ITC 2/54
Information Technology and Control
Vol. 54 / No. 2/ 2025
pp. 593-612
DOI 10.5755/j01.itc.54.2.40247

SAEDF: A Synthetic Anomaly-Enhanced Detection Framework for Detection of Unknown Network Attacks

Received 2025/01/20 | Accepted after revision 2025/04/07

HOW TO CITE: Liang, K., Li, C., Duan, Q. (2025). SAEDF: A Synthetic Anomaly-Enhanced Detection Framework for Detection of Unknown Network Attacks. *Information Technology and Control*, 54(2), 593-612. https://doi.org/10.5755/j01.itc.54.2.40247

SAEDF: A Synthetic Anomaly-Enhanced Detection Framework for Detection of Unknown Network Attacks

Kai Liang, Chuanfeng Li, Qiong Duan

School of Computer and Information Engineering, Luoyang Institute of Science and Technology, 471023, Luoyang, China

Corresponding author: Chuanfeng Li, e-mail: lcf@lit.edu.cn

Detecting unknown cyber-attacks (i.e., zero-day) is difficult because network environments change frequently and there are few labeled examples of anomalies. Traditional methods for detecting anomalies often struggle to handle unknown attack types and work effectively with complex, high-dimensional data. To overcome these problems, we propose a new approach called the synthetic attack-enhanced detection framework (SAEDF). SAEDF combines synthetic anomaly generation, flexible feature extraction, and unsupervised anomaly detection. The framework employs a model known as the adaptive and dynamic generative variational autoencoder (ADGVAE). This model generates realistic synthetic attacks and adapts its structure to work effectively with datasets of varying complexity. This helps the model work well with a wide range of attack patterns while still being efficient. Tests on benchmark datasets show that SAEDF performs better than other methods. It achieves higher scores for F1, Recall, and has a much lower rate of false positives. These results show that SAEDF is effective in finding unknown attacks, improving detection accuracy, and handling complex and changing network traffic.

KEYWORDS: Unknown attack detection, synthetic attack anomalies, deep generative model, intrusion detection, network security

1. Introduction

Network security is confronting an increasingly complex challenge as cyberattacks become more sophisticated and develop in unpredictable ways. The risk posed by previously unknown attacks, often referred to as zero-day attacks, continues to rise. Conventional intrusion detection systems (IDS) that are anchored in signature-based detection methods are proving to be progressively ineffective at identifying these emerging threats [13]. Such systems are primarily limited to recognizing established attack patterns and are unable to generalize to new, unrecognized attack types. Consequently, there is a critical need for more adaptive and intelligent detection mechanisms.

Anomaly detection is an important component of network security to identify novel threats by analyzing patterns in network traffic data with the help of machine learning (ML). As a result, ML based anomaly detection systems are able to model normal behavior and raise an alarm at the onset of any anomaly that may indicate an attack [21]. Among these techniques, generative models such as Variational Autoencoders (VAEs) [20] and Generative Adversarial Networks (GANs) [16] have been found to be very useful. These models are particularly useful in learning the probability distribution of high dimensional data and therefore are able to capture the characteristics of the normal behaviour and any point that lies far from the learned distribution is an anomaly [8,19,40,41]. GAN have been successfully applied to unsupervised network anomaly detection and provided improved detection performance since they are able to identify anomalous traffic without needing any labelled data [35]. VAEs when combined with a deep neural network model (DNN) [39] outperformed traditional DNN in terms of accuracy for detection on standard datasets like NSL-KDD [34] and UNSW-NB15[25].

While machine learning techniques present significant potential for detecting network anomalies, several substantial challenges continue to hinder the development of effective systems. A critical barrier is the scarcity of labeled data for unidentified attacks. In the absence of labeled datasets that encompass examples of all possible attack types, it is challenging for machine learning models to accurately identify novel, previously unrecognized threats [33]. The problem is further exacerbated by the prevalence of imbalanced datasets. Typically, in network traffic, normal activity significantly outnumbers attack traffic, creating difficulties for training models to effectively differentiate between standard behavior and infrequent, anomalous occurrences [3]. Additionally, adapting existing intrusion detection models to recognize unknown attack types remains a complex issue [26]. A lot of studies use experiments that are based on certain assumptions or artificial data, but this does not really capture the messy and unpredictable reality of actual zero-day attacks. As a result, the true effectiveness of these systems in identifying actual zero-day attacks remains uncertain. To address these challenges, we employ adaptive and dynamic generative variational autoencoder (ADGVAE). Unlike GAN or conditional VAE, ADGVAE provides a dynamic and adaptive architecture that aligns with the evolving nature of network traffic features, offering improved generalization to unknown attack patterns.

1.1 Contribution of Paper

We propose a framework called synthetic anomaly-enhanced detection framework (SAEDF). The main idea behind SAEDF is to train a variational autoencoder (VAE) using a specialized data normalization method and feature selection approach to learn the latent representation of network attack traffic, enabling it to reconstruct anomalous traffic with high accuracy.

The contributions of this work are as follows:

Design and Development of the SAEDF: The SAEDF represents an innovative detection framework that integrates Variational Autoencoder (VAE)-based modeling of attack traffic, the synthesis of anomalous data, and unsupervised anomaly detection methodologies, in order to tackle the complexities associated with the identification of previously unknown threats.

Proposal for an ADGVAE model: The ADGVAE model features a projection layer along with a dynamic layer count mechanism (DLCM), which empowers it to adapt to datasets characterized by

diverse input dimensions and varying feature complexities. These attributes facilitate the model's capacity for dynamic structural adjustment, thereby enhancing feature extraction, mitigating the risks of overfitting and underfitting through adaptive depth, and improving both scalability and the quality of synthetic anomaly generation in varied and dynamic environments.

Cross-Dataset Generalization and Robustness of SAEDF: Comprehensive empirical investigations performed on established benchmark datasets, such as NSL-KDD, UNSW-NB15, and CICIDS2017, substantiate the holistic efficacy of SAEDF in the identification of previously unrecognized categories of attacks. The model demonstrates superior generalization capability, achieving consistently higher F1-Scores, Recall, and lower False Positive Rates (FPR) compared to baseline methods. Notably, cross-dataset evaluations highlight SAEDF's ability to adapt to unseen attack patterns and maintain robustness in minimizing false alarms, showcasing its practical applicability in dynamic and diverse real-world network environments.

1.2 Organization of Paper

This paper is organized as follows: Section 2 provides a review of pertinent literature on supervised learning, unsupervised learning, and generative methodologies pertinent to network attack detection. Section 3 details the proposed SAEDF framework, encompassing its architecture and methodology. Section 4 delineates the experimental setup and procedures utilized for the assessment of the framework. Section 5 engages in a discussion of the results, offering a comprehensive analysis of the strengths, limitations, and practical implications inherent to the proposed approach. Finally, Section 6 concludes the paper and delineates prospective avenues for future research.

2. Related Work

Machine learning-based intrusion detection (ID) methods have become fundamental to contemporary network security, providing automated and adaptive solutions for identifying and addressing both known and unknown cyber threats.

2.1. Supervised Learning-Based Methods

Supervised learning methodologies for network intrusion detection utilize labeled datasets to train models capable of categorizing network activities as either benign or malicious. Das et al. [6] proposed an ensemble framework t that integrates both supervised and unsupervised methodologies in order to identify distributed denial-of-service (DDoS) attacks, encompassing zero-day variants. Their approach integrated supervised models for identifying known attacks with unsupervised novelty detection models for zero-day threats, achieving high accuracy across datasets such as NSL-KDD, UNSW-NB15, and CICIDS2017. Ban et al. [4] introduced an IoT intrusion detection model leveraging convolutional neural networks with attention mechanisms (CNN-SE) and particle swarm optimization (APSO), which significantly improved classification accuracy on benchmark datasets such as UNSW-NB15 and NSL-KDD. Alashhab et al. [1] presented an ensemble online machine learning approach for DDoS detection in software-defined networks (SDN), demonstrating improved detection rates on datasets like CIC-DDoS2019 and custom SDN traffic data.

Despite their high detection accuracy, supervised learning-based methods depend heavily on labeled data, making them less effective in handling zero-day attacks or adapting to rapidly evolving threats. The labeling process is often time-consuming and labor-intensive, limiting the scalability of these methods in dynamic environments.

2.2. Unsupervised and Semi-Supervised Learning-Based Methods

Researchers have explored unsupervised and semi-supervised methods, which focus on detecting anomalies without requiring extensive labeled datasets. Pinto et al. [29] presented a novel methodology that integrates variational autoencoders (VAE) with long short-term memory (LSTM) architectures for the purpose of real-time anomaly detection within industrial Internet of Things (IIoT) systems, enhancing efficacy through the application of KL-divergence regularization. Truong and Le [36] introduced a privacy-preserving collaborative intrusion detection system (MetaCIDS)

using federated learning and unsupervised autoencoders, achieving detection accuracies between 96% and 99% across multiple datasets. Zhang et al. [42] introduced a two-stage intrusion detection system that uses light gradient boosting machine (LightGBM) and autoencoder technologies. This system employs recursive feature elimination (RFE) for feature selection and incorporates focal loss within the LightGBM framework to improve its learning efficiency. Falowo et al. [11] conducted a decadal longitudinal analysis of malware and DDoS attack evolution using unsupervised autoregressive integrated moving average (ARIMA) models to predict future attack trends.

Unsupervised and semi-supervised approaches demonstrate strong capabilities in identifying emerging threats and minimizing reliance on labeled datasets. Nevertheless, they often encounter considerable obstacles, such as overfitting, enhanced false positive rates, and a limited availability of anomalous traffic or unfamiliar attack data samples. These issues finally restrict the model's generalization ability, which may not sufficiently meet the demands of real-world attack scenarios.

