Theory and Practice of Science and Technology
2026, VOL. 7, NO. 2, 116-121
DOI: 10.47297/taposatWSP2633-456923.20260702

A Technical Survey of Network Intrusion Detection Based on Deep Learning

Ruibo Shi

College of Computer Science and Mathematics, Anyang University, Xinxiang, Henan, 453500, China

ABSTRACT
Traditional Intrusion Detection Systems (IDS) exhibit notable limitations in addressing complex network attacks, failing to meet the requirements of practical intrusion detection scenarios.Leveraging its robust feature learning capability, deep learning offers an effective solution to overcome the shortcomings of traditional IDS and enhance the performance of network intrusion detection. This paper systematically reviews deep learning-based network intrusion detection technologies, summarizes mainstream models, various learning paradigms, and key challenges in practical deployment, proposes future research directions based on the limitations of existing studies, and provides insights for the development of efficient and deployable intelligent intrusion detection systems.

KEYWORDS
Deep learning; Network intrusion detection; Representation learning; Interpretability; Federated learning; Adversarial examples

1 Introduction

1.1 Research Background

1.1.1 Evolution and Complexification of Cyber Threat Situation

With the explosive growth of network scale and the number of interconnected devices, cyber threats have evolved from single and explicit attacks into a complex situation characterized by polymorphism and concealment. Traditional explicit attacks have gradually been replaced by more sophisticated and covert attack methods, where attackers make full use of distributed resources and multi-stage covert strategies, resulting in diluted detection signals and increased noise.

The novel communication patterns and cross-domain interactions introduced by the Internet of Things and cloud-native architectures have expanded the attack surface and lateral movement paths, and the frequent emergence of zero-day vulnerabilities and customized attack samples has rendered signature-based detection methods— which rely on rule bases, suffer from update lags, and struggle to cope with unknown attacks and encrypted traffic— ineffective. Faced with the ever-changing threats, intrusion detection systems are required to process high-dimensional traffic data and long-term temporal dependencies simultaneously, and their models are demanded to have real-time response capability and maintain robustness under the conditions of scarce samples and distribution drift.

1.1.2 Limitations of Traditional Intrusion Detection Methods

Traditional intrusion detection methods are dominated by rule matching and statistical features. Although they achieve high accuracy in identifying known threats, they exhibit significant limitations when confronted with unknown attacks, multi-stage attacks, and variant samples. Signature-based approaches rely on pre-defined rule bases, and their delayed updates make it difficult to detect zero-day attacks. Statistical methods are sensitive to changes in feature distributions and vulnerable to interference from encrypted traffic and high-dimensional noise, resulting in elevated false positive and false negative rates. Furthermore, these methods typically assume that samples are independently and identically distributed, ignoring temporal correlation and contextual information, which limits their ability to identify complex attack chains. In resource-constrained environments, the overhead of rule engines and feature calculation makes it hard to guarantee real-time performance. Manually constructed features depend heavily on expert experience and are difficult to adapt to new scenarios. Their insufficient generality and scalability restrict widespread deployment in large-scale network environments.

1.1.3 The Emergence and Potential of Deep Learning in the Field of Cybersecurity

With its powerful multi-layer nonlinear mapping and representation capability over massive data, deep learning has rapidly become an important tool in the field of cybersecurity. Compared with traditional feature engineering-driven approaches, deep models can automatically learn high-level features from raw traffic or packet sequences and effectively handle spatial and temporal information in data, which greatly promotes the development of security defense by enabling more accurate capture of potential security risks.

Deep learning exhibits unique potential in adversarial detection, traffic generation, attack traceability, and other directions. It enables detection systems to evolve from single signature recognition toward intelligent perception and

* **Corresponding Author:** Ruibo Shi, udmadh@qq.com

active defense, promoting the systematic upgrading of cybersecurity capabilities.

1.2 Research Objectives and Research Questions

1.2.1 Research Objective: To Explore the Application and Challenges of Deep Learning in Network Intrusion Detection

This study aims to systematically evaluate the applicability and limitations of deep learning techniques in Network Intrusion Detection Systems (NIDS), focusing not only on the improvement of detection performance but also on the feasibility of engineering deployment.

