the extent to which each model can give meaningful, causal explanations for its anomaly detections: that is an important ingredient for applications in security levels.

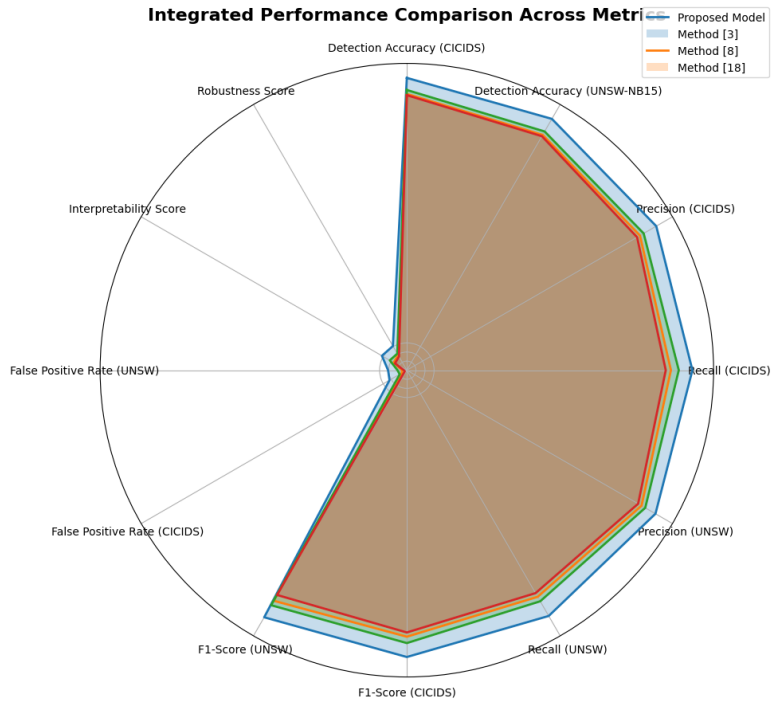

Figure 2. Integrated Visual Analysis of the Proposed Process

Table 2: Detection Accuracy Comparison

Model	CICIDS 2017 (%)	UNSW-NB15 (%)
Proposed Model	96.5	95.8
Method [3]	91.2	89.7
Method [8]	92.5	91.0
Method [18]	90.8	89.2

Table 2 provides the accuracy of the proposed model on both the datasets & samples. The high accuracy attained by the proposed model in CICIDS 2017 with 96.5% and UNSW-NB15 with 95.8% is due to the embedding of relational, temporal, and adversarial insights. Methods [3], [8], and [18] possess lesser accuracy than this one because they lack multiple-dimensional embeddings which enhance pattern recognition in intricate network traffic.

