Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 9
AIDML 2025

Application Review of Network Attack Detection Based on Deep Learning

Yixuan Liu[1], *Jiaohang Yu[2]

[1]School of Information Technology and Management, Hunan University of Finance and Economics, Changsha, China

[2]School of Mathematics, Southwest Jiaotong University, Chengdu, China

*Corresponding author: liuyixuan0850@163.com

Abstract. With the rapid development of the Internet, network attacks (NA) have become increasingly diverse and sophisticated. Traditional detection methods relying on rules systems and statistical analysis now face significant challenges, including strong dependency on feature engineering and limited generalization ability. The effectiveness of detection systems is not only limited, but the difficulty of addressing new threats is also increased. Deep learning has been recognized as a novel solution for NA detection due to its powerful nonlinear feature extraction capabilities and end-to-end learning paradigm. A concise overview of deep learning fundamentals is first provided. Subsequently, the paper proceeds with a systematic review of prevailing deep learning approaches for NA detection, followed by an analysis of their current limitations and challenges in NA. Finally, promising future research directions are proposed to address the aforementioned limitations and challenges.

Keywords: Deep learning; Network attack detection; Neural network; Supervised learning.

1. Introduction

With the rapid development of information technology on a global scale, the Internet has not only become the core channel for information exchange but also gradually evolved into the basic platform for economic activities and social services. However, the increasing societal dependence on network infrastructure has elevated cybersecurity threats to a critical concern. In recent years, the increasing frequency of attack incidents, such as personal information leakage, corporate data theft, and attacks on critical infrastructure (such as power grids and transportation systems), has revealed that the network security situation is facing severe challenges. Therefore, network attack detection has been prioritized as a critical concern today.

The concept of intrusion detection was first proposed by James P. Aderson in 1980[1]. An intrusion attempt or threat is defined as a potential, premeditated, and unauthorized access to or manipulation of information. Such actions may result in system unreliability or unavailability. An abstract model of the Intrusion Detection System (IDS) was proposed. Real-time network monitoring is implemented through misuse or anomaly detection methods. Unauthorized or malicious behaviors can be identified, and an effective defense against intrusion can be established.

Traditional defense systems rely on predefined security rule sets and historical data for statistical analysis. Potential security threats are identified and responded to through this process. By combining with machine learning, supervised machine learning methods such as support vector machines[2], k-nearest neighbor(KNN) method[3], decision trees, and unsupervised machine learning methods such as K-means clustering analysis[4]and principal component analysis[5]have been developed. Basic predictions with a certain degree of accuracy are achieved, and real-time monitoring and response requirements in network environments are met through the aforementioned method.

However, with the popularization of emerging technologies such as the Internet of Things, 5G communication, and cloud computing, network attack forms and methods are increasingly complex and diverse. Traditional machine-learning-based defense measures are difficult to deal with these new

threats. With the continuous development of the machine - learning field, deep learning, as a branch of it, has gradually been applied to network attack detection models. Researchers have gradually developed network attack detection models based on deep learning. For example, deep -learning models such as CNN, RNN, Transformer, auto-encoders, and generative adversarial networks have been gradually applied to network attack detection systems. Compared with machine -learning methods, deep learning has higher efficiency in model construction and feature extraction and can improve the accuracy when processing large-scale data.

In this paper, the basics of deep learning are first introduced. Subsequently, several mainstream deep - learning models that are commonly employed in network intrusion detection systems are listed. Each of these models is then analyzed in detail. Following the analysis, the current problems associated with the application of deep-learning in network intrusion detection are discussed. Finally, a forward-looking perspective on the future development in this field is presented.

2. Main Body

2.1. Fundamentals of Deep Learning

As a branch of machine learning, the core of deep learning lies in automatically learning complex feature representations from data by constructing multi-layer neural network models. This process depends on the design of the neural network architecture, the choice of optimization algorithms, and the support of large-scale datasets. In network intrusion detection systems, deep learning shows great potential and can perform well when processing large-scale and high-dimensional datasets. Especially when the data contains rich context information, deep learning can effectively mine valuable information to achieve more accurate and effective defense.