The specific objectives are as follows:

To compare the representation capabilities of typical deep learning models (e.g., CNN, RNN, GNN, and self-attention mechanisms) for spatial and temporal traffic features.

To investigate robust learning strategies under label scarcity or distribution drift, including supervised, semi-supervised, and self-supervised learning methods. To evaluate model performance in resource-constrained environments, analyze model interpretability and adversarial example defense, and propose feasible evaluation metrics and improvement approaches.

1.2.2 Research Questions: Issues of Model Selection, Feature Representation, Real-time Performance, and Interpretability

The research questions are reflected in four interrelated technical challenges.

First, model selection: different deep learning architectures (such as CNN, RNN, GNN, and self-attention models) have different emphases on capturing spatial, temporal, and correlation features of traffic, requiring a trade-off between accuracy and computational cost according to attack types and deployment environments. Second, feature representation: how to construct effective representations across byte-level, packet-level, and session-level, and improve generalization ability through embedding vectors or self-supervised pre-training methods, is the key to increasing detection rate. Third, real-time performance: intrusion detection systems demand low latency and online inference on streaming data, which requires optimization between model complexity and throughput, with a common constraint expressible as $T_{latency} \leqslant T_{threshold}$. Finally, interpretability: security operation and maintenance require understanding the decision basis of the model, so interpretable feature importance metrics and visualization methods should be designed to support event traceability and response decisions.

1.3 Research Significance

1.3.1 Theoretical Significance: Promote the Innovation of Network Intrusion Detection Methodologies

The introduction of deep learning into network intrusion detection has not only brought performance improvements, but also promoted innovations in concepts and paradigms at the methodological level. Traditional methods focus on feature engineering and rule matching, whereas deep learning emphasizes end-to-end representation learning, enabling models to automatically extract multi-scale and multi-granularity features from raw traffic or packet sequences and reduce reliance on manual prior knowledge. Theoretically, this encourages researchers to revisit the modeling framework of anomaly detection. For instance, intrusion detection can be formulated as a sequence-to-sequence mapping, density estimation, or contrastive learning task, with typical formulations including maximum likelihood estimation or the contrastive loss $L_{contrast}$. The transferability and hierarchical representation of deep models have facilitated theoretical investigations into cross-domain knowledge transfer and few-shot learning, driving the adaptation and development of statistical learning theory, generalization bounds, and robustness analysis in cybersecurity scenarios, thereby enriching the theoretical foundation of intrusion detection.

1.3.2 Practical Significance: Improve Detection Effectiveness and Response Efficiency, and Support Security Operation And Maintenance Decisions

The practical significance of deep learning in network intrusion detection is reflected in multiple aspects.

It can significantly improve detection effectiveness. Through end-to-end feature learning, models can capture high-dimensional and nonlinear correlations that are difficult to cover by traditional rules, thereby increasing the detection rate and reducing the false positive rate.

Automated feature extraction and online learning mechanisms help improve response efficiency, enabling security systems to adapt and update more quickly against new attacks and shortening the delay from threat detection to response. For operation and maintenance decisions, deep models can output behavioral representations and risk scores. For example, session risk can be quantified as $R \in [0,1]$, and alarm and isolation actions can be triggered by combining the threshold strategy $I(R > \tau)$, facilitating the formation of a traceable disposal process.

The visualization and interpretable modules of the model can provide decision support for security analysts, enabling post-incident traceability and policy optimization, thereby improving the overall security operation efficiency and defense capability.

2 Literature Review

2.1 Foreign Research Status

2.1.1 The Development of Classical Machine Learning and Deep Learning in Intrusion Detection

Traditional intrusion detection based on statistical methods and classical machine learning excels in known attack scenarios but lacks generalization to unknown and variant threats. This bottleneck has spurred the shift to deep learning, whose strong feature learning capability enables more robust detection against complex network traffic.

With the expansion of network scale and the escalation of attack methods, researchers introduced deep learning. Through multi-layer nonlinear transformation, it automatically learns high-level traffic features, captures complex temporal dependencies and cross-session semantics, and improves the ability to recognize complex attacks and noisy environments.

