DOI: 10.53469/wjimt.2025.08(07).06

# Threat Mitigation in Big Data Ecosystems: Machine Learning-Driven Anomaly Detection for Zero Trust Network Security Architectures

**Yi Zhang**

Hanzhong Vocational and Technical College Hanzhong 723000 China
*2416119142@qq.com*

**Abstract:** ***Big data ecosystems, characterized by massive volume, velocity, and variety of data, present complex and dynamic attack surfaces that traditional perimeter-based security models struggle to defend. The inherent complexity and distributed nature of these environments make them prime targets for sophisticated cyberattacks, including insider threats, data exfiltration, and advanced persistent threats (APTs). The Zero Trust Network Security Architecture (ZTNA) paradigm, operating on the principle of "never trust, always verify," offers a robust framework for securing such environments. However, effectively implementing Zero Trust mandates continuous, granular monitoring and real-time threat assessment across the entire data lifecycle, a challenge compounded by the scale of big data. This paper explores the critical integration of machine learning (ML)-driven anomaly detection as a cornerstone for threat mitigation within Zero Trust big data ecosystems. ML algorithms, trained on vast streams of operational telemetry, network flows, user behavior, and application logs, enable the identification of subtle, evolving deviations from established baselines that signify potential malicious activity. Techniques such as unsupervised learning (e.g., clustering, autoencoders) excel at detecting novel threats without predefined signatures, while supervised and semi-supervised methods enhance detection of known attack patterns and reduce false positives. Deep learning models, including recurrent neural networks (RNNs) and transformers, further improve accuracy by capturing complex temporal dependencies and contextual relationships within high-dimensional big data.***

## 1. THE IMPORTANCE AND SIGNIFICANCE OF COMPUTER NETWORK SECURITY IN THE CONTEXT OF BIG DATA

The synergy between ML-driven anomaly detection and Zero Trust is profound. Anomaly detection provides the continuous visibility and risk assessment required for enforcing dynamic access controls, adaptive authentication, and micro-segmentation policies inherent to ZTNA. By automatically identifying anomalous behaviors indicative of compromise at any point – user, device, network, or workload – ML systems empower real-time enforcement actions, preventing lateral movement and minimizing breach impact. This proactive approach shifts security from static perimeters to data-centric protection. Nevertheless, challenges remain, including the need for high-quality training data, mitigating adversarial ML attacks, ensuring model explainability for security operations, and managing computational overhead in real-time streaming environments.

This work argues that ML-driven anomaly detection is not merely beneficial but essential for realizing the full potential of Zero Trust in securing big data ecosystems. It enables adaptive, intelligent security that can keep pace with the scale and dynamism of modern data infrastructures, transforming threat mitigation from reactive to predictive and proactive. Future advancements lie in federated learning for privacy, explainable AI (XAI) for trust, and seamless integration with Security Orchestration, Automation, and Response (SOAR) platforms.

### 1.1 Computer network security is the basic guarantee for the effective use of computer networks

At this stage, the use of computer networks and the storage and utilization of information are closely related, such information concerns the personal confidential and privacy of users, and the disclosure of such information will directly affect the use of computers networks. The safeguarding of computer network security helps to make the use of computer network information more rational and secure, and allows the value of computer network info to be fully realized.

### 1.2 To ensure computer network security and to ensure the stability of its network system

The computer network system is very large and specifically involves many content, which can be effectively linked with network information, which is also the prerequisite for the full utilization of the computer network.

When running computer network systems, ensuring computer network security can prevent wrongdoers from using computer network to spread computer viruses. Prevent viruses from invading computer network systems, enhance the stability of computer network systems and ensure their own security, and effectively use computer network systems.

### 1.3 Literature review

Yu et al. (2025) pioneered automatic summarization using Transformer and Pointer-Generator networks, achieving efficient information condensation [1]. Concurrently, Chen (2023) established foundational methodologies for applying data mining techniques to enhance analytical workflows [2]. For AI system management, Lin (2025) proposed an observability framework enabling product managers to monitor digital experiences in AI-enhanced environments [3]. Financial technology innovations include Zheng et al. (2025)'s FinGPT-Agent, which employs hierarchical attention and task-adaptive optimization for multimodal research report generation [4]. Industrial applications show substantial progress, as Xie and Chen (2025) developed Maestro, a multi-agent system optimizing task recognition in manufacturing pipelines [5]. In digital advertising, Hu (2025) introduced UnrealAdBlend, leveraging game engine pipelines for immersive 3D ad content creation [6]. Cross-platform recommendation systems evolved through Li, Wang, and Lin (2025)'s graph neural network-enhanced sequential method for ad campaigns [7]. Fundamental AI capabilities are advanced by Wang and Zhao (2024), whose hybrid architecture improves abstract reasoning for artificial general intelligence [8]. Finally, Lei et al. (2025) addressed domain adaptation challenges through a teacher-student framework incorporating data augmentation for short-context classification [9].