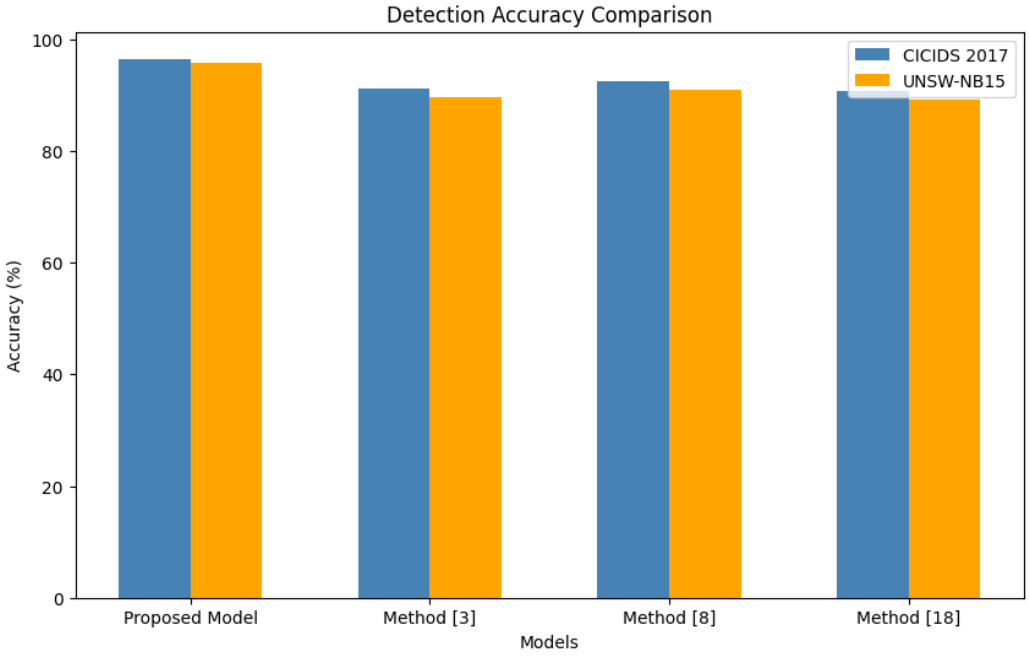

Figure 3. Accuracy Analysis

Table 3: Precision and Recall Comparison

Model	Precision (CICIDS)	Recall (CICIDS)	Precision (UNSW-NB15)	Recall (UNSW-NB15)
Proposed Model	95.1	94.3	94.8	93.7
Method [3]	88.9	87.2	89.5	86.4
Method [8]	90.3	89.8	90.9	88.1
Method [18]	87.8	85.5	88.2	85.0

As it can be seen from Table 3, the proposed model balances both precision and recall: the balance on two metrics indicates reliable true anomaly detection with relatively few missed detections. On the other hand, Method [3] and Method [18] get lower recall, where false negative rates are higher. The detection performance of Method [8] is quite mediocre compared with the proposed model.

