Proceedings of CONF-SEML 2025 Symposium: Machine Learning Theory and Applications
DOI: 10.54254/2755-2721/2025.TJ23838

Research Progress on Network Security Threat Detection and Defense Technology Based on Artificial Intelligence

Chenyuan Zhang

School of Information and Technology, PSB Academy of Singapore, Singapore
1558140400@qq.com

Abstract: Traditional defense mechanisms are no longer sufficient in dealing with the growing complexity of cyber threats—such as advanced persistent threats (APTs), zero-day exploits, and supply chain attacks. Artificial intelligence (AI) offers solutions to these issues by supporting automated threat identification, adaptive response measures, and predictive analysis techniques. The exploration in this article focuses on cybersecurity implementations driven by AI, highlighting machine learning techniques such as supervised classifiers, which address known threats, and unsupervised clustering used for detecting anomalies; along with deep learning methods like long short-term memory (LSTM) networks suited for analyzing temporal patterns, alongside graph neural networks (GNNs) aiding in designing attack path models. Additionally, this article looks at automated defense systems, like SOAR platforms and zero-trust architectures, and evaluates them with case studies on preventing ransomware and protecting cloud systems. Although AI's role is promising, it still faces significant obstacles such as adversarial tactics targeting vulnerable AI frameworks, privacy conflicts, particularly under regulations like GDPR, and challenges related to scalability. Future plans look at federated learning projects that promote teamwork without central control and include AI in new cryptographic methods to protect against risks from quantum computing. Through compiling recent research trajectories while focusing on practical usages, this paper serves to outline how to create secure, robust cybersecurity structures powered by AI safeguarding vital public infrastructure within an ever more competitive digital realm.

Keywords: Artificial Intelligence, Cybersecurity, Threat Detection, Machine Learning, Deep Learning, Adversarial Attacks, Zero Trust, Federated Learning.

1. Introduction

Due to the rapid development of network threats, traditional security measures find it difficult to keep up with the pace of complex attacks such as APT and zero-day vulnerabilities. This article examines the current artificial intelligence-based threat detection technology, analyzes key issues, and explores the future direction of supporting the development of more sustainable network security systems.

With the complexity of network attack methods (such as APT attacks and zero-day vulnerability exploitation), traditional security Defense technology is difficult to cope with in dynamic threat

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.TJ23838

environments. Artificial intelligence (AI) technology provides a new solution for network security through automated analysis, real-time response, and high-precision detection.

This study summarizes the application status of artificial intelligence technology in network threat detection and Defense, analyzes the advantages and limitations of different methods, and explores the future research direction.

2. Core types of network security threats

2.1. Traditional network attack

Denial of Service (DoS) is a traditional network attack. Denial of Service (DoS) attacks constitute one of the major threats and among the hardest security problems in today's internet. The computer network literature has clearly demonstrated the seriousness of Denial of Service (DoS) attacks on the Internet. The main aim of a DoS is the disruption of services by attempting to limit access to a machine or service instead of subverting the service itself. This kind of attack aims to render a network incapable of providing normal service by targeting either its bandwidth or its connectivity. These attacks achieve their goal by sending at a victim a stream of packets that swamps his network or processing capacity denying access to his regular clients [1]. A different flavor of DoS is Distributed DoS, or DDoS, involves a group of machines attacking a specific service. Distributed Denial of Service (DDoS) is a relatively simple yet very powerful technique to attack Internet resources. DDoS attacks add the many-to-one dimension to the DoS problem, making the prevention and mitigation of such attacks more difficult and the impact proportionally severe. DDoS exploits the inherent weakness of the Internet system architecture, its open resource access model, which ironically, also happens to be its greatest advantage [1]. Distributed Denial of Service (DDoS) attack is one such serious attack in the cloud space [2]. More than 20% of enterprises in the world saw at least one reported DDoS attack incident on their infrastructure [3]. Yang et al. [4] anticipated that DDoS attackers would increasingly target cloud infrastructure and services. This prediction has been supported by numerous attacks in recent years, some of which have drawn significant attention from the research community due to their scale and impact. In June 2023, the hacker organization Anonymous Sudan launched a DDoS attack on Microsoft's services, including Azure Portal, OneDrive and Microsoft 365. These attacks cause service interruptions and affect the access of global users. Microsoft confirmed these attacks and pointed out that the attackers used a variety of technical means, including HTTP(S) flood attacks and cache bypass. Another example is in August 2023, Google Cloud reported that it had suffered the largest DDoS attack so far, with a peak of 398 million requests per second (RPS). The attack exploits a new vulnerability in the HTTP/2 protocol, called "Rapid Reset", which allows attackers to crush the server by quickly resetting the connection. Despite the unprecedented scale of the attack, Google successfully mitigated the attack without long-term impact on the service. A report by Verisign Defense Security Intelligence Services [5] shows that the most attacked target of DDoS attacks in the last number of quarters is the cloud and SaaS (Software as a Service) sector. From these, the cloud and SaaS industries have always preferred DDoS attacks.