2.3. Generative-based Methods

Generative models became an important focus in network intrusion detection because of their capability to produce synthetic data, improve model training with uneven datasets, and identify detailed attack patterns. These methods are especially useful for addressing zero-day attacks and handling rare or unseen threats in cybersecurity.

Dunmore et al. [9] conducted a comprehensive survey on the application of GANs in cybersecurity intrusion detection. They emphasized that GANs demonstrate exceptional capability in creating realistic synthetic samples for imbalanced datasets, thereby enhancing detection rates for infrequent attack types. Aldhaheri and Alhuzali [2] presented the SGAN-IDS framework, which uses GAN and self-attention mechanisms to produce adversarial attack traffic for IDS. Their approach evaluated various IDSs' detection rates against these synthetic attack flows, demonstrating an average reduction of false alarm rates by 15.93%. Peppes et al. [28] proposed the Zero-Day GAN (ZDGAN) to generate near-realistic synthetic data for zero-day attacks. By integrating this synthetic data with original datasets, their method improved the accuracy and robustness of deep learning classifiers while minimizing validation loss.

VAE-based approaches that improve anomaly detection through sophisticated feature extraction and representation techniques have been proposed [12, 22, 31, 37]. These studies employ VAE to tackle issues such as imbalanced datasets [12, 37], a scarcity of anomalous traffic samples [22], and the extraction of important latent features [31]. VAE-based methods have shown considerable promise in increasing detection accuracy and resilience against zero-day threats by refining feature representation and addressing data imbalance. Generative-based methods present substantial benefits; however, they face challenges associated with high computational complexity and restricted scalability when implemented in practical settings. The synthetic data generated by these models frequently fails to short in accurately representing the complexity and diversity of real-world attack scenarios, limiting their effectiveness in practical applications. Table 1 shows the summary of the related works discussed above.

Through the analysis of the aforementioned literature, it was found that while GANs are powerful for generating realistic data, they often suffer from training instability and mode collapse, which can pose significant challenges when dealing with high-dimensional network traffic data characterized by sparse anomalies. On the other hand, conditional VAEs provide conditional control during the generation process but rely on a fixed architecture, which limits their adaptability to datasets with varying feature complexities.

Therefore, the ADGVAE model proposed in this study is designed to address the limitations of models like GAN and CVAE when generating samples. It is particularly suitable for scenarios involving unknown input data characteristics, ensuring scalability across various network traffic scenarios. Moreover, it is well-suited for structured data scenarios, such as network anomaly attack traffic. Its variational framework maintains stable training dynamics and reduces training time.

Table 1
Summary of the related work.

Reference	Existing Technique	Summary
Das et al., (2024) [6]	Ensemble Learning: Supervised and Unsupervised	**Classifier:** ensemble model (SVM, NN, DT, One-class SVM) **Dataset:** NSL-KDD, UNSW-NB15, CICIDS2017 **Measures:** TN FN FP TP
Ban et al., (2024) [4]	Adaptive CNN with APSO	**Classifier:** APSO, CNN, Channel Attention Mechanism **Dataset:** UNSW-NB15, NSLKDD **Measures:** TP FN FP TN
Alashhab et al., (2024) [1]	Ensemble Online Machine Learning	**Classifier:** Ensemble Online Machine Learning **Dataset:** CICDDoS2019, InSDN, Slowread-DDoS, Custom dataset **Measures:** Accuracy Precision Recall F1score FAR
Pinto et al., (2024) [29]	Unsupervised Learning: VAE and LSTM	**Classifier:** VAE, LSTM mode **Dataset:** SWAT **Measures:** Precision Recall F1score AUC TP TN FP FN
Truong et al., (2023) [36]	Federated Learning and Blockchain	**Classifier:** Autoencoder, Attention Classifier and Federated Learning (FL) **Dataset:** 4 Network Intrusion Datasets **Measures:** Accuracy Precision Recall
Zhang et al., (2023) [42]	Two-stage Intrusion Detection: LightGBM and Autoencoder	C**lassifier:** LightGBM and Autoencoder **Dataset:** NSL-KDD, UNSW-NB15 **Measures:** Accuracy Precision Recall F1
Falowo et al., (2024) [11]	Time Series Analysis: ARIMA	**Classifier:** ARIMA **Dataset:** CSIS Database, DBIR
Dunmore et al., (2023) [9]	Generative Adversarial Networks (GANs)	**Classifier:** GANs **Dataset:** Multiple Public Datasets
Aldhaheri et al., (2023) [2]	Self-Attention GAN for IDS	**Classifier:** SGAN, Self-Attention Mechanism **Dataset:** Not specified **Measures:** Detection rate Precision Recall F1
Peppes et al., (2023) [28]	GANs for Zero-Day Attack Data Generation	**Classifier:** GAN, Neural Network **Dataset:** SWAT **Measures:** Precision Recall F1-score
Wang et al., (2024) [37]	VAE-LSTM-DRN for Encrypted Traffic	**Classifier:** VAE, LSTM, Deep Residual Network (DRN) **Dataset:** Tor, VPN Datasets **Measures:** Accuracy Recall Precision F1-score
Prabakaran et al., (2023) [31]	Deep Learning for Phishing Detection	**Classifier:** VAE **Dataset:** ISCX-URL2016, Kaggle Datasets **Measures:** Confusion matrix Precision Recall F1-score
Fathima et al., (2024) [12]	Hybrid Framework: GRU-VAE	**Classifier:** GRU-VAE **Dataset:** CIC-IDS-2017, CIC-IDS2018 **Measures:** Accuracy Precision Recall F1-score Temporal Correlation Index (TCI)
Liu, (2023) [22]	AI-based DDoS Attack Protection	**Classifier:** VAE **Dataset:** Not specified **Measures:** Accuracy Precision Recall F1-score

3. Proposed Framework: SAEDF

Hence to accurately identify zero-day anomalies, we combined synthetic anomaly generation with VAE-based modeling and an anomaly-enhanced unsupervised detection process. A novel zero-day attack detection framework (SAEDF) was developed by integrating data resampling, synthetic anomaly generation, and anomaly-enhanced unsupervised detection. To improve detection performance, we applied advanced data preprocessing techniques, as discussed in Section 3.2, to normalize and optimize the feature space for model training.

3.1. Workflow

SAEDF enhances the detection of unknown attack patterns through three stages: data resampling, synthetic anomaly generation, and anomaly-enhanced detection, as shown in Figure 1. The procedure commences with data resampling, during which a GMM-based approach is employed for the normalization and quantification of selected attack samples. This is followed by a four-step feature selection process aimed at enhancing the dataset for synthetic attack samples.

Synthetic anomaly generation is then performed, with a VAE sampling regions beyond the normal distribution to create realistic synthetic anomalies. This step enriches the dataset and enhances the model's ability to detect unknown attack patterns. Anomaly-enhanced detection integrates both synthetic and real anomalies to train an unsupervised detection model, improving its generalization to unknown attacks. Detailed algorithms and settings for each stage are provided in subsequent sections.

3.2. Data Preprocessing

The benchmark datasets (NSL-KDD, UNSW-NB15, and CICIDS2017) contain diverse features representing various network traffic characteristics. These datasets encompass numerical and categorical attributes, varying scales, and occasionally missing or inconsistent values, necessitating a systematic preprocessing approach.

3.2.1. Normalization

The benchmark datasets exhibit a mix of continuous and discrete features, reflecting the diverse nature of network traffic data. Continuous features, such as packet size and flow duration, span wide numerical

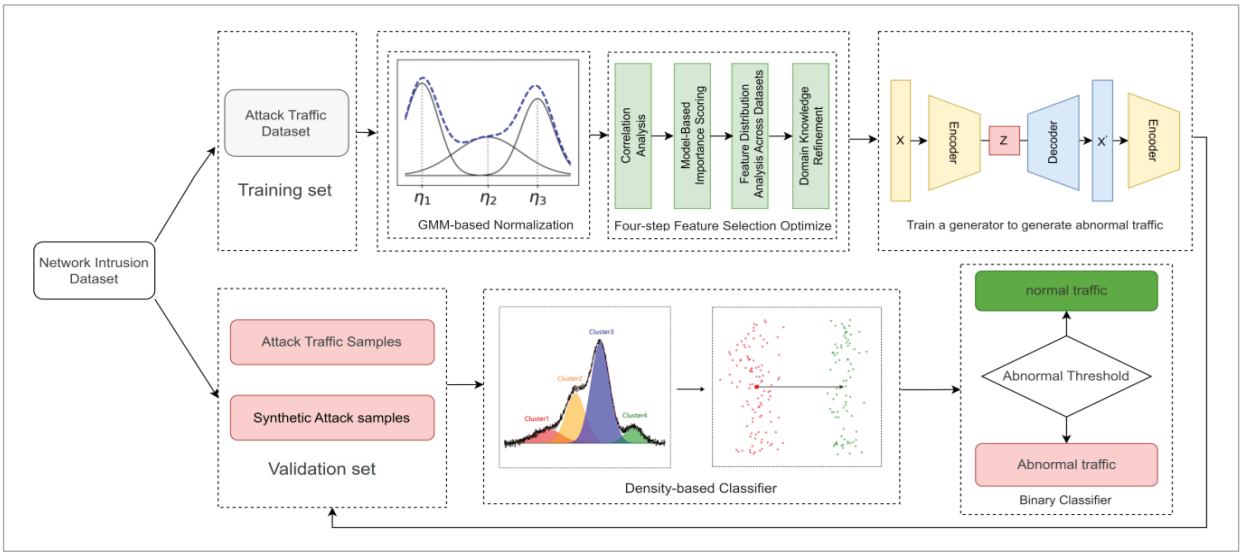
Figure 1

Workflow of SAEDF Data Resampling:

GMM-based normalization, quantification, and a four-step feature selection optimize.