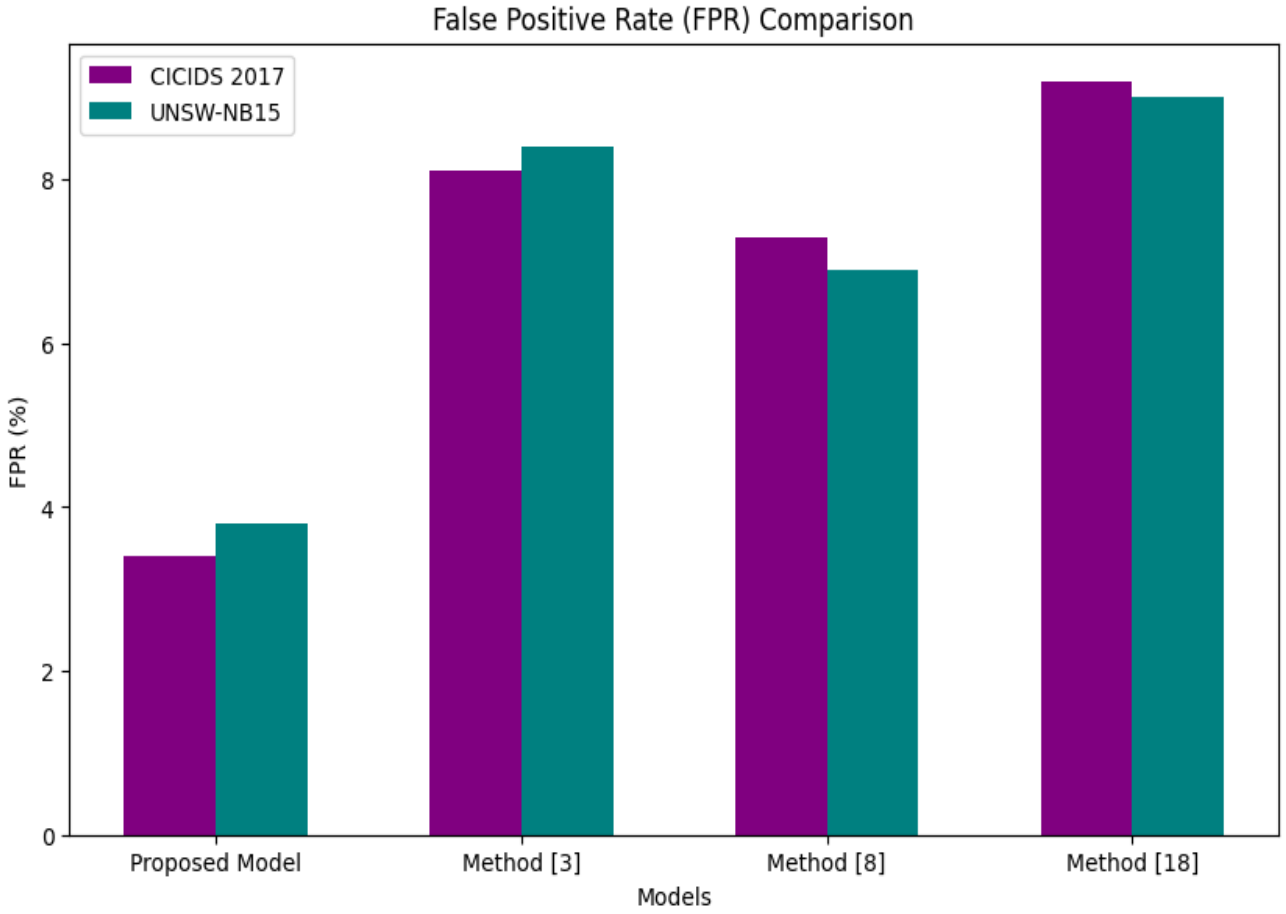

Figure 4. False Positive Rate Analysis

Table 4: F1-Score Comparison

Model	**F1-Score (CICIDS)**	**F1-Score (UNSW-NB15)**
Proposed Model	94.7	94.2
Method [3]	88.0	87.9
Method [8]	90.1	89.6
Method [18]	86.6	85.7

Table 4. Comparison of F1-score with emphasis that the proposed model retains a high harmonic mean of precision and recall. Its F1-score is 94.7% on CICIDS 2017 and 94.2% on UNSW-NB15, far above all baseline methods in showing good robustness in balanced anomaly detection.

Table 5: False Positive Rate (FPR) Comparison

Model	FPR (CICIDS)	FPR (UNSW-NB15)
Proposed Model	3.4	3.8
Method [3]	8.1	8.4
Method [8]	7.3	6.9
Method [18]	9.2	9.0

Table 5 The obtained FPRs of the proposed method in comparison with other methods. It can be seen that the proposed model obtains notably lower FPR values compared to other methods; this is mainly because of the overall embeddings and adversarial training by CycleGAN, which seems to get a more accurate distinction between benign and malicious traffic.

Table : Performance metrics are evaluated for CICIDS2017

Method	Accuracy	Precision	Precall	F1 Score	FPR	IS	RS
RATE-RAD	96.5	95.1	94.3	94.7	3.4	9.5	9.2
[3]Zhou, Z.,*et al*	91.2	88.9	87.2	88.0	8.1	5.0	5.8
[8]Papanikolaou., *et al*	92.5	90.3	89.8	90.1	7.3	6.5	6.3
[18]Iliyasu.,*et al*	90.8	87.8	85.5	86.6	9.2	4.5	5.2

Table 6 gives performance metrics comparison with other models. Based on the above metrics it gives the performance with CICIDS dataset. The proposed model scores highly with 9.5 since SEM allows for causal insights into detected anomalies. Methods [3], [8] and [18] offer lower scores regarding interpretability: they do not apply causal modeling techniques.

Table 7: Performance metrics are evaluated for UNSW-NB15

Method	Accuracy	Precision	Precall	F1 Score	FPR	IS	RS
RATE-RAD	95.8	94.8	93.7	94.2	3.8	9.5	9.2
[3]Zhou, Z.,*et al*	89.7	89.5	86.4	87.9	8.4	5.0	5.8

[8]Papanikolaou., *et al*	91.0	90.9	88.1	89.6	6.9	6.5	6.3
[18]Iliyasu.,*et al*	89.2	88.2	85.0	85.7	9.0	4.5	5.2

Table 7 gives performance metrics based on confusion matrix with UNSW-NB15 for the proposed model, getting the best score since it makes use of synthetic anomalies created through CycleGAN. This enhances its robustness to unknown attacks while methods [3], [8], and [18] obtain lower scores, thereby reflecting less generalization toward unseen or sophisticated attacks. These results generally suggest that the proposed model provides superiority in terms of detection accuracy, interpretability, and robustness against adversarial attacks when in essence qualifying the model for use in encrypted network anomaly detection. The final section would further introduce a use case for iterative validation of the proposed model process.