2.2. Main deep learning methods for detecting network attacks

2.2.1. Supervised Learning Methods

1. Convolutional Neural Networks (CNN) were developed from multilayer perceptions and are categorized as deep feedforward neural networks incorporating convolutional operations. Local spatial features are automatically extracted by these networks, enabling suitability for structured or image-based data. Common applications encompass network traffic-to-image conversion, port scanning, DDoS attack detection, and malware identification[6]. Strengths include sensitivity to local patterns and capability in processing high-dimensional data, while limitations involve dependency on data spatial structure assumptions and restricted performance in temporal attack detection. Core CNN Formula: let the input data I (e.g.,a traffic matrix or malware image) be defined with dimensions $H \times W$. The learnable weight matrix W, representing the convolution kernel, is assigned a size of $K_H \cdot K_{\mathrm{w}}$. The stride parameter s (e.g., s =1 or 22) controls the sliding step size, p denotes the padding size, and b is the scalar bias term. The formula can be applied to detect DDoS attacks, where the input I is a traffic matrix, and the kernel W learns high-frequency request patterns[7].

$$O(i,j) = (\sum_{m=0}^{k_H-1} \sum_{n=0}^{k_W-1} W(m,n) \cdot I(i \cdot s + m - p, j \cdot s + n - p)) + b \tag{1}$$

2. Recurrent Neural Networks (RNN) are a type of recursive neural network designed to capture temporal dependencies in sequential data using cyclic structures. Sequence data analysis, such as detection of long-term Advanced Persistent Threat (APT) attacks, represents a primary application domain. Robust modeling of both long- and short-term dependencies stands as a critical capability. However, Challenges such as slow training speeds and sensitivity to sudden noise are encountered. Recursive operations occur along the temporal progression of each sequence. The chain structure of RNN nodes is illustrated in Fig.1.

Unlike traditional neural networks, RNN incorporate self-connected weights in hidden layer neurons. Allowing state information from previous training steps to be passed forward, preserving sequential context and effectively addressing sequence data prediction problems that conventional neural networks struggle to handle[8].

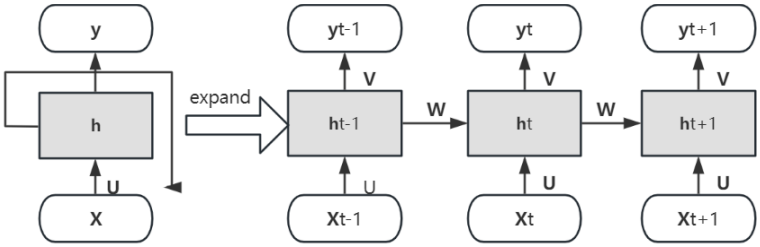

Fig.1 Structural Example of RNN

3. The Transformer architecture leverages the self-attention mechanism to model contextual dependencies across sequences, eliminating the need for recursive operations [9]. It is widely used in scenarios like large-scale network log analysis and encrypted traffic classification. Its strengths include high parallel computing efficiency and effectiveness in modeling long sequences. However, drawbacks such as the requirement for large amounts of labeled data and high computational resources are observed. Self-Attention Formula: The attention weights are calculated as follows: Query (Q), Key (K), and Value (V) matrices are generated by applying linear transformations to the input. Compute the scaled dot product of Q and K: Here, Q is the dimension of K. The SoftMax normalization generates attention weights, and the weighted sum of V reflects correlations between elements.

$$Attention(Q,K,V) = Soft\max(\frac{QK^2}{\sqrt{d_k}})V \tag{2}$$

2.2.2. Unsupervised/Semi-Supervised Learning Methods

1. The normal distribution of data is learned by autoencoders through an encoder-decoder structure, with anomalies identified via reconstruction errors[10]. Primary applications encompass zero-day attack detection (discovering unknown malicious traffic patterns without prior labels) and low-frequency anomaly recognition in industrial control systems (ICS). Advantages include compatibility with limited labeled data, while limitations involve sensitivity to high-variance noise and potential oversight of subtle anomalies. A representative framework is the Variational Autoencoder (VAE)-based IoT anomaly traffic detection system[11].

2. Generative Adversarial Networks (GAN) utilize adversarial training between two components: a generator designed to create synthetic data and a discriminator trained to differentiate real from generated data. Key applications involve data augmentation for mitigating class imbalance and adversarial attack defense by enhancing model robustness through synthetic adversarial sample generation. Despite improvements in model generalization enabled by diverse data synthesis, challenges persist, including unstable training dynamics and difficulties in controlling output fidelity. For instance, rare abnormal user behavior distributions are modeled by the generator, while synthetic samples are assessed by the discriminator. High-quality synthetic samples resembling real anomalies are produced through this adversarial process, with the GAN architecture detailed in Fig.2[12].