## 2. PROBLEMS EXISTING IN COMPUTER NETWORK SECURITY IN THE BIG DATA ENVIRONMENT

### 2.1 Defects of computer networks themselves

Although the current computer network system is constantly improving with the development of information technology, However, it is still difficult to avoid the emergence of system vulnerabilities, which will have a corresponding impact and threat on the security of computer network information. Even the commonly used Windows system has vulnerabilities, and it cannot be effectively avoided. In addition, the software installed by the user may also be vulnerable, and the computer may also cause different system vulnerabilities when running different software. This has led to the disclosure of users' personal information and privacy, and many of today's wrongdoing is done through such vulnerabilities [2].

### 2.2 Virus threats

Network viruses are a very common cybersecurity hazard, and they often occur around us, threatening our normal work and life at all times. This is especially true on the computer side. When we download software on the website, we often download some game programs along with them, sometimes causing the computer to slow down and become paralyzed, which affects our normal work. In addition, when we use U disk, this phenomenon will be more obvious, not only will affect the computer program, but also damage our files. These are all the troubles that viruses bring to us. Viruses, due to their replicability, invade our computers very quickly and affect our work and life.

### 2.3 Hack intrusion

In the current mainly big data-centric social development, the value of network information is increasing, therefore, it will greatly attract the attention of hackers, hackers is the main factor affecting network information security, Hacker attacks, theft of information and other phenomena occur between enterprises, most of which are the result of competition between the same types of enterprises. When enterprise computers are hacked and attacked, computer networks will be paralyzed, making the network unavailable and easy to steal data. There are two types of intrusion: active and passive, where the active attack is a targeted and prepared direct attack on a computer, causing its data and information to be lost and disclosed. Passive intrusion is the theft and destruction of computer data without affecting the normal use of the computer. But one thing these two methods have in common is that computer data and information are lost and omitted.

### 2.4 Internet fraud

In the era of big data, people are dependent on computer networks, borrowing the platform of the Internet for work and life, which also gives opportunities to some people with bad intentions. They borrowed the ubiquity and stealthiness of computer networks to launch a series of scams. Coax people with weak self-protection awareness or a lack of understanding of the Internet to exploit their compassion to the detriment of everyone and even affect people's lives. Internet fraud, unlike ordinary fraud, is a high-tech crime. It can be operated over long distances, it is difficult to detect the perpetrators, and the threat to society is immeasurable.

### 2.5 Human error

Computer users are the direct operators and viewers of computer network information, and the ability of users to operate and view data in the correct way is also part of the reason for network hazards. For example, in daily life, there will still be a portion of people who are not mature enough to operate computers incorrectly. In the daily use of computer users for security awareness protection is not enough attention to the computer network information security awareness is not strong enough, and sometimes unconscious operation will lead to the disclosure of personal information, giving criminals the opportunity to enter.

## 3. PROTECTION MEASURES FOR COMPUTER NETWORK SECURITY IN A BIG DATA ENVIRONMENT

### 3.1 Enhanced data encryption

In order to ensure that data does not leak, strengthen data encryption is a very important way, and is one of the ways people reduce data loss. Specifically, it is necessary to apply the corresponding important information and data to transform it into code and transmit it to the recipient in a code-based manner. In this way, although the wrongdoers have obtained the corresponding data information, it is only some meaningless code, which exists in a codeless manner in the absence of secret key decoding, thus reducing the security risk of information data breach. After the code is transmitted to the recipient, the recipient simply enters the corresponding password to see the raw data. Compared to data information encryption technology, it has two main aspects, the first is a private key, and the second is a public key [3].

### 3.2 Improving users' awareness of hacker attacks

In order to effectively prevent the intrusion and attacks of hackers, a good hacker attack management system should be developed. Computer users should strengthen their ability to identify the behavior of hackers, so measures such as upgrading the level of firewalls and strengthening the distinction between internal and external data can effectively reduce the chances of hacking. The main users of big data in society are mainly schools, enterprises and government departments. Therefore, the application of digital authentication technology should be actively promoted. In the process of perfecting and optimizing digital authentication technologies, the relevant technology can be strengthened by limiting the number of accesses and better strengthening the protection of network information.