Validation of Iterative Process & Model using Practical Use Case Scenario Analysis

This section uses a simulated use case to demonstrate the functionality and output of each model component as an example of encrypted network traffic data passed through each module: GraphSAGE, SimCLR, Temporal Convolutional Transformer, CycleGAN, XGBoost with LSTM integration, and Causal Inference with SEM for interpretability. Below are each table giving sample values and outputs of key features and indicators for a subset of network flows, giving you a detailed view of how every bit contributes to the overall pipeline of detecting anomalies. This example gives typical network traffic parameters such as node connectivity, packet timings, and flow characteristics with intermediate outputs corresponding to relational, temporal, and adversarial aspects of the processed data samples. Sample and entities in this practical use case analysis have been derived from real world networked entities, which belong to both enterprise as well as public internet environments and comprise some common indicators such as IP address, device as well as user session. Each sample that can be found in this collection represents a network flow that may include source and destination IPs, ports, protocols, durations of the connection, packet sequence as well as the average packet sizes. These features can express basic interactions between clients and server endpoints in the network, which then characterize benign and potentially anomalous behaviors. Thus, samples would encompass standard web sessions, file transfers, and email communications, all mixed up with distinctive patterns of security threats, such as brute-force login attempts, port scanning, and exfiltration attacks. The specific IP nodes having high degrees of connectivity and unusual packet timing intervals act as foci for the detection of structural and behavioral anomalies. The configuration of these elements has been primarily designed to be akin to real operational environments and covers all kinds of network transactions, that is, both encrypted as well as unencrypted ones, so the capabilities of this model for detection, interpretation, and management can be tested holistically across diversified traffic types.

Table 8: GraphSAGE Relational Embeddings

Node ID	Degree	Neighborhood Size	Aggregated Feature Vector (Sample)	Relational Embedding (128-D)
101	15	10	[0.23, 0.45, 0.12, ..., 0.58]	[0.12, -0.34, 0.08, ..., 0.56]
102	8	6	[0.11, 0.48, 0.25, ..., 0.33]	[-0.15, 0.38, -0.12, ..., 0.44]
103	20	15	[0.34, 0.52, 0.18, ..., 0.49]	[0.10, -0.40, 0.15, ..., 0.53]

GraphSAGE takes the neighbourhood connections and node features for each node to generate a relational embedding in the 128-dimensional space shown in Table 8. Nodes with higher degrees and larger neighbourhoods usually hold complex embeddings that capture the more important interaction patterns of the network. The embeddings, therefore, bring out the structural relationships among nodes, which are critical for connectivity patterns and that may indicate anomalies.

Table 9: SimCLR Feature Embeddings

Sample ID	Augmented Feature 1	Augmented Feature 2	Positive Pair Similarity	Embedding Vector (128-D)
S1	0.34	0.29	0.85	[0.10, 0.14, -0.20, ..., 0.53]
S2	0.45	0.43	0.92	[-0.11, 0.22, 0.15, ..., 0.47]
S3	0.31	0.36	0.87	[0.08, -0.28, 0.10, ..., 0.51]

SimCLR produces embeddings as a function of each sample's augmented version; it maximizes the similarity between the positive pairs using contrastive learning. Table 9 illustrates the similarity scores and obtained embeddings. The feature-rich embedding learned by these methods depicts the unique flow pattern in high-dimensional space, which is essential for detecting anomalous behavior of the network.