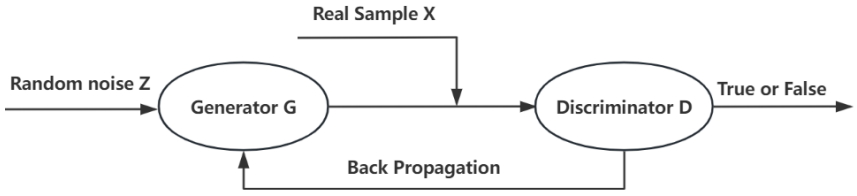

Fig.2 Structural Example of GAN

2.2.3. Hybrid Models and Emerging Methods

1. Graph Neural Networks (GNN) model topological relationships between network entities (e.g., IP addresses, users, and devices). Key applications include phishing detection in social networks and botnet node identification[13]. GNN use three steps as shown in the table 1 to process information integration updates. Strengths include capturing complex correlations in large-scale networks, but challenges include high graph construction costs and difficulty updating dynamic graphs.

Table 1. GNN iteratively perform three steps

The core of GNN is a three-step loop of message passing-aggregation-update		
Message Generation	Neighbor Aggregation	Feature Update
Each node generates messages based on its own and its neighbors' characteristics	Each node generates messages based on its own and its neighbors' characteristics	Update node representation based on its own characteristics and aggregated messages
Formula: $m_{ij} = \phi(h^{(k)}{}_i, h^{(k)}{}_j, e_{ij})$	Formula: $m_i^{(k)} = \oplus_{j \in N(i)} m_{ij}$	Formula: $h_i^{(k+1)} = \varphi(h_i^{(k)}, m_i^{(k)})$
$m_i^{(k)}$: characteristics of node i at the K-th layer	$\oplus$: Aggregation function (such as summation, mean, maximum)	φ : update function (such as GRU, linear transformation +activation function)
e_{ij} : feature from edge i to j		
ϕ : message functions (such as MLP)		

2. Multimodal fusion models integrate multi-source data (e.g., network traffic, system logs, and system calls). Key applications encompass advanced persistent threat (APT) detection[14]and cloud security enhancement through the combination of virtual machine (VM) monitoring data with network traffic analysis. Enhanced detection coverage is achieved via multi-source integration; however, significant challenges arise from data heterogeneity across modalities.

2.2.4. Shortcomings and Challenges

Recent research regarding deep learning-based network intrusion detection systems has revealed significant effectiveness. Deep-learning-based network intrusion detection systems exhibit stronger processing performance in comparison to traditional machine - learning methods. Nevertheless, multiple challenges and limitations persist.

At the data level, high labeling costs and severe class imbalance are critical issues. In real-world scenarios, normal traffic dominates. Rare attack samples are scarce. Manual labeling heavily depends on expert experience, which struggles to address rapidly evolving attacks[15]. The complexity of network environments further reduces data generalization capabilities. Models often experience drastic performance drops in real-world traffic[16].

Complex models such as Transformers and GNN demand substantial computational resources. These models struggle to satisfy the real-time demands of edge devices. Additionally, dynamic multi-phase defense lag persists as a prominent problem. Traditional static models are devoid of adaptive adjustment capabilities in response to environmental changes.

3. Future and Prospects

The rapid development of deep learning-based network intrusion detection systems has enabled widespread application across various domains. However, several technical challenges remain unresolved. A discussion of three aspects-data, models, and applications-will occur, along with an analysis of future development directions.

At the data-driven level, the integration of federated learning with differential privacy shows promises for achieving secure cross-organizational collaboration, addressing both privacy concerns and data

silo issues[17]. Emerging generative technologies like diffusion models can synthesize high-quality attack traffic, alleviating data scarcity challenges for zero-day attacks.

In model architecture, edge deployment of detection systems will be facilitated by lightweight designs and knowledge distillation techniques to meet low-latency requirements[18]. Dynamic attack graphs could be constructed by multimodal GNN that integrate multi-source data including traffic flows, logs, and network topology to enhance detection comprehensiveness.