Synthetic Anomaly Generation: The VAE generates synthetic anomalies by sampling abnormal regions in the latent space.

Anomaly-Enhanced Detection: A binary classifier is trained on real and synthetic anomalies to detect known and unknown attacks



ranges, while discrete features, such as protocol type and connection status, are represented by categorical values. Handing these differences is critical for ensuring uniformity in data representation and optimizing model performance.

To address this variability, the specialized normalization technique introduced in CTGAN [38] was employed. This method demonstrated considerable effectiveness in managing mixed-type data by normalizing continuous variables and encoding categorical variables while maintaining their intrinsic distributions and relationships. Further details regarding this approach are outlined in Section 3.3.2.

3.2.2 Feature Selection

Feature selection is a fundamental process that substantially improves the effectiveness of predictive models by identifying the most pertinent features, thereby reducing both computational expenses and training time. In this study, a systematic approach was employed to refine features from the benchmark datasets, ensuring adaptability across their varying characteristics and complexities. The main methods applied in this study are outlined below.

Correlation Analysis: Statistical techniques were used to compute the correlation coefficients between independent features and target labels [17]. Features with low correlation were eliminated to reduce redundancy and irrelevant information. A heatmap was generated to visually represent the relationships and assist in feature interpretability.

Model-Based Importance Scoring: An extra trees classifier [15] was employed to assess feature importance scores. Features with higher scores were considered essential due to their strong contribution to the target variable. This method provided a quantitative basis for feature prioritization.

Feature Distribution Analysis Across Datasets: A comparative analysis was performed to evaluate the consistency of feature significance across the benchmark datasets [27]. This guaranteed that the chosen features were resilient and preserved their ability to differentiate.

Domain Knowledge Refinement: Domain-specific expertise in network security was applied to manually refine the feature subset. This step incorporated features such as packet entropy and flow duration [32]. Through the systematic application of the aforementioned methodologies, a refined feature subset was derived, effectively capturing the key attributes of network traffic within the benchmark datasets. Detailed insights into the selected features and their statistical significance are presented in Section 4.2.

3.3. Synthetic Anomaly Generation

The ADGVAE focused on generating synthetic anomalies specifically from attack categories in benchmark datasets. To present the structure of ADGVAE concisely and accurately, we defined the relevant notations and used them to describe the model design in detail.

3.3.1. Notations

$x_1 \oplus x_2 \oplus \ldots$ Concatenate vectors x_1, x_2 ...

$FC_{u\to v}(x)$: Perform a linear transformation on a u-dimensional input to produce a v-dimensional output.
BN: batch normalization [18].

r: A row of the dataset, including both continuous and categorical features.

z: Latent variable representing the compressed representation of r.

N_c: Number of continuous features in the dataset.

N_d: Number of categorical features in the dataset.

α_i, β_i, d_i: Representations of the i-th continuousfeature (scalar and one-hot) and categorical feature (one-hot), respectively.

3.3.2. Anomaly Sampling and Feature Representation

The attack category data was first sampled from the dataset, rather than using the full dataset. This ensured that the ADGVAE model focused on learning the distribution of attack data, thereby improving its ability to generate diverse and high-quality synthetic anomalies. For continuous features, a variational gaussian mixture model (VGM) [38] was used to estimate modes, normalize values within selected modes, and represent them as a combination of a scalar and one-hot mode vector. For categorical features, one-hot encoding was applied, and the final representation combined normalized continuous features and one-hot encoded categorical features for each row.

In this study, the VGM was employed to capture the multimodal nature of continuous features by modeling their underlying distributions with a mixture of Gaussian components. This approach ensured a precise rep-

resentation of complex data patterns. For each value, probability densities across modes were computed, a mode was sampled, and the feature value was normalized within the selected mode using the formula:

$$\alpha = (c - \eta)/(4\varphi) \tag{1}$$

$$\beta_i^{(k)} = \begin{cases} 1 & \text{if k = selected mode index} \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where c represents the feature value, η denotes the mean of the selected mode computed from the VGM, and φ indicates the standard deviation of the selected mode computed by the VGM. Represent the continuous feature as a concatenation of α (scalar) and a one-hot vector β indicating the sampled mode.

For each row of data, the normalization is expressed as:

$$r = \alpha_i \oplus \beta_i \oplus \ldots \oplus \alpha_{N_c} \oplus \beta_{N_c} \oplus d_i \oplus \ldots \oplus d_{N_d} \tag{3}$$

3.3.3. ADGVAE Model Structure

The ADGVAE model incorporated three key components: a projection layer (d_{proj}), a dynamic hidden layer structure, and a dynamic layer count mechanism, which together enabled adaptability to diverse datasets. The projection layer (d_{proj}) standardized input dimensions from different datasets, ensuring consistent input representation for the model.

Projection Layer (d_{proj}): The projection layer was introduced to standardize the input dimensions of the datasets, mapping the raw input features r (with dataset-specific dimensionality $|r|$) to a fixed dimension d_{proj}, which served as the starting point for the encoder. This modification ensured that the model could effectively process datasets with variable input dimensions, thereby providing a uniform input format for the subsequent network.

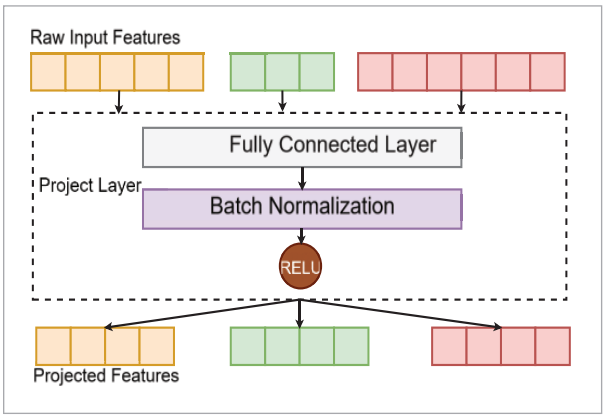
As shown in Figure 2, the transformation is defined as:

$$r_{\text{proj}} = \text{ReLU}\left(BN\left(FC_{|r|\to d_{\text{proj}}}(r)\right)\right). \tag{4}$$

Here, d_{proj} is a predefined projection dimension and the detailed dimensionality reduction methods are provided in Section 4.2.2.

$FC_{|r|\to d_{\text{proj}}}$ is a fully connected layer that maps the original input to the projection space.

Figure 2
Principle of projection layer.



Encoder ($q_\varphi(z|r)$): The encoder mapped the input projection r_{proj} to the latent space z of dimension d_{latent}. It consisted of dynamically determined hidden layers (n_{hidden}) where each layer reduced the dimensionality by half until d_{latent} was reached. The encoder transformation is defined as:

$$h_1 = \text{ReLU}\left(FC_{d_{\text{proj}}\to\frac{d_{\text{proj}}}{2}}(r_{\text{proj}})\right) \tag{5}$$

$$h_2 = \text{ReLU}\left(FC_{\frac{d_{\text{proj}}}{2}\to\frac{d_{\text{proj}}}{4}}(h_1)\right) \tag{6}$$

$$h_{n_{\text{hidden}}} = \text{ReLU}\left(FC_{\frac{d_{\text{proj}}}{2^{n_{\text{hidden}}\cdot(n_{\text{hidden}}-1)}}\to d_{\text{latent}}}(h_{n_{\text{hidden}}-1})\right) \tag{7}$$

At the final layer, the encoder output the parameters of the latent distribution:

$$\mu = FC_{\frac{d_{\text{proj}}}{2^{n_{\text{hidden}}\cdot-1}}\to d_{\text{latent}}}(h_{n_{\text{hidden}}}) \tag{8}$$

$$\sigma = \exp\left(\frac{1}{2}\cdot FC_{\frac{d_{\text{proj}}}{2^{n_{\text{hidden}}\cdot-1}}\to d_{\text{latent}}}(h_{n_{\text{hidden}}})\right), \tag{9}$$

where μ and σ are the mean and standard deviation of the latent variable z. The latent variable is sampled as:

$$q_\phi(z \mid r) \sim \mathcal{N}(\mu, \sigma^2 I). \tag{10}$$

Decoder ($p_\theta(r|z)$): The decoder replicated the architecture of the encoder, methodically augmented the dimensionality from d_{latent} back to d_{proj}, and ultimately reverted to the original input dimensionality $|r|$.The decoder transformation is defined as:

$$h_1 = \text{ReLU}\left(FC_{d_{\text{latent}} \to \frac{d_{\text{proj}}}{2^n \cdot \text{hidden}^{-1}}}(z)\right) \quad (11)$$

$$h_2 = \text{ReLU}\left(FC_{\frac{d_{\text{proj}}}{2^n \cdot \text{hidden}^{-1}} \to \frac{d_{\text{proj}}}{2^n \cdot \text{hidden}^{-2}}}(h_1)\right). \quad (12)$$

The hidden layer $h_{n_{\text{hidden}}}$ of the decoder is the same as equation (7). At the final layer, the decoder reconstructed the original input:

$$r_{\text{reconstructed}} = FC_{d_{\text{proj}} \to r_i}\left(h_{n_{\text{hidden}}}\right) \quad (13)$$

Loss Function: The ADGVAE model was trained by optimizing the evidence lower bound (ELBO) [32]. The ELBO is given as:

$$\mathcal{L} = E_{q_\phi(z|r)}[\log p_\theta(r|z)] - \text{KL}\left(q_\phi(z|r)|p(z)\right), \quad (14)$$

where the first term represents the reconstruction loss, and the second term is the Kullback-Leibler (KL) divergence, which regularizes the latent variable z to follow a standard Gaussian distribution $p(z) = N(0,I)$.