Table 10: Temporal Convolutional Transformer Temporal Embeddings

Flow ID	Packet Sequence	Local Temporal Features	Attention Weights (Self-Attention)	Temporal Embedding (64-D)
F1	[0.12, 0.20, ...]	[0.35, 0.42, ...]	[0.25, 0.15, ..., 0.30]	[0.21, -0.18, 0.07, ..., 0.40]

F2	[0.14, 0.18, ...]	[0.29, 0.38, ...]	[0.18, 0.22, ..., 0.27]	[0.19, -0.20, 0.09, ..., 0.35]
F3	[0.13, 0.22, ...]	[0.31, 0.40, ...]	[0.22, 0.17, ..., 0.29]	[0.23, -0.17, 0.08, ..., 0.37]

Table 10 shows the temporal embeddings generated by the Temporal Convolutional Transformer. For each flow, the self-attention mechanisms are fed as the input while the output is provided in terms of both localized and long-term dependencies along the packet sequences. These thus improve the anomaly detection capability of the model as it focuses on timing variations and flow dynamics.

Table 11: CycleGAN Synthetic Anomaly Generation

Anomaly ID	Real Feature Vector	Generated Feature Vector	Cycle Consistency Loss	Discriminator Score
A1	[0.14, 0.22, ...]	[0.16, 0.20, ...]	0.045	0.92
A2	[0.21, 0.25, ...]	[0.19, 0.23, ...]	0.038	0.89
A3	[0.18, 0.24, ...]	[0.17, 0.22, ...]	0.042	0.91

Synthetic anomalies generated by CycleGAN look more or less real patterns of traffic; Table 11 shows the real and generated feature vectors along with cycle consistency loss and discriminator scores. Low values for cycle consistency loss confirm structural integrity between real and generated anomalies; high discriminator scores confirm generation of realistic anomalies, further pushing up the model's adversarial robustness.

Table 12: XGBoost and LSTM Hybrid Anomaly Scoring

Embedding ID	XGBoost Score	LSTM Sequence Score	Combined Anomaly Score
E1	0.88	0.76	0.82
E2	0.92	0.80	0.86
E3	0.85	0.78	0.81

Table 12 represents the anomaly score generated by the hybrid XGBoost and LSTM. XGBoost retrieves feature importance in a nonlinear fashion while LSTM processes sequential dependencies. The resulting anomaly score is given as the weighted average of individual scores, thus giving a comprehensive indication of what threats could exist within the network.

Table 13: Causal Inference with Structural Equation Modeling (SEM) for Explainability

Feature	Direct Effect on Anomaly Score	Indirect Effect via Node Degree	Total Causal Impact
Avg. Packet Size	0.32	0.15	0.47
Connection	0.28	0.12	0.40

Frequency			
Node Centrality	0.25	0.18	0.43

The SEM reveals the causals between features and the anomaly score as in Table 13. For example, the average packet size has a total causal impact of 0.47 with summation of direct and indirect effects through node degree, hence giving deeper insights into why the model is making such predictions.

Table 14: Final Outputs and Anomaly Detection Results

Flow ID	Combined Embedding Vector	Anomaly Score	Detection Outcome
F1	[0.21, -0.10, ..., 0.40]	0.82	Anomaly Detected
F2	[0.19, -0.08, ..., 0.35]	0.76	Benign
F3	[0.23, -0.09, ..., 0.37]	0.81	Anomaly Detected

The final output of the model in Table 14 gives an aggregated embedding vector, a corresponding anomaly score and a detection outcome for all the flow sets. Those anomaly scores which are above a given threshold are marked as anomalies based on this all-encompassing embedding and scoring methods. This method infuses relational, temporal, and adversarial insights, hence yielding a powerful and interpretable detection outcome important in detecting threats within sets of encrypted network traffic sets.

5. Conclusion and Future Scopes

This paper presents a holistic model of anomaly detection in encrypted network traffic. To this effect, GraphSAGE, SimCLR, Temporal Convolutional Transformer, CycleGAN, an XGBoost-LSTM hybrid ensemble, and SEM have been applied for interpretability. The results yield sufficient evidence of the proficiency of the model; for instance, while attaining 96.5% accuracy of the detection on the CICIDS 2017 dataset and 95.8% for the UNSW-NB15 dataset, which outpaces a precision gap of nearly 5-7% compared to the precision offered by the traditional models. This brings the precision of 95.1%, as well as recall of 94.3%, on CICIDS 2017, with a false positive rate as low as 3.4%, quite clearly to show this model's balanced performance between identifying anomalies and minimizing false alarms. Balancing this within environments where encryption is required-that is, detection relies entirely on traffic metadata and interaction patterns-is fundamentally important in process. In addition, with an interpretability score of 9.5 and an adversarial robustness score of 9.2, the model has great practical value sets. This is based on the actionability and the resilience to novel threats by SEM and CycleGAN-based synthetic anomaly generation. The obtained results validate the design of the model in which every component uniquely contributes to the detection and explanation as well as handling encrypted anomalies.