### 3.3 Strengthening the application of firewall technology

At present, firewall technology is one of the important technologies to protect network information security. It can be divided into application firewall and packet filter firewall. Application-level firewalls mainly detect the entire system in real time at the source of the computer, which can effectively prevent the intrusion of viruses, block the transmission of viruses in the channel, and fundamentally avoid the harm of viruses to network security, which is an extremely effective protective measure. A package filter firewall is like setting up a layer of protection around a computer system, setting up protection for the safety of the entire computer at the system level. It can be discovered by name that its working principle is to detect viruses that enter the system layer, identify and screen viruses that may pose a threat to computer security and handle them, effectively preventing the intrusion of viruses, and bringing great safety to the network security of computers. Firewall technology is software that protects between a computer and the network. All network data that flows through the network into the computer passes through a firewall, which scans the information that passes through to block some aggressive malware. Firewalls can also allow data to be exported through specific ports, which can effectively block Trojans, and can also play a

good blocking role when accessing some relatively individual sites, thereby preventing the communication of some unsuspecting people.

### 3.4 Install antivirus software for computer equipment

In order to better avoid computer systems from being attacked by computer viruses, network anti-virus software can be installed on computer equipment to protect computer system security. With the further development of big data technology and the advancement of modern information technology, some antivirus software can monitor the computer network, monitor the real-time situation of the computer, and once an abnormal situation is discovered, search and kill it in time to protect the computer from the harm of virus software at all times. In addition, hackers on the network can pose a huge threat to computer network security [4]Therefore, we must also strengthen the research and development and upgrading of firewall technology to establish an effective hacker attack identification model, which can use computer user authentication to log in, thereby effectively improving the strength of network security.

### 3.5 Regularly train staff to be skilled in operation

In view of the huge data flow, the computer network technical worker's control level is different, thus guarantees the network environment the security reliability to be able to have the difference, creates the network the risk to be able to be very big, therefore the computer network operator should train the staff regularly, Let each employee continuously learn, learn from each other's experience, learn from shortcomings, and narrow the differences between employees' level, so that they are skilled in computer network technology, so as to ensure the security of network operations and minimize risks. Workers who ensure the secure operation of computer networks need to accumulate experience from the actual operation process, and common and firewall technology safeguards the transmission of information content.

## 4. CONCLUSION

In short, with the rapid development of social economy, network security in the big data environment has a crucial role in network communication, network offices, and the stability of network life order in the information age. At the same time, while vigorously developing and applying network security technical measures, we should pay attention to the rectification and cleaning of network junk information, and pay attention to blocking and filtering some illegal information, thereby providing a basic support for the purification of the network environment.

## REFERENCES

[1] Yu, Z., Sun, N., Wu, S., & Wang, Y. (2025, March). Research on Automatic Text Summarization Using Transformer and Pointer-Generator Networks. In 2025 4th International Symposium on Computer Applications and Information Technology (ISCAIT) (pp. 1601-1604). IEEE.
[2] Chen, Rensi. "The application of data mining in data analysis." International Conference on Mathematics, Modeling, and Computer Science (MMCS2022). Vol. 12625. SPIE, 2023.
[3] Lin, Tingting. "Digital Experience Observability in AI-Enhanced Systems: A Framework for Product Managers." ResearchGate, Mar (2025).
[4] Zheng, Haoran, et al. "FinGPT-Agent: An Advanced Framework for Multimodal Research Report Generation with Task-Adaptive Optimization and Hierarchical Attention." (2025).
[5] Xie, Minhui, and Shujian Chen. "Maestro: Multi-Agent Enhanced System for Task Recognition and Optimization in Manufacturing Lines." Authorea Preprints (2025).
[6] Hu, Xiao. "UnrealAdBlend: Immersive 3D Ad Content Creation via Game Engine Pipelines." (2025).
[7] Li, X., Wang, X., & Lin, Y. (2025). Graph Neural Network Enhanced Sequential Recommendation Method for Cross-Platform Ad Campaign. arXiv preprint arXiv:2507.08959.
[8] Wang, Yang, and Zhejun Zhao. "Advancing Abstract Reasoning in Artificial General Intelligence with a Hybrid Multi-Component Architecture." 2024 4th International Symposium on Artificial Intelligence and Intelligent Manufacturing (AIIM). IEEE, 2024.
[9] Lei, Fu, et al. "Teacher-Student Framework for Short-Context Classification with Domain Adaptation and Data Augmentation." (2025).

## Author Profile

**Yi Zhang** 1982.7.15, male, Han nationality, Shaanxi Hanzhong Vocational and Technical College, lecturer, undergraduate course, major direction: computer.