Literatures show that although soft computing methods such as self-organizing maps provide insights for feature selection and preliminary anomaly detection, they rely on manual design or shallow representations and are difficult to adapt to large-scale scenarios. In contrast, deep models can reduce the dependence on manual feature engineering in an end-to-end manner, learn more expressive semantic representations from raw traffic, and exhibit stronger robustness and generalization.

Practice emphasizes that dataset quality and labeling strategies are critical to model performance. Under data imbalance and scarce labels, semi-supervised or self-supervised learning should be combined to compensate for the limitations of traditional supervised learning.

Classical machine learning has laid a methodological foundation for intrusion detection, while deep learning has promoted the evolution of this field toward higher-level representation learning and automated feature extraction, providing a new technical path for addressing complex network threats [1].

2.1.2 Research on Traffic Feature Extraction Based on Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) have been introduced into the field of network traffic analysis by virtue of their advantages in image feature extraction, enabling automatic learning of local patterns and hierarchical representations from raw byte streams or structured traffic matrices.

Hounsou et al. (2019) emphasized the importance of feature representation and adaptive selection for detection performance, which indirectly supports the idea of replacing manual feature engineering with deep convolution [1].

Researchers usually map packets or sessions into two-dimensional arrays ("images"), and use convolution kernels of different scales to capture local byte correlations and protocol field combination features, enhancing sensitivity to malicious patterns.

Unlike traditional methods that rely on manually designed statistical features, CNN abstracts low-level byte patterns into high-level semantic representations through multi-layer convolution and pooling, helping to distinguish various attacks such as scanning and DDoS.

Practice shows that strategies such as data augmentation and sliding windows can alleviate generalization problems caused by traffic diversity and class imbalance, but the model still has limitations in encrypted traffic processing and fine-grained behavior recognition.

The key direction for future research remains to improve the ability to detect covert attacks and cross-domain scenarios while ensuring real-time performance and interpretability [1].

2.1.3 Research on Sequential Behavior Detection Based on Recurrent Neural Networks (RNN/LSTM/GRU)

Recurrent Neural Networks (RNNs) and their variants (LSTM, GRU) are widely used in sequential behavior detection due to their inherent strength in capturing temporal dependencies in network traffic.

Researchers model traffic or sessions as time series and use RNNs to learn temporal patterns, identifying anomalous behaviors that require temporal context for discrimination, such as persistent scanning and slow data exfiltration.

Early studies focused on encoding raw traffic or statistical features into RNN-compatible inputs to verify their effectiveness in short-term attack recognition.

Subsequently, LSTM and GRU have become mainstream due to their superior ability to preserve long-term dependencies and training stability, achieving better performance in long-duration low-speed attacks and multi-stage intrusion detection.

To address the problems of high noise in network data and uneven session lengths, scholars have proposed multi-scale modeling and segmented aggregation strategies to improve detection robustness.

In the face of scarce annotations and imbalanced samples, methods such as sequence autoencoders and contrastive learning enable the RNN family to maintain satisfactory performance in unsupervised and semi-supervised scenarios.

In engineering deployment, inference latency and resource consumption are optimized through model compression, knowledge distillation, and other strategies.

Some studies have also integrated RNNs with rule engines and graph neural networks to compensate for the shortcomings of single temporal models.

At present, RNN-based sequence detection has formed a complete technical chain, but in-depth research is still required in terms of interpretability, adversarial robustness, and cross-domain transferability.

2.1.4 Application of Self-Supervised Learning and Representation Learning in Low-Label Scenarios

Self-supervised learning and representation learning provide a new solution for network intrusion detection scenarios with scarce annotations.

Şen (2023) pointed out that the reliance of traditional supervised methods on large amounts of high-quality labeled data is particularly prominent in the security field, while intrusion samples are rare and labeling is costly, making it difficult for models to generalize effectively to unknown attacks or heterogeneous traffic [2].

To this end, researchers adopt self-supervised pre-training strategies: general representations are learned on large-scale unlabeled traffic by designing proxy tasks (such as sequence reconstruction, contrastive learning, or predicting the features of the next packet), and then fine-tuned with a small amount of labeled data, thereby significantly improving detection performance and robustness under limited label conditions.