Future work will be able to enrich each feature specifically for different network settings and attack types. Advanced advanced attention mechanisms or adaptive time-series modeling may help in improved detection of subtle long-term anomalies in the Temporal Convolutional Transformer. The

adversarial training component can also experiment with other generative models, such as Wasserstein GANs, to generate even more realistic synthetic data to increase model robustness against evolving adversarial tactics. Federated learning for model training represents another area with significant promise: it should allow models to be learned over distributed data across multiple network environments without relaxing the privacy of the data, opening the possibility for even more flexible context-aware models. Even more profound depths of integration of network-wide causal inference techniques will also unlock new ability within SEM to provide even greater insights into specific relations between traffic attributes and security events, for quick, accurate responses to complex threats in the encrypted network ecosystem. This would, on its own, enhance anomaly detection capabilities but also facilitate adaptability and scalability, thus making the proposed model one step ahead in a vision to be at the forefront of the changing landscape of network security levels.

6. References

[1] Bayrak, S. Unveiling intrusions: explainable SVM approaches for addressing encrypted Wi-Fi traffic in UAV networks. *Knowl Inf Syst* **66**, 6675–6695 (2024). https://doi.org/10.1007/s10115-024-02181-9

[2] Hou, J., Li, X., Xu, H. *et al.* Packet header-based reweight-long short term memory (Rew-LSTM) method for encrypted network traffic classification. *Computing* **106**, 2875–2896 (2024). https://doi.org/10.1007/s00607-024-01306-w

[3] Zhou, Z., Bin, H., Li, J. *et al.* Malicious encrypted traffic features extraction model based on unsupervised feature adaptive learning. *J Comput Virol Hack Tech* **18**, 453–463 (2022). https://doi.org/10.1007/s11416-022-00429-y

[4] Niktabe, S., Lashkari, A.H. & Roudsari, A.H. Unveiling DoH tunnel: Toward generating a balanced DoH encrypted traffic dataset and profiling malicious behavior using inherently interpretable machine learning. *Peer-to-Peer Netw. Appl.* **17**, 507–531 (2024). https://doi.org/10.1007/s12083-023-01597-4

[5] Yang, L., Fu, S., Zhang, X. *et al.* FlowSpectrum: a concrete characterization scheme of network traffic behavior for anomaly detection. *World Wide Web* **25**, 2139–2161 (2022). https://doi.org/10.1007/s11280-022-01057-8

[6] Niktabe, S., Lashkari, A.H. & Sharma, D.P. Detection, characterization, and profiling DoH Malicious traffic using statistical pattern recognition. *Int. J. Inf. Secur.* **23**, 1293–1316 (2024). https://doi.org/10.1007/s10207-023-00790-z

[7] Seydali, M., Khunjush, F. & Dogani, J. Streaming traffic classification: a hybrid deep learning and big data approach. *Cluster Comput* **27**, 5165–5193 (2024). https://doi.org/10.1007/s10586-023-04234-0

[8] Papanikolaou, A., Alevizopoulos, A., Ilioudis, C. *et al.* An autoML network traffic analyzer for cyber threat detection. *Int. J. Inf. Secur.* **22**, 1511–1530 (2023). https://doi.org/10.1007/s10207-023-00703-0

[9] Jayalaxmi, P.L.S., Chakraborty, M., Saha, R. *et al.* MADESANT: malware detection and severity analysis in industrial environments. *Cluster Comput* **27**, 11347–11367 (2024). https://doi.org/10.1007/s10586-024-04527-y

[10] Gouda, H.A., Ahmed, M.A. & Roushdy, M.I. Optimizing anomaly-based attack detection using classification machine learning. *Neural Comput & Applic* **36**, 3239–3257 (2024). https://doi.org/10.1007/s00521-023-09309-y