Regarding security enhancement, manual intervention will be reduced by adaptive adversarial training frameworks and Auto-ML-driven automated pipelines while model self-optimization and rapid iteration will be enabled[19]. Log semantics could be interpreted and human-readable reports could be generated by the incorporation of large language models (LLMs), and explain ability could be improved by being complemented with causal reasoning techniques, and ultimately, human-machine collaborative intelligent defense systems will be formed. For long-term perspectives, deep learning will be deeply integrated with cybersecurity. Self-evolution, quantum-safe security, and collaborative defense against global threats will be evolved towards.

4. Conclusion

Excellent performance can be achieved by applying deep-learning technology to network intrusion detection systems, which benefits from its strong advantages in model construction and feature extraction and its ability to maintain high accuracy when processing large-scale data. The research status of deep-learning-based network intrusion detection systems is described in this paper from aspects such as the description of mainstream methods, existing challenges, and future prospects. This review is expected to provide references and insights for future research in related fields.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Anderson J P. Computer security threat monitoring and surveillance. Technical Report, James P. Anderson Company, 1980.

[2] Al-Qatf M, Lasheng Y, Al-Habib M, et al. Deep learning approach combining sparse autoencod-er with SVM for network intrusion detection. IEEE Access, 2018, 6: 52843-52856.

[3] Ding Hongwei, Chen Leiyang, Dong Liang, et al. Imbalanced data classification: A KNN and generative adversarial networks-based hybrid approach for intrusion detection. Future Generation Computer Systems, 2022, 131: 240-254.

[4] Al-Yaseen W L, Othman Z A, Nazri M Z A. Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Expert Systems with Applications, 2017, 67: 296-303.

[5] Kravchik M, Shabtai A. Efficient Cyber Attack Detection in Industrial Control Systems Using Lightweight Neural Networks and PCA. IEEE Transactions on Dependable and Secure Computing, 2022, 19(4): 2179-2197.

[6] Chua L O, Roska T. The CNN paradigm. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 1993, 40(3): 147-156.

[7] Otter D W, Medina J R, Kalita J K. A survey of the usages of deep learning for natural language processing. IEEE transactions on neural networks and learning systems, 2020, 32(2): 604-624.

[8] Kheddar H. Transformers and large language models for efficient intrusion detection systems: A comprehensive survey. arXiv preprint arXiv:2408.07583, 2024.

[9] Shen Bingbing, Ge Zhiqiang. Supervised nonlinear dynamic system for soft sensor application aided by variational auto - encoder. IEEE Transaction Instrumentation and Measurement, 2020, 69: 6132–6142.

[10] Gouda W, Tahir S, Alanazi S, et al. Unsupervised outlier detection in IOT using deep VAE. Sensors, 2022, 22(17): 6617.

[11] Sen M A. Attention-GAN for Anomaly Detection: A Cutting-Edge Approach to Cybersecurity Threat Management. arXiv preprint arXiv:2402.15945, 2024.

[12] Dutta A, Chatterjee S, Bhattacharya A, et al. Deep reinforcement learning for cyber system defense under dynamic adversarial uncertainties. arXiv preprint arXiv:2302.01595, 2023.

[13] Zhou J, Xu Z, Rush A M, et al. Automating botnet detection with graph neural networks. arXiv preprint arXiv:2003.06344, 2020.

[14] Mutalib N H A, Sabri A Q M, Wahab A W A, et al. Explainable deep learning approach for advanced persistent threats (APTs) detection in cybersecurity: a review. Artificial Intelligence Review, 2024, 57(11): 297.

[15] Ding H, Sun Y, Huang N, et al. TMG - GAN: Generative adversarial networks-based imbalanced learning for network intrusion detection. IEEE Transactions on Information Forensics and Security, 2023, 19: 1156-1167.

[16] Apruzzese G, Andreolini M, Ferretti L, et al. Modeling realistic adversarial attacks against network intrusion detection systems. Digital Threats: Research and Practice (DTRAP), 2022, 3(3): 1-19.

[17] Avishka K P S, Gunasekara G K A, Maduwantha K K, et al. FedSec: Advanced Threat Detection System for Federated Learning Frameworks. 2023 5th International Conference on Advancements in Computing (ICAC). IEEE, 2023: 71-76.

[18] Fusco P, Rimoli GP, Ficco M. TinyIDS-an IoT intrusion detection system by tiny machine learning. Computational Science and ITS Applications-ICCSA 2024 Workshops, PT II, 2024, 14816: 71-82.

[19] Jiang Q, Chen C, Zhao H, et al. Understanding and constructing latent modality structures in multi - modal representation learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7661-7671.