Dynamic Hidden Layer Count Mechanism: The dynamic hidden layer count mechanism determined the optimal number of hidden layers (n_{hidden}) for the encoder and decoder by training, based on the input dimensions of the dataset. Unlike fixed architectures, this mechanism learned the number of layers that yielded the best performance for datasets with varying complexities. The determination of n_{hidden} was guided by the input dimension of the dataset (d_{input}), the latent space dimension (d_{latent}), and a dataset-specific complexity factor (w_{dataset}).

The number of hidden layers is dynamically computed as follows:

$$n_{\text{hidden}} = \left\lceil w_{\text{dataset}} \cdot \log_2\left(\frac{d_{\text{input}}}{d_{\text{latent}}}\right)\right\rceil. \quad (15)$$

The specific determination process of w_{dataset} is elaborated in Section 4.2.2

3.4. Anomaly-Enhanced Detection

The anomaly-enhanced detection module was developed to detect anomalies by using a comprehensive dataset that includes both original traffic data and synthetic anomalies produced by the ADGVAE module. This module enhanced the decision boundary between normal and anomalous samples by modeling their distributions and computing scores based on density and distance metrics. The detection process consists of two main steps: density-based modeling and anomaly scoring and classification.

In the density-based modeling step, a gaussian mixture model (GMM) [5] was used to estimate the probability distribution of latent representations. GMM modeled the data distribution as a mixture of Gaussian components, each characterized by a mean, variance, and weight. The expectation-maximization (EM) algorithm [7] was employed to learn these parameters, allowing the GMM to identify high-density regions where normal samples resided, while anomalies, including synthetic samples, are located in lower-density areas.

The anomaly scoring and classification step computed an anomaly score for each sample by combining density-based [10] and distance-based metrics [14]. The density-based score evaluated how well a sample fit the learned Gaussian components, with lower scores indicating potential anomalies. The distance-based score measured the proximity of a sample to the nearest Gaussian center, with larger distances suggesting a higher likelihood of being anomalous. Table 2. shows the complete parameters and strategy used in the proposed approach.

Table 2
Anomaly attack detection model parameters.

Model	Parameters	Value
GMM	Number of Components (K)	5
Anomaly Scoring	Density Weight (λ1)	0.7
Anomaly Scoring	Mahalanobis Distance Weight (λ2)	0.3
Anomaly Classification	Threshold (τ)	95th Percentile
Dataset Split	Training Set Percentage	67%
Dataset Split	Validation Set Percentage	33%

As shown in Table 2, the anomaly-enhanced detection approach uses five Gaussian components (K=5) in the GMM as an initial value, which is optimized based on evaluation metrics during the experiments. The density weight ($\lambda 1$) is set to 0.7, and the Mahalanobis distance weight ($\lambda 2$) is set to 0.3. The threshold (τ) for anomaly classification is set to the 95th percentile of anomaly scores. The dataset has been allocated 67% for training purposes and 33% for validation. This configuration of parameters facilitates the accurate identification of both normal and anomalous samples.

4. Experiments

4.1. Datasets

This study employed three prominent network intrusion detection datasets: NSL-KDD, UNSW-NB15, and CICIDS2017, serving as benchmark datasets. The selection of these datasets was based on a comprehensive literature review, which indicated that approximately 75% of recent studies on intrusion detection employed one or more of these datasets [8, 13, 21]. Their widespread use is due to their extensive coverage of attack scenarios, a diverse array of features, and their efficacy in benchmarking anomaly detection models. A summary of the datasets is presented in Table 3.

Table 3
Summary of benchmark datasets

Dataset	Samples	Features	Attack Types
NSL-KDD	125,973	41	DoS, Probe, U2R, R2L
UNSW-NB15	257,673	49	Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode
CICIDS2017	2,830,743	78	Brute Force, Heartbleed, Botnet, DoS, DDoS, Infiltration, Web Attacks

These datasets encompass a mix of normal and anomalous traffic, covering modern and traditional attack types. NSL-KDD focuses on traditional attacks, UNSW-NB15 introduces hybrid traffic scenarios, and CICIDS2017 provides a realistic representation of modern network attacks.

The datasets were standardized using the normalization process described in Section 3.2.1, ensuring all features were scaled to a uniform range to improve model performance. Feature selection were performed following the four-step process outlined in Section 3.2.2.

Correlation analysis was performed to identify and remove features with high collinearity (correlation coefficient > 0.9), reducing redundancy in the dataset. Table 4 presents the highly correlated features in each dataset.

Table 4
Highly correlated features identified in datasets.

Dataset	Feature Pairs with Correlation (> 90%)
NSL-KDD	dst_bytes & src_bytes, srv_serror_rate & serror_rate , srv_rerror_rate & rerror_rate
UNSW-NB15	sttl & dttl , ct_dst_sport_ltm & ct_dst_src_ltm , ct_src_dport_ltm & ct_src_src_ltm
CICIDS2017	Flow Bytes/s & Flow Packets/s , Total Length of Fwd Packets & Total Fwd Packet Length , Fwd IAT Mean & Fwd IAT Std

Then a Random Forest (RF)-based importance scoring method [15] was applied to rank features based on their contribution to classification performance. The parameters used in the RF model for importance scoring are listed in Table 5.

Table 5
RF-based feature importance scoring parameter settings.

Parameter	Value	Description
Number of Trees	100	Number of decision trees in the forest
Max Depth	None	Unlimited tree depth
Min Samples Split	2	Minimum samples required to split an internal node
Min Samples Leaf	1	Minimum samples required to be at a leaf node
Max Features	sqrt	Number of features to consider when looking for the best split
Bootstrap	True	Whether bootstrap sampling is used when building trees

In the results, features with scores below the dynamic threshold were excluded as summarized in Table 6.

Table 6
Features excluded by dynamic thresholds.

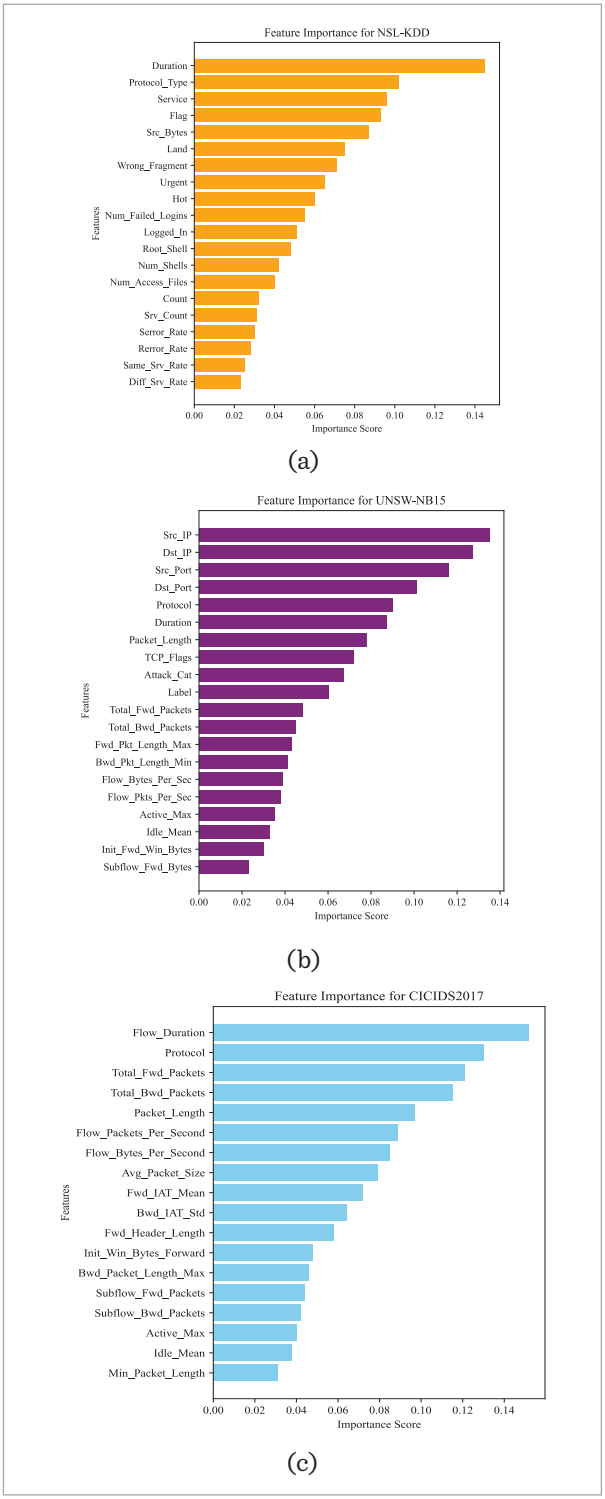
Datasets	Threshold	Removed Features
NSL-KDD	0.018	7
UNSW-NB15	0.03	9
CICIDS2017	0.04	14

The feature importance scores for the three datasets are shown in Figure 3, These charts illustrate the relative importance of selected features, sorted in descending order based on their contribution to classification performance.