Over one-third of reported DDoS mitigations have occurred on cloud services, highlighting their vulnerability. A key consequence of such attacks in cloud environments is significant economic loss —averaging around $444,000 per incident [3], with some reports showing losses over $66,000 per hour [6]. In the cloud context, DDoS attacks behave differently, largely because of the impact on virtualized environments like IaaS, where client services run inside virtual machines.

DOI: 10.54254/2755-2721/2025.TJ23838

2.2. New threats

Over the past couple of years, a new class of threats has been seen – so-called Advanced Persistent Threats (APTs) [7] Advanced Persistent Threat (APT) is a term coined over the past couple of years for a new breed of insidious threats that use multiple attack techniques and vectors and that are conducted by stealth to avoid detection so that hackers can retain control over target systems unnoticed for long periods of time [7]. This feature makes the traditional threats different from the new threats. Table 1 shows the differences between them [8].

Table 1: The differences between traditional attack and APTs attack

	Traditional Attack	APTs Attack
Attacker	Mostly single person	Highly organized, sophisticated, determined and well-resourced group
Target	Unspecified, mostly individual systems	Specific organizations, governmental institutions, commercial enterprises
Purpose	Financial benefits, demonstrating abilities	Competitive advantages, strategic benefits
Approach	Single-run, “smash and grab”, short period	Repeated attempts, stays low and slow, adapts to resist defenses, long term

And even some large companies have not noticed these new threats. For example, the problem of CrowdStrike in recent years. Starting from noon on July 19, 2024, Beijing time, the update of the CrowdStrike problem caused a large-scale blue screen of Windows around the world, resulting in flight suspensions, train delays, bank abnormalities, Paris Olympic services, etc., affecting at least 20 countries around the world. CrowdStrike was founded in 2011 by two executives of the traditional anti-virus software McAfee. The team members are mainly from the information security industry, such as Microsoft and Amazon. The company is a world-renowned next-generation terminal security manufacturer. Its core products include the cloud-based Falcon platform and its modules, which cover multiple fields such as endpoint protection, threat intelligence, IT asset management, and malware search. The market value once exceeded 80 billion US dollars, second only to Palo Alto Networks, the largest network security company at that time. CrowdStrike has more than 24,000 customers, covering most of the world's top 500 enterprises. The malfunction of the Falcon platform's core component driver was the cause of this accident. Starting from 2 p.m. on Friday, July 19, 2024, Beijing time, a large number of Windows users around the world posted computer blue screen images on social media, and a large number of Windows computers crashed, displayed blue screen death, and could not be restarted. Because the Asia-Pacific region was during the day and the United States and Europe were at night at the time of the incident, the initial feedback on social media was mainly concentrated in the Asia-Pacific region, mainly in Japan and Australia. With the progress of time, European and American users have also received a large

DOI: 10.54254/2755-2721/2025.TJ23838

amount of feedback on service interruptions. Service interruptions at a large number of airports, hospitals, media outlets, and banks have been caused by system collapses. Tens of thousands of flights have been delayed and cancelled. Some hospitals have to transfer patients, and many affected enterprises have to take early holidays. The incident also affected Microsoft's cloud service, mainly because a large number of Windows-based application instances were running on Microsoft's cloud service, some of which were installed with CrowdStrike software, so even these virtual machines also crashed. Undoubtedly, the impact of CrowdStrike may also extend to Microsoft's management cloud application system.