Correspondingly, representation learning compresses high-dimensional raw traffic into low-dimensional but semantically rich vector representations, which helps to enhance the separability between normal and abnormal patterns and reduce the impact of noise.

Şen further emphasized that the combination of self-supervised and representation learning can not only reduce the dependence on manual annotation, but also lay a foundation for cross-domain transfer and online updating, enabling intrusion detection systems to have stronger adaptability and continuous learning ability when dealing with new attacks and traffic evolution [2].

2.1.5 Advances in Interpretability, Security Robustness, and Adversarial Examples Research

Research on interpretability, security robustness, and adversarial examples has become a critical issue in deep learning-based intrusion detection.

Hounsou et al. (2019) pointed out that although black-box models have advantages in detection rate, their decision paths and feature contributions are difficult to interpret, making it challenging for security analysts to judge the credibility and forensic value of alerts. Therefore, it is urgent to introduce interpretability mechanisms in practical deployment to improve trust and usability [1].

From a methodological perspective, Şen (2023) noted that models exhibit vulnerability when facing abnormal inputs and distribution drift, and recommended enhancing robustness by improving model structures and training strategies—such as adversarial training and uncertainty estimation—while using interpretability techniques to locate misjudgment sources, so as to build a more reliable defense closed-loop [2].

Focusing on engineering practice, Tufféry (2022) summarized experimental methods for evaluating the impact of adversarial examples and suggested introducing adversarial example testing, post-compression interpretability verification, and online monitoring into the model evaluation and deployment process simultaneously, so as to improve the robustness and maintainability of models in real network environments while maintaining performance [3].

2.2 Domestic Research Status

2.2.1 Research Hotspots of Domestic Academia in the Field of Deep Learning-Based Intrusion Detection

Domestic academic research in the field of deep learning-based intrusion detection is both diverse and practical. Researchers have continuously made attempts in model construction and improvement by organically integrating convolutional neural networks, recurrent neural networks and various attention mechanisms, supplemented by traditional feature engineering, so as to achieve both higher detection accuracy and lower false positive rate in traffic representation and behavioral sequence modeling. Such work continues the consensus that deep learning can significantly improve anomaly detection capability, while focusing on addressing data noise and class imbalance in real network environments [4].

Faced with scarce annotations and privacy constraints, domestic scholars have actively promoted research on paradigms such as semi-supervised learning, self-supervised learning and federated learning. They strive to reduce the dependence on manual annotation while improving the effect of representation learning, and explore practical paths to realize collaborative modeling and privacy protection in scenarios where data cannot be centrally shared, reflecting the practical demand for transforming methodologies into industrial applications [5].

2.2.2 Exploration of Dataset Construction and Traffic Labeling Methods

In the field of network intrusion detection, dataset construction and traffic labeling methods directly affect model training effectiveness and evaluation reliability.

Chan and Yeoh (2018) emphasized the importance of collecting traffic close to real network environments and simulation capabilities by building a platform for exploring intrusion detection systems. They pointed out that relying solely on laboratory-synthesized traffic weakens the model's generalization in practical applications, and recommended introducing diverse normal behaviors and attack scenarios into datasets to improve representativeness [6].

Gupta et al. (2025) focused on the relationship between the labeling process and classification performance, proposing a combination of automated detection and manual review to improve labeling efficiency while ensuring label accuracy, thereby enhancing the performance of supervised learning methods. They also stressed the key role of public and reproducible datasets in promoting academic comparison and method iteration [7].

Both indicate that future dataset construction should balance scenario realism, diverse coverage, and strict quality control to provide a more robust foundation for training and evaluating deep learning models.

2.2.3 Industrial Practices in Deployment and Performance Optimization

Industrial practices in applying deep learning techniques to network intrusion detection have demonstrated the feasibility of transformation from research to engineering, while also exposing several practical constraints.

In practice, operation and engineering teams pay close attention to the real-time performance and resource consumption of models. Therefore, engineering methods such as model compression, quantization, and inference acceleration are often adopted to reduce computational and memory overhead to meet the requirements of high throughput and low latency.