Table 7
Final selected features

Dataset	Feature Count	Selected Features
NSL-KDD	24	service, flag, src_bytes, dst_bytes, land, urgent, hot, logged_in, root_shell, is_host_login, serror_rate, srv_serror_rate, rerror_rate, same_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_srv_serror_rate, dst_host_serror_rate, num_failed_logins, num_root
UNSW-NB15	27	srcip, sport, dstip, dsport, proto, state, is_ftp_login, res_bdy_len, ct_dst_src_ltm, ct_dst_sport_ltm, Sload, Dload, is_sm_ips_ports, dttl, ct_src_ltm, ct_src_dport_ltm, dur, ct_ftp_cmd, ct_srv_dst, Dintpkt, Ltime, Sintpkt, synack, ct_srv_src, ct_dst_ltm, Djit, Stime
CICIDS 2017	33	flow duration, destination port, total fwd packets, total backward packets, ece flag count, flow packets/s, flow bytes/s, average packet size, fwd iat mean, bwd iat std, fwd header length, init_win_bytes_forward, bwd packet length max, bwd packet length mean, subflow fwd packets, subflow bwd bytes, active max, active std, min packet length, fwd iat std, bwd iat total, bwd header length, active min, fwd packets/s, total length of fwd packets, bwd packet length std, total length of bwd packets, flow iat mean, cwe flag count, flow iat std, fin flag count, active mean, idle mean

Figure 3
Feature importance for benchmark datasets: (a) Feature importance for NSL-kDD; (b) Feature importance for UNSW-NB15; (c)Feature importance for CICIDS2017.



Subsequently, the feature distribution analysis across datasets method was applied to ensure consistency of selected features across NSL-KDD, UNSW-NB15, and CICIDS2017. This process involved comparing the statistical metrics (mean, variance, maximum, and minimum) of each feature across datasets. Features with significant variance instability (variance ratio > 5) or mean deviation (mean offset > 30%) were deemed inconsistent and removed. Following this analysis, NSL-KDD had 4 features removed due to distribution differences, UNSW-NB15 also had 4 such features removed. In CICIDS2017, 10 features were excluded, such as Init_Win_Bytes_Backward and Bwd_Header_Length, due to significant cross-dataset variability.

The remaining features were further refined using domain knowledge. Features unrelated to network security tasks were removed, including those with no direct correlation to attack behavior. Features characterized by high dynamic variability, marked by rapid or unpredictable value changes, were also excluded. Statistical features (e.g., Packet_Length, Flow_Bytes_Per_Second) and attack-related features (e.g., Protocol_Type, Service) were retained to ensure stability and relevance. The final selected features for all datasets are summarized in Table 7.

4.2. Experimental Setup

4.2.1. Baseline Methods

To evaluate the performance of SAEDF, we compared it with four baseline methods, including traditional detection techniques and advanced generative-based approaches. These baselines were selected to cover a range of methodologies, including unsupervised detection, supervised learning, and generative models for data augmentation. Specifically, we included isolation forest (IF) [23] and local outlier factor (LOF) [24] as unsupervised anomaly detection methods, SGAN-IDS [2] as a GAN-based generative model for anomaly detection. These methods were chosen based on their relevance to the scope of this study, as they either represent state-of-the-art generative approaches or provide benchmark techniques commonly used for intrusion detection in the existing literature. Furthermore, they are well-suited for the datasets used in this study, ensuring a fair and meaningful comparison.

In order to assess the performance of SAEDF, we conducted a comparative analysis against four foundational methodologies, which encompassed both traditional detection techniques and sophisticated generative-based strategies. These baselines methods were selected to encompass a spectrum of methodologies, including unsupervised detection, supervised learning, and generative models utilized for data augmentation. Specifically, our selection included the isolation forest (IF) [23] and local outlier factor (LOF) [24] as unsupervised anomaly detection techniques, SGAN-IDS [2] as a GAN-based generative model for anomaly detection, and a hybrid supervised two-stage detection approach that combines LightGBM and an autoencoder [42]. These methodologies were chosen based on their robust performance documented in existing literature and their suitability for the datasets employed in this research.

4.2.2. Implementation Details

This section delineates the structural specifics of the ADGVAE model (as detailed in Section 3.3.3), encompassing its network architecture, primary components, hyperparameters, and training configurations. The model integrates three fundamental components: a projection layer (d_{proj}), a dynamic hidden layer structure, and a mechanism for dynamic layer counting. A summary of the parameters defining the network architecture of ADGVAE for each dataset is provided in Table 8.

Table 8
Network architecture for each dataset.

Dataset	d_{in}	d_{proj}	Hid-Layers	Nodes	w_{dataset}
NSL-KDD	28	16	6	[64, 128, 256, 256, 128, 64]	1.5
UN-SW-NB15	34	20	8	[128, 256, 512, 512, 256, 128, 64, 32]	2.0
CIC-IDS2017	50	32	10	[256, 512, 1024, 1024, 512, 256, 128, 64, 32, 16]	2.5

For each dataset, was determined experimentally based on the dataset size and feature complexity, balancing the trade-off between detection performance and computational efficiency.

The training process was configured to ensure stable convergence and effective anomaly detection. The key hyperparameters are shown in Table 9.

Table 9
Hyperparameters for training ADGVAE.

Parameters	NSL-KDD	UNSW-NB15	CICIDS2017
Learning rate	0.001	0.0005	0.0001
Batch size	32	64	128
Epochs	100	150	200
Latent space	8	16	32
Dropout rate	0.2	0.3	0.3
Regularization	0.01	0.005	0.001
Optimizer	Adam	AdamW	AdamW
Learning rate decay	0.95	0.9	0.85

5. Results and Discussion

5.1 Evaluation metrics

The evaluation metrics we designed in this study focus on both the quality of the generated anomalous samples and the detection performance of the framework. The Fréchet Inception Distance (FID) was employed to assess the quality of the generated anomalous samples, representing the similarity between the distributions of the generated samples and the original samples through the distance between the two distributions.

The FID is defined as:

$$\mathrm{FID} = \|\mu_r - \mu_g\|^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}\right), \quad (16)$$

where μ_r and μ_g are the mean feature representations of real and generated anomalous samples, respectively, and Σ_r and Σ_g are their covariance matrices. A lower FID indicates higher similarity, reflecting better fidelity and diversity in the generated samples.

The detection performance of the framework was evaluated using metrics derived from the confusion matrix [30]. Specifically, the metrics included Precision, which measures the proportion of correctly identified anomalies among predicted anomalies; Recall, which reflects the model's ability to detect all true anomalies; F1-Score, the harmonic mean of Precision and Recall; and AUC-ROC, which assesses the model's ability to distinguish between normal and anomalous samples.

5.2. Results

5.2.1. Attack Sample Generation Quality

The FID values in Table 10 summarize the quality of the generated samples for each dataset. On average, CICIDS2017 achieved the lowest mean FID (8.96), indicating the highest similarity between the generated and real samples, followed by UNSW-NB15 (10.78) and NSL-KDD (12.34). Across all datasets, the maximum FID values remain low (<19), reflecting consistent fidelity and diversity in the generated samples. The standard deviation of FID values is also small, suggesting stable performance across features.

The observed low FID values in this study reflects the high fidelity and diversity of the generated samples. The CICIDS2017 dataset achieved the lowest mean FID value (8.96), demonstrating that the ADGVAE model effectively captures the underlying distribution of complex modern network attacks. This allows the model to learn more realistic decision boundaries and improves its ability to detect previously unseen anomalies.

The small standard deviation (STD) of FID values across datasets, such as 2.47 for CICIDS2017, emphasizes the stability of the ADGVAE model in generating consistent data distributions. This consistency minimizes fluctuations in the training process, guaranteeing that the model is less susceptible to biases resulting from inconsistent or low-quality synthetic samples.

Table 10
FID results for generated samples.

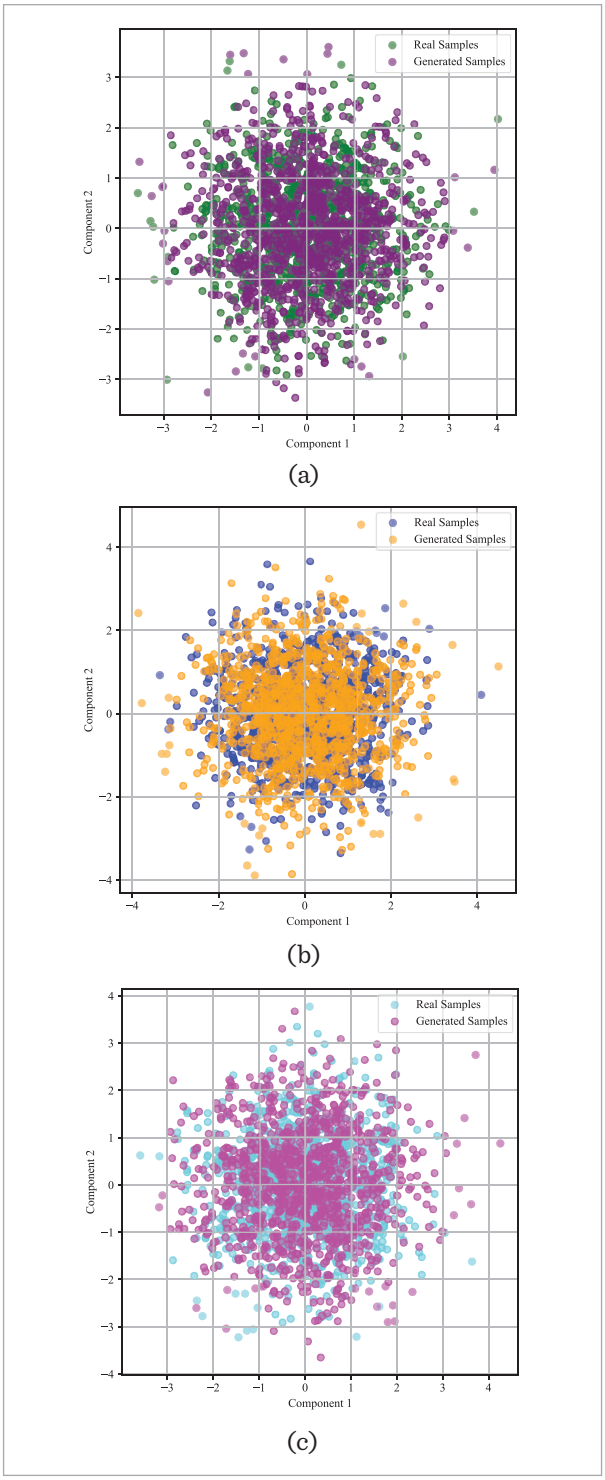
Dataset	Mean FID	Max FID	Min FID	STD
NSL-KDD	12.34	18.56	6.78	3.21
UNSW-NB15	10.78	15.89	5.43	2.95
CICIDS2017	8.96	14.32	4.87	2.47

In order to visually demonstrate the caliber of the produced anomalous samples, we performed a comparative assessment of their distributions relative to genuine anomalies through the utilization of scatter plots. Figure 4 depicts the distribution of both authentic and generated samples for the benchmark datasets.