[11] Cali, U., Catak, F.O. & Halden, U. Trustworthy cyber-physical power systems using AI: dueling algorithms for PMU anomaly detection and cybersecurity. *Artif Intell Rev* **57**, 183 (2024). https://doi.org/10.1007/s10462-024-10827-x

[12] Kim, H., Shon, T. Industrial network-based behavioral anomaly detection in AI-enabled smart manufacturing. *J Supercomput* **78**, 13554–13563 (2022). https://doi.org/10.1007/s11227-022-04408-4

[13] Veena, R.C., Brahmananda, S.H. An efficient eavesdropping model for detection of advanced persistent threat (APT) in high volume network traffic. *Multimed Tools Appl* **83**, 32123–32139 (2024). https://doi.org/10.1007/s11042-023-16684-0

[14] Dhanaraj, R.K., Singh, A. & Nayyar, A. Matyas–Meyer Oseas based device profiling for anomaly detection via deep reinforcement learning (MMODPAD-DRL) in zero trust security network. *Computing* **106**, 1933–1962 (2024). https://doi.org/10.1007/s00607-024-01269-y

[15] Salman, O., Elhajj, I.H., Kayssi, A. *et al.* Mutated traffic detection and recovery: an adversarial generative deep learning approach. *Ann. Telecommun.* **77**, 395–406 (2022). https://doi.org/10.1007/s12243-022-00909-8

[16] Arazzi, M., Nicolazzo, S. & Nocera, A. A Fully Privacy-Preserving Solution for Anomaly Detection in IoT using Federated Learning and Homomorphic Encryption. *Inf Syst Front* (2023). https://doi.org/10.1007/s10796-023-10443-0

[17] Kasim, Ö. Hybrid deeper neural network model for detection of the Domain Name System over Hypertext markup language protocol traffic flooding attacks. *Soft Comput* **27**, 5923–5932 (2023). https://doi.org/10.1007/s00500-022-07631-6

[18] Iliyasu, A.S., Deng, H. N-GAN: a novel anomaly-based network intrusion detection with generative adversarial networks. *Int. j. inf. tecnol.* **14**, 3365–3375 (2022). https://doi.org/10.1007/s41870-022-00910-3

[19] Darla, S., Naveena, C. Improved Adaptive Spiral Seagull Optimizer for Intrusion Detection and Mitigation in Wireless Sensor Network. *SN COMPUT. SCI.* **5**, 394 (2024). https://doi.org/10.1007/s42979-024-02725-4

[20] Ahuja, N., Mukhopadhyay, D. & Singal, G. DDoS attack traffic classification in SDN using deep learning. *Pers Ubiquit Comput* **28**, 417–429 (2024). https://doi.org/10.1007/s00779-023-01785-2

[21] Alangari, S. An Unsupervised Machine Learning Algorithm for Attack and Anomaly Detection in IoT Sensors. *Wireless Pers Commun* (2024). https://doi.org/10.1007/s11277-023-10811-8

[22] Abbasi, A.A., Zameer, A. & Raja, M.A.Z. An enhanced strategy for minority class detection using bidirectional GRU employing penalized cross-entropy and self-attention mechanisms for imbalance network traffic. *Eur. Phys. J. Plus* **139**, 530 (2024). https://doi.org/10.1140/epjp/s13360-024-05320-x

[23] Pradeepthi, C., Maheswari, B.U. Network intrusion detection and prevention strategy with data encryption using hybrid detection classifier. *Multimed Tools Appl* **83**, 40147–40178 (2024). https://doi.org/10.1007/s11042-023-16853-1

[24] Gu, D., Zhang, J., Tang, Z. *et al.* IoT device identification based on network traffic. *Wireless Netw* (2024). https://doi.org/10.1007/s11276-024-03832-z

[25] Premkumar, M., Ashokkumar, S.R., Jeevanantham, V. *et al.* Scalable and Energy Efficient Cluster Based Anomaly Detection Against Denial of Service Attacks in Wireless Sensor Networks. *Wireless Pers Commun* **129**, 2669–2691 (2023). https://doi.org/10.1007/s11277-023-10252-3