Enterprises tend to take deep learning models as a supplement in a multi-layer defense system, deploying them collaboratively with traditional mechanisms including rule-based detection, threat intelligence, and behavioral analysis, so as to improve the overall detection rate while controlling the false positive rate [8].

Faced with distributed data and privacy constraints brought by the Internet of Things and edge computing, the industry is also experimenting with distributed training, federated learning, and other schemes.

By completing representation learning locally and exchanging only model parameters or updates, these approaches alleviate the data island problem and protect data privacy as much as possible.

However, in actual deployment, it is necessary to balance communication overhead, convergence speed, and system complexity, as well as strengthen the protection and monitoring against security risks such as adversarial examples [9].

2.2.4 Research Challenges Under the Constraints of Regulations, Privacy and Data Sharing

Under the constraints of regulations, privacy and data sharing, researchers worldwide generally agree that legal and compliance requirements impose significant restrictions on the data acquisition and exchange of network intrusion detection systems, thereby affecting the consistency and reliability of model training and evaluation.

In their research on system implementation, Ourida et al. emphasized that intrusion detection systems must balance the legality of data collection and user privacy protection during deployment; otherwise, it is difficult to obtain sufficient and realistic traffic samples to support model development and validation [10]. From the theoretical and practical perspectives of deep learning methods, Santry pointed out that training data quality determines the generalization ability of models. Under data-limited conditions, models are prone to overfitting and evaluation bias. Therefore, researchers need to adopt stricter validation strategies and privacy-preserving techniques to ensure the credibility of conclusions [11].

In their systematic review, Thakkar and Lohiya summarized the challenges of intrusion detection in feature selection, model evaluation and practical deployment. They particularly emphasized that under regulatory and privacy constraints, incomplete data labeling, unavailability of cross-domain data and the lack of unified evaluation benchmarks hinder fair comparisons between methods and impede the adoption of academic achievements in industry [12].

To solve the "data island" dilemma caused by regulations and privacy, it is necessary to promote feasible technologies such as federated learning and differential privacy at the technical level, and establish standardized data sharing protocols and compliance frameworks at the institutional level. Only through the coordination of both can the transformation of research results into practical applications be promoted.

3 Conclusions and Prospects

3.1 Conclusion Summary

3.1.1 Deep Learning has Significantly Improved the Capabilities of Anomaly Detection and Complex Attack Identification

With multi-layer nonlinear transformation and end-to-end learning capabilities, deep learning can automatically learn high-level representations from massive traffic, capture implicit temporal patterns, spatial correlations, and protocol semantics, significantly improving the recognition effect of network anomalies and complex attacks compared with traditional methods that rely on manual features and fixed rules. Among them, Convolutional Neural Networks (CNNs) are good at extracting local traffic patterns, Recurrent Neural Networks (RNNs) and attention mechanisms can effectively model session-level temporal dependencies, and Graph Neural Networks (GNNs) perform prominently in depicting host and traffic relationships. These models all show higher detection rates and lower false positive rates in public datasets and practical scenarios, laying a solid foundation for building an adaptive defense system.

3.1.2 Federated Learning, Few-shot Learning, and Cross-domain Transfer are Key Directions for the Future

Faced with the dilemma of data silos, scarce annotations and distribution differences in real network environments, federated learning, few-shot learning and cross-domain transfer are three key directions for the practical application of intrusion detection. Federated learning enables collaborative modeling without sharing raw data, ensuring privacy and improving cross-organizational threat awareness; few-shot learning achieves rapid generalization from limited labeled samples, reducing annotation costs and adapting to the identification of new attacks; cross-domain transfer narrows the distribution gap between training and deployment environments to realize effective knowledge transfer. The three approaches are complementary to each other, and their collaborative development can significantly improve the adaptability and generalization ability of detection systems, providing a more flexible and robust technical path for practical deployment in industrial scenarios.