From Figure 4, the horizontal (Component 1) and vertical (Component 2) axes represent the reduced dimensions obtained through PCA, preserving the

Figure 4

Synthetic sample distribution for benchmark datasets: (a) sample distribution for NSL-KDD; (b) sample distribution for UNSW-NB15; (c) sample distribution for CICIDS2017.



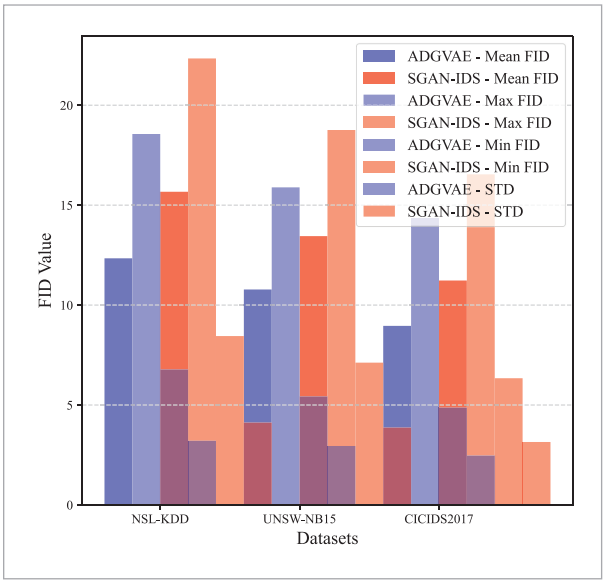
primary relationships in the feature space. The generated samples closely align with the real anomalies in feature space, reflecting the high fidelity and diversity of our ADGVAE model.

We conducted a comparative analysis of the FID metrics pertaining to ADGVAE alongside those of SGAN-IDS for the samples produced across the benchmark datasets.

Figure 5 presents a comprehensive comparison of FID metrics (Mean, Max, Min, and STD) between ADGVAE and SGAN-IDS. ADGVAE consistently outperforms SGAN-IDS in all metrics, with significantly lower Mean FID values, reflecting higher fidelity in generated samples. Additionally, the lower Max FID and STD values for ADGVAE indicate better distribution consistency and reduced variance, highlighting its superiority in generating high-quality anomaly samples. These results demonstrate that ADGVAE more effectively captures the underlying feature distributions of real anomalies compared to SGAN-IDS.

Figure 5

Comparison of FID metrics across datasets.



5.2.2. Detection Performance for Unknown Attacks

We summarize the detection performance results for unknown attacks in Table 11, comparing SAEDF with baseline methods across three datasets.

Table 11
Best performance for unknown attacks detection.

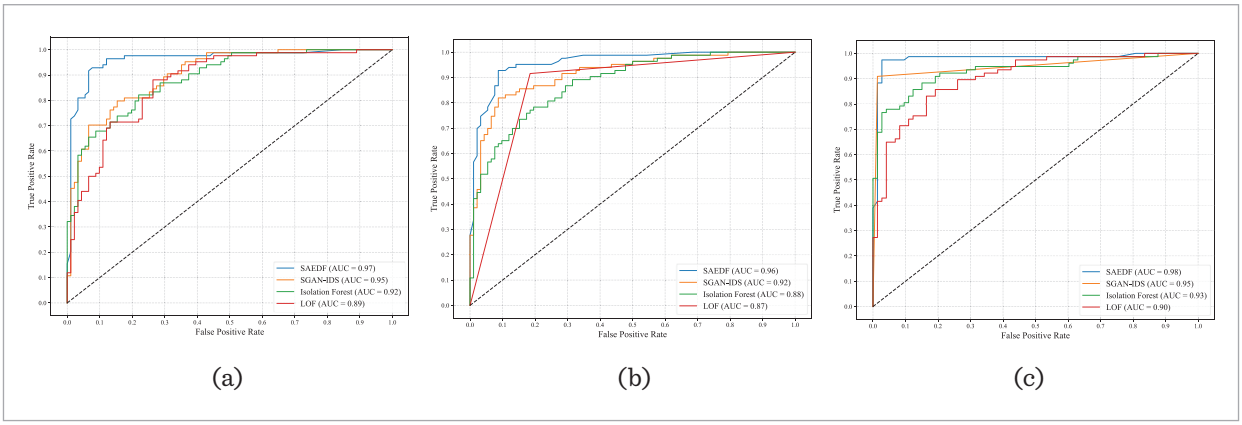
Dataset	Model	F1	Acc.	Prec.	Rec.	FPR
NSL-KDD	SAEDF	0.971	0.975	0.965	0.978	0.003
	SGAN-IDS	0.943	0.952	0.931	0.955	0.007
	IF	0.891	0.916	0.902	0.882	0.015
	LOF	0.872	0.894	0.880	0.865	0.020
UNSW-NB15	SAEDF	0.945	0.962	0.940	0.951	0.004
	SGAN-IDS	0.912	0.938	0.900	0.924	0.012
	IF	0.861	0.887	0.870	0.853	0.018
	LOF	0.848	0.873	0.860	0.837	0.022
CICIDS 2017	SAEDF	0.984	0.988	0.981	0.987	0.002
	SGAN-IDS	0.957	0.965	0.950	0.964	0.006
	IF	0.902	0.925	0.910	0.895	0.012
	LOF	0.889	0.910	0.895	0.882	0.020

For the NSL-KDD dataset, SAEDF achieves the most significant improvement in Recall, outperforming LOF by 13.02% and demonstrating its ability to detect anomalies effectively. On UNSW-NB15, SAEDF shows the largest increase in F1-Score, with a 11.45% improvement compared to LOF. For CICIDS2017, SAEDF achieves the highest improvement in Recall, surpassing LOF by 11.05%. From Figure 6, we can observe that SAEDF's ROC curves consistently stay closer to the top-left corner compared to other models, reflecting its superior ability to achieve higher True Positive Rates at lower False Positive Rates. These results highlight SAEDF's unmatched ability to detect unknown attacks, particularly in handling imbalanced datasets, and emphasize its significant performance gains over baseline models.

5.2.3. Cross-Dataset Unknown Attack Detection

To further validate the effectiveness of SAEDF in detecting unknown attacks, particularly zero-day attacks, we designed a cross-dataset unknown attack detection experiment. This experiment evaluates the generalization capability of SAEDF by training the model on one dataset and testing it on entirely

Figure 6
ROC curves of benchmark datasets: (a) NSL-KDD; (b) UNSW-NB15; (c) CICIDS2017.



(a) (b) (c)

different datasets. This setup mimics real-world scenarios where the model encounters previously unseen attack patterns from diverse environments. The model was trained on the NSL-KDD dataset using all normal samples and a subset of known attack samples, and it was tested on two separate datasets (UNSW-NB15 and CICIDS2017) containing unknown attack types and normal traffic, with the results shown in Table 12.

Table 12
Cross-dataset detection performance for unknown attacks.

Dataset	Model	F1	Prec.	Rec.	FPR
UNSW-NB15	SAEDF	0.922	0.915	0.930	0.005
	SGAN-IDS	0.876	0.860	0.893	0.013
	IF	0.831	0.845	0.818	0.018
	LOF	0.812	0.820	0.805	0.022
CICIDS2017	SAEDF	0.937	0.930	0.945	0.004
	SGAN-IDS	0.892	0.880	0.905	0.011
	IF	0.850	0.865	0.837	0.016
	LOF	0.828	0.835	0.820	0.019

From Table 12, It is observed that SAEDF achieves significantly higher F1-Score and Recall compared to baseline methods, demonstrating its ability to generalize to unknown attack patterns across datasets. On the UNSW-NB15 dataset, SAEDF outperforms SGAN-IDS by 4.4% in F1-Score and 3.5% in Recall, while on the CICIDS2017 dataset, it achieves an F1-Score of 0.938, which is 4.2% higher than SGAN-IDS. Additionally, the False Positive Rate (FPR) of SAEDF remains remarkably low across both datasets, with values of 0.005 (UNSW-NB15) and 0.004 (CICIDS2017), indicating its robustness in minimizing false alarms.

5.2.4. Ablation Study

We conducted an ablation study to understand the importance of each component in our model. Table 13 shows the performance impact, measured as the average F1-score and training time across three datasets, when specific components are removed or replaced.

As shown in Table 13, ADGVAE provides superior performance while incurring only a minor decrease in training time (approximately 8% shorter than the simpler autoencoder).

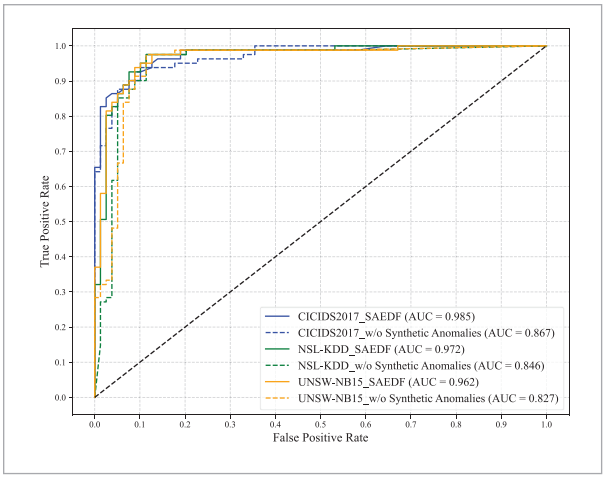
Table 13
Performance decrease (%) in ablation study.

	Synthetic Anomalies	Feature Representation	Network Architecture
Model Performance	w/o -12.5%	w/o -18.7%	w/o ADGVAE -19.3%
Training Time	——	——	w/o ADGVAE -8 %

The study focuses on the following components:

Synthetic Anomaly Sampling: We analyzed the impact of removing synthetic anomaly samples from the training process. Without these samples, the model relies solely on real data, which limits its ability to generalize to unknown attacks. Synthetic anomalies expand the decision boundary by introducing diverse and augmented anomalous patterns that real-world data alone cannot provide. These patterns allow the model to better capture the variability of unknown attacks, avoiding overfitting to seen data and improving its generalization performance. Figure 7 illustrates this comparison, showing the ROC curves for benchmark datasets. In each dataset, the performance of the model trained with synthetic anomalies (SAEDF) significantly outperforms the one trained without synthetic anomalies (w/o Synthetic Anomalies), as evidenced by the higher AUC values and ROC curves closer to the top-left corner.

Figure 7
ROC curves for ablation.



Feature Representation: To evaluate the importance of feature representation, we replaced the feature transformation module with raw input features. This tests the contribution of representation learning to anomaly detection, and the training time increased by 8% when using raw features.

ADGVAE Model: We substituted the ADGVAE model with a simpler autoencoder (a unified 3-layer FC mirrored structure) to evaluate its role in generating high-quality synthetic anomalies and learning robust latent representations. In addition to performance degradation, we also compared the training and inference times of the two models to evaluate computational efficiency.

5.3 Analysis and Discussion

5.3.1. SAEDF's Capability to Detect Zero-Day Attacks

The above experiments conclusively show that SAEDF significantly outperforms baseline methods, primarily due to its ability to detect unknown attack types, improved generalization through synthetic anomaly generation, and scalability to high-dimensional network traffic data. By generating synthetic anomalies that extend beyond the normal distribution, SAEDF effectively simulates potential unknown attack patterns, enabling the model to generalize to previously unseen threats. This capability is particularly critical for detecting zero-day attacks, which are inherently unknown during training.

1 **Synthetic Anomaly Generation for Generalization:** The use of synthetic anomalies is a critical component of SAEDF′s success, as it enables the model to effectively learn decision boundaries that distinguish normal behavior from diverse and unseen attack patterns. The synthetic anomaly generation process leverages the latent space of the VAE to model the normal data distribution effectively. By sampling regions of low density in the latent space, SAEDF generates realistic synthetic anomalies that mimic unknown attack patterns, enhancing the model's capacity to detect zero-day attacks. This process exposes the model to a broader range of abnormal scenarios during training, significantly improving its generalization ability and reducing overfitting to known attack types.

2 **Density-Based Modeling with GMM:** The classification process further strengthens SAEDF′s ability to detect unknown attacks through the use of GMM-based density modeling. The GMM identifies low-density regions in the latent space, corresponding to potential unknown attack patterns. This probabilistic approach ensures that anomalies, including synthetic ones, are effectively separated from normal samples, enabling robust detection of unknown threats. By modeling the data distribution as a mixture of Gaussian components, the GMM isolates high-density (normal) regions while flagging low-density (abnormal) samples. This not only aids in detecting synthetic anomalies but also provides a robust mechanism for identifying real-world unknown attack patterns.

3 **Scalability and Adaptability:** Another key strength of SAEDF lies in its adaptability and scalability to various datasets with diverse feature complexities. The ADGVAE model incorporates a projection layer (d_{proj}), a dynamic hidden layer structure, and a dynamic layer count mechanism, which together ensure consistency in input representation and adaptability to different dataset characteristics. These features are essential for detecting unknown attacks across varying network environments, as they allow the model to maintain high performance even when dealing with high-dimensional and heterogeneous network traffic data. The dynamic layer count mechanism enables the model to adjust its depth based on the complexity of the input data, ensuring that sufficient representational capacity is allocated for intricate patterns while avoiding overfitting on simpler data. This flexibility is critical for addressing the diverse and evolving nature of network traffic anomalies.

5.3.2. Limitations

1 **Framework Limitations:** While SAEDF demonstrates strong performance, it is not without limitations. Its reliance on high-quality, anomaly-free training data is crucial. If the training data contains contamination or mislabeled samples, the model's ability to accurately detect true anomalies can be significantly hindered. This limitation highlights the importance of robust data preprocessing and careful curation of training datasets to ensure that they are free from noise or anomalies.

The effectiveness of synthetic anomaly generation depends heavily on the quality and diversity

of the generated samples. Poorly generated anomalies may negatively impact decision boundaries and generalization, potentially reducing the model's ability to detect unknown attacks.

2 **The Challenge of True Zero-Day Attacks:**

It should be noted that these simulated scenarios, while effective in controlled experiments, may not fully reflect the complexity and diversity of real-world zero-day attacks. Real-world zero-day attacks often involve highly sophisticated techniques, rapid evolution, and adaptive adversarial behaviors that are challenging to replicate in experimental settings· These limitations provide opportunities for future research. Expanding the evaluation to include real-world datasets and live network traffic, as well as incorporating adaptive adversarial testing, will be essential to further demonstrate the robustness and practicality of SAEDF in real-world zero-day scenarios.

6. Conclusion

In this study, we proposed a flexible and robust framework for detecting unknown attacks in network traffic data. Through extensive experiments, we demonstrated the effectiveness of SAEDF in addressing the challenge of unknown attack detection, highlighting the critical role of synthetic anomalies in enhancing detection performance and improving generalization. By leveraging synthetic anomaly generation, SAEDF establishes a more comprehensive decision boundary and achieves scalability to high-dimensional network traffic data, outperforming baseline methods across multiple datasets.

As future work, we aim to explore more deep learning models and distributed learning within a single ADGVAE framework to simplify the architecture and enhance learning efficiency. Adapting SAEDF for real-time detection in dynamic network environments, such as those found in IoT, cloud computing, and 5G networks, is a key future direction. Further investigation will involve validating SAEDF on larger datasets and diverse network settings to assess its real-world robustness and scalability.

Acknowledgement

This research was supported by the Key Research and Development Project of Henan Province (Big Data and Artificial Intelligence-Based Decision Support Platform, Grant No. 251111211800) and by the Henan Higher Education Teaching Reform Research and Practice Project (Inclusion Research, Grant No. 2024SJGLX0188).

References

1. Alashhab, A. A., ZaFhid, M. S., Isyaku, B., Elnour, A. A., Nagmeldin, W., Abdelmaboud, A., Abdullah, T. A. A., Maiwada, U. D. Enhancing DDoS Attack Detection and Mitigation in SDN Using an Ensemble Online Machine Learning Model. IEEE Access 2024, 12, 51630-51649. https://doi.org/10.1109/ACCESS.2024.3384398
2. Aldhaheri, S., Alhuzali, A. SGAN-IDS: Self-Attention-Based Generative Adversarial Network Against Intrusion Detection Systems. Sensors, 2023, 23(18), 7796. https://doi.org/10.3390/s23187796
3. Bagui, S., Li, K. Resampling Imbalanced Data for Network Intrusion Detection Datasets. Journal of Big Data, 2021, 8(1), 1-22. https://doi.org/10.1186/s40537-020-00390-x
4. Ban, Y., Zhang, D., He, Q., Shen, Q. APSO-CNN-SE: An Adaptive Convolutional Neural Network Approach for IoT Intrusion Detection. Cmc-Computers Materials & Continua, 2024, 81(1), 567-601. https://doi.org/10.32604/cmc.2024.055007
5. Bin, Y., Chen, Y., Ren, Q., Zhang, R., Smith, P. J., Wang, X., Ma, A., Gao, H. SCGMAI: A Gaussian Mixture Model for Clustering Single-Cell RNA-Seq Data Based on Deep Autoencoder. Briefings in Bioinformatics,2020, 21(3), 123-135. https://doi.org/10.1093/bib/bbaa316
6. Das, S., Ashrafuzzaman, M., Sheldon, F. T., Shiva, S. Ensembling Supervised and Unsupervised Machine Learning Algorithms for Detecting Distributed Denial of Service Attacks. Algorithms, 2024, 17(3), 99. https://doi.org/10.3390/a17030099
7. Dongqing, W., Shuo, Z., Min, G., Jianlong, Q. A Novel EM Identification Method for Hammerstein Systems with Missing Output Data. IEEE Transactions on Industrial Informatics, 2020, 16(4), 2500-2508. https://doi.org/10.1109/TII.2019.2931792