3.2 Future Outlook

3.2.1 Establishment of Standardized, High-quality Public Datasets and Evaluation Benchmarks

Constructing standardized, high-quality public datasets and evaluation benchmarks is the cornerstone for promoting deep learning-based network intrusion detection research from the laboratory to engineering implementation. Priority should be given to collecting multi-source, multi-level network traffic samples covering normal services, known attacks, and emerging threats, while providing clear and consistent labeling specifications and metadata descriptions. Evaluation benchmarks require unified definitions of preprocessing pipelines, partitioning strategies, and performance metrics, balancing practical concerns such as detection rate, false positive rate, latency, and resource consumption, while supporting testing of cross-domain transfer and online update capabilities. Establishing an open data-sharing platform and privacy-preserving mechanisms, and encouraging academia and industry to jointly maintain data quality and benchmark evolution, will lay a solid foundation for algorithm comparison, fair evaluation, and long-term reproducible research.

3.2.2 Strengthen Industry-academia Collaboration to Improve the Deployability and Maintainability of Models in Production Environments

To improve the deployment capability of deep learning-based intrusion detection technologies, collaborative innovation between industry and academia should be strengthened. Academia excels in theoretical and methodological exploration, while industry possesses real traffic, engineering implementation, and operational experience. Their complementarity can address the challenges of data distribution drift, latency constraints, and resource limitations that models face in production environments. Joint experimental platforms and open benchmarks should be established to promote real-scenario-based data sharing and privacy-preserving mechanisms, and to conduct cross-domain validation and long-term evaluation. Joint efforts are encouraged to tackle engineering issues such as model compression, online learning, and fault self-healing, forming reusable deployment frameworks and interface standards.

Through sustained industry-academia interaction, the transformation of models from proof-of-concept to stable and maintainable industrial systems can be accelerated, ultimately enhancing the practical effectiveness and sustainable operation capability of network defense.

References

[1] Hounsou, J.T., Nsabimana, T., Degila, J. (2019) Implementation of Network Intrusion Detection System Using Soft Computing Algorithms (Self Organizing Feature Map and Genetic Algorithm). Journal of Information Security, 10: 1–24.

[2] Şen, Z. (2023) Deep Learning. In: Şen, Z. (Ed.), Shallow and Deep Learning Principles. Springer International Publishing, Cham. pp. 9.

[3] Tufféry, S. (2022) Deep Learning: From Big Data to Artificial Intelligence with R. Wiley, Hoboken.

[4] Ahmad, Z., Khan, A. S., Shiang, C. W., Abdullah, J., Ahmad, F. (2021) Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches. Transactions on Emerging Telecommunications Technologies, 32: e4150.

[5] Alsharafat, W. (2019) A New FLAME Selection Method for Intrusion Detection (FLAME-ID). Communications and Network, 11: 11-20.

[6] Chan, C.K., Yeoh, A.W.T. (2018) Development of a Platform to Explore Network Intrusion Detection System (NIDS) for Cybersecurity. Journal of Computer and Communications, 06: 1-11.

[7] Gupta, C., Kumar, A., Jain, N.K. (2025) Intelligent Intrusion Detection System Based on Crowd Search Optimization for Attack Classification in Network Security. EURASIP Journal on Information Security, 2025: 22.

[8] Liao, H. J., Lin, C. H. R., Lin, Y. C., Tung, K. Y. (2013) Intrusion Detection System: A Comprehensive Review. Journal of Network and Computer Applications, 36: 16-24.

[9] Mahjoub, C., Hamdi, M., Alkanhel, R. I., Mohamed, S., Ejbali, R. (2024) An Adversarial Environment Reinforcement Learning-Driven Intrusion Detection Algorithm for Internet of Things. EURASIP Journal on Wireless Communications and Networking, 2024: 21.

[10] SAIDI BEN BOUBAKER, O. (2012) Implementation of an Intrusion Detection System, 9: 3.

[11] Santry, D.J. (2023) Demystifying Deep Learning: An Introduction to the Mathematics of Neural Networks. Wiley, Hoboken.

[12] Thakkar, A., Lohiya, R. (2022) A Survey on Intrusion Detection System: Feature Selection, Model, Performance Measures, Application Perspective, Challenges, and Future Research Directions. Artificial Intelligence Review, 55: 453-563.