8. Di Mattia, F., Galeone, P., Simoni, M. D., Ghelfi, E. A Survey on GANs for Anomaly Detection. arXiv:1906.11632, 2019. https://doi.org/10.48550/arXiv.1906.11632
9. Dunmore, A., Jang-Jaccard, J., Sabrina, F., Kwak, J. A Comprehensive Survey of Generative Adversarial Networks (GANs) in Cybersecurity Intrusion Detection. IEEE Access, 2023, 11, 76071-76094. https://doi.org/10.1109/ACCESS.2023.3296707
10. Elad, A., Rami, B., Daniel, R., Alex, B. Noise Estimation Using Density Estimation for Self-Supervised Multimodal Learning. Proceedings of the AAAI Conference on Artificial Intelligence, (AAAI 2021), Virtual Conference, February 2-9, 2021, 35 (8), 6644-6652. https://doi.org/10.1609/aaai.v35i8.16822
11. Falowo, O. I., Ozer, M., Li, C., Abdo, J. B. Evolving Malware and DDoS Attacks: Decadal Longitudinal Study. IEEE Access, 2024, 12, 39221-39237. https://doi.org/10.1109/ACCESS.2024.3376682
12. Fathima, A. N., Ibrahim, S. S., Khraisat, A. Enhancing Network Traffic Anomaly Detection: Leveraging Temporal Correlation Index in a Hybrid Framework. IEEE Access, 2024, 12, 136805-136824. https://doi.org/10.1109/ACCESS.2024.3458903
13. Fernandes, G., Rodrigues, J. J., Carvalho, L. F., Al-Muhtadi, J. F., Proença, M. L. A Comprehensive Survey on Network Anomaly Detection. Telecommunication Systems, 2019, 70, 447-489. https://doi.org/10.1007/s11235-018-0475-8
14. Feiniu, Y., Lin, Z., Xue, X., Qinghua, H., Xuelong, L. A Wave-Shaped Deep Neural Network for Smoke Density Estimation. IEEE Transactions on Image Processing, 2020, 29, 2301-2313. https://doi.org/10.1109/TIP.2019.2946126
15. Gang, Y., Xiaojian, H., Taiyun, Z., Yue, Z. Enterprise Credit Risk Prediction Using Supply Chain Information: A Decision Tree Ensemble Model Based on the Differential Sampling Rate, Synthetic Minority Oversampling Technique and AdaBoost. Expert Systems, 2022, 39(6), 1-15. https://doi.org/10.1111/exsy.12953
16. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y. Generative Adversarial Networks. Communications of the ACM,2020, 63(11), 139-144. https://doi.org/10.1145/3422622
17. Gang, K., Yong, X., Yi, P., Feng, S., Yang, C., Kun, C., Shaomin, K. Bankruptcy Prediction for SMEs Using Transactional Data and Two-Stage Multiobjective Feature Selection. Decision Support Systems, 2021, 140, 113429. https://doi.org/10.1016/j.dss.2020.113429
18. Hongwei, Y., Jianqiang, H., Deyu, M., Xian-Sheng, H., Lei, Z. Momentum Batch Normalization for Deep Learning with Small Batch Size. Proceedings of the European Conference on Computer Vision (ECCV 2020), Glasgow, UK, August 23-28, 2020, 224-240. https://doi.org/10.1007/978-3-030-58610-2_14
19. Kim, J. Y., Bu, S. J., Cho, S. B. Zero-Day Malware Detection Using Transferred Generative Adversarial Networks Based on Deep Autoencoders. Information Sciences, 2018, 460, 83-102. https://doi.org/10.1016/j.ins.2018.04.092
20. Kingma, D. P., Welling, M. An Introduction to Variational Autoencoders. Foundations and Trends in Machine Learning, 2019, 12(4), 307-392. https://doi.org/10.1561/2200000056
21. Kwon, D., Kim, H., Kim, J., Suh, S. C., Kim, I., Kim, K. J. A Survey of Deep Learning-Based Network Anomaly Detection. Cluster Computing, 2017, 22, 949-961. https://doi.org/10.1007/s10586-017-1117-8
22. Liu, C. Design and Implementation of Computer Network Security Protection System Based on Artificial Intelligence Technology. Applied Mathematics and Nonlinear Sciences, 2023, 8 (2), 1491-1508. https://doi.org/10.2478/amns.2023.1.00049
23. Liang, Z., Lingyun, L. Data Anomaly Detection Based on Isolation Forest Algorithm. Proceedings of the 2022 International Conference on Computation, Big-Data and Engineering (ICCBE 2022), Yunlin, Taiwan, May 27-29, 2022, 87-89. https://doi.org/10.1109/ICCBE56101.2022.9888169
24. Mahmood, S., Abul Samad, I., Hassan, C., Maha, D., Wadii, B., Shahla, A., Mitra, S. Standalone Noise and Anomaly Detection in Wireless Sensor Networks: A Novel Time-Series and Adaptive Bayesian-network-based Approach. Software - Practice and Experience, 2020, 50(4), 428-446. https://doi.org/10.1002/spe.2785
25. Moustafa, N., Slay, J. UNSW-NB15: a Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set). Proceedings of the Military Communications and Information Systems Conference, (MilCIS 2015), Canberra, ACT, Australia, November 10-12, 2015, 1-6. https://doi.org/10.1109/MilCIS.2015.7348942
26. Mulyanto, M., Faisal, M., Prakosa, S. W., Leu, J. S. Effectiveness of Focal Loss for Minority Classification in Network Intrusion Detection Systems. Symmetry 2020, 13(1), 1-13. https://doi.org/10.3390/sym13010004
27. Melika, A., Hamid, M. Stable Feature Selection Based on Probability Estimation in Gene Expression Datasets.

Expert Systems with Applications, 2024, 248, 123372. https://doi.org/10.1016/j.eswa.2024.123372

28. Peppes, N., Alexakis, T., Adamopoulou, E., Demestichas, K. The Effectiveness of Zero-Day Attacks Data Samples Generated via GANs on Deep Learning Classifiers. Sensors, 2023, 23(2), 900. https://doi.org/10.3390/s23020900
29. Pinto, A., Herrera, L.-C., Donoso, Y., Gutierrez, J. A. Enhancing Critical Infrastructure Security: Unsupervised Learning Approaches for Anomaly Detection. International Journal of Computational Intelligence Systems, 2024, 17(1), 236. https://doi.org/10.1007/s44196-024-00644-z
30. Piya, L., Nirattaya, K., Cholwich, N. Gait Recognition and Re-Identification Based on Regional LSTM for 2-Second Walks. IEEE Access, 2021, 9, 112057-112068. https://doi.org/10.1109/ACCESS.2021.3102936
31. Prabakaran, M. K., Meenakshi Sundaram, P., Chandrasekar, A. D. An Enhanced Deep Learning-Based Phishing Detection Mechanism to Effectively Identify Malicious URLs Using Variational Autoencoders. IET Information Security, 2023, 17 (3), 423-440. https://doi.org/10.1049/ise2.12106
32. Phanindra Reddy, K., Noorullah Shariff, C., Rajkumar Laxmikanth, B. An Anomaly-Based Intrusion Detection System Using Recursive Feature Elimination Technique for Improved Attack Detection. Theoretical Computer Science, 2022, 931, 56-64. https://doi.org/10.1016/j.tcs.2022.07.030
33. Sun, X., Dai, J., Liu, P., Singhal, A., Yen, J. Using Bayesian Networks for Probabilistic Identification of Zero-Day Attack Paths. IEEE Transactions on Information Forensics and Security, 2018, 13(10), 2506-2521. https://doi.org/10.1109/TIFS.2018.2821095
34. Tavallaee, M., Bagheri, E., Lu, W., Ghorbani, A. A. A Detailed Analysis of the KDD CUP 99 Data Set. Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA 2009), Ottawa, ON, Canada, July 8-10, 2009, 1-6. https://doi.org/10.1109/CISDA.2009.5356528
35. Truong-Huu, T., Dheenadhayalan, N., Kundu, P. P., Ramnath, V., Liao, J., Teo, S. G., Kadiyala, S. P. An Empirical Study on Unsupervised Network Anomaly Detection Using Generative Adversarial Networks. Proceedings of the 1st ACM Workshop on Security and Privacy on Artificial Intelligence, (SPAI 2020), Taipei, Taiwan, October 6, 2020, 20-29. https://doi.org/10.1145/3385003.3410924
36. Truong, V. T., Le, L. B. MetaCIDS: Privacy-Preserving Collaborative Intrusion Detection for Metaverse Based on Blockchain and Online Federated Learning. IEEE Open Journal of the Computer Society, 2023, 4, 253-266. https://doi.org/10.1109/GCWkshps58843.2023.10464435
37. Wang, H., Yan, J., Jia, N. A New Encrypted Traffic Identification Model Based on VAE-LSTM-DRN. Computational Materials and Continua, 2024, 78 (1). https://doi.org/10.32604/cmc.2023.046055
38. Xu, L., Skoularidou, M., Cuesta-Infante, A., Veeramachaneni, K. Modeling Tabular Data Using Conditional GAN. Advances in Neural Information Processing Systems, (NeurIPS 2019), Vancouver, BC, Canada, December 8-14, 2019, 32. https://doi.org/10.48550/arXiv.1907.00503
39. Yang, Y., Zheng, K., Wu, C., Yang, Y. Improving the Classification Effectiveness of Intrusion Detection by Using Improved Conditional Variational Autoencoder and Deep Neural Network. Sensors, 2019, 19(11), 2528. https://doi.org/10.3390/s19112528
40. Zavrak, S., Iskefiyeli, M. Anomaly-Based Intrusion Detection from Network Flow Features Using Variational Autoencoder. IEEE Access, 2020, 8, 108346-108358. https://doi.org/10.1109/ACCESS.2020.3001350
41. Zenati, H., Foo, C.-S., Lecouat, B., Manek, G., Chandrasekhar, V. Efficient GAN-Based Anomaly Detection. arXiv:1802.06222, 2018. https://doi.org/10.48550/arXiv.1802.06222
42. Zhang, H., Ge, L., Zhang, G., Fan, J., Li, D., Xu, C. A Two-Stage Intrusion Detection Method Based on Light Gradient Boosting Machine and Autoencoder. Mathematical Biosciences and Engineering, 2023, 20(4), 6966-6992. https://doi.org/10.3934/mbe.2023301