3. Threat detection and defense technology based on artificial intelligence

3.1. Threat analysis driven by deep learning

LSTM helps address the vanishing gradient problem [9-11] and is capable of learning dependencies across more than 1000 time steps [11]. In LSTM networks, traditional hidden units are replaced by memory blocks, each containing at least one memory cell. Figure 1 illustrates a basic LSTM memory cell[13].

Memory cells are regulated by gates, which manage the flow of information in and out. Positioned between the input and output gates, the forget gate allows the model to reset the cell state when stored data is no longer useful. These gates use sigmoid activation functions that output values between 0 and 1.

The output $ycj(t)$ of an LSTM memory cell shown in Figure 1is computed as:

$$y^{cj}(\mathrm{t}) = y^{out_j}(\mathrm{t})\mathrm{h}\left(s_{cj}(\mathrm{t})\right) \quad (1)$$

where y^{out_j} is the output gate activation, s_{cj} is the internal state of the output gate, and h is the hidden layer output [12].

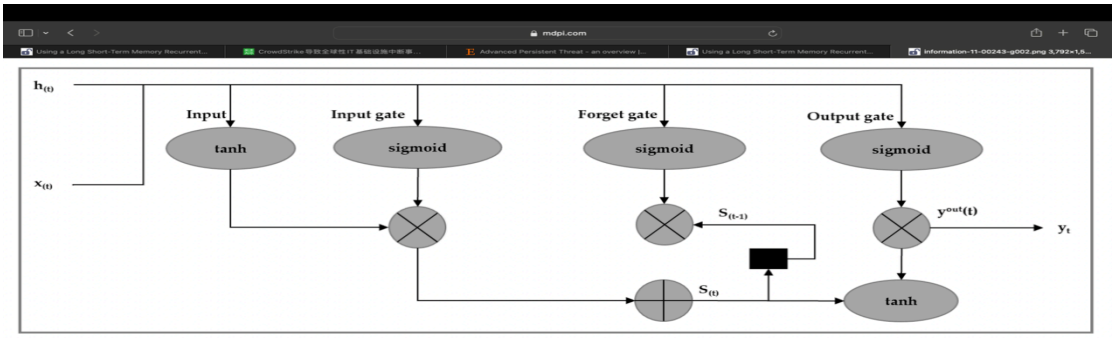

Figure 1: One cell in a basic LSTM network [13]

3.2. Automated response technology

Security orchestration, automation, and response (SOAR) tools are a new kind of technology that help improve efficiency, consistency, and security by automating common manual tasks. According to Gartner, SOAR tools can take in data from many sources and use workflows that follow company processes to work more efficiently. Unlike SIEM tools, which mainly collect and search through

DOI: 10.54254/2755-2721/2025.TJ23838

data, SOAR tools have flexible workflows or playbooks that help guide analysts and carry out response actions automatically. Both tools can bring in data from many sources and show real-time dashboards.

By automating routine tasks, ranking and linking alerts, adding extra context (such as Blacklist), and helping teams inside and across Security Operations Centers (SOCs) work together, SOAR tools aim to improve the speed and consistency of security operations. They do this by offering a single, easy-to-use platform where analysts can manage alerts supported by smart features, and by using playbooks to keep response steps consistent.This main point comes from[14]

4. Typical application scenarios and case analysis

4.1. Ransomware detection and blocking

Many studies have been carried out to tackle ransomware attacks, which are generally divided into three main categories: detection, prevention, and prediction. These efforts focus on identifying ransomware either while it is happening or after it has occurred. Ransomware detection methods are usually split into two types: structural and behavioral [15,16]. Mercaldo et al. [17] proposed a bytecode-based method to detect and remove ransomware from Android devices. However, it had a limited scope and struggled to disassemble all samples. Maiorca et al. [18] developed R-Pack Droid, a lightweight detection tool using supervised learning and API calls to classify apps. It detected known ransomware but lacked accuracy due to not analyzing runtime behavior and was not fully optimized for Android. Study [19] introduced a deep learning-based approach using a Deep Belief Network for embedded and IoT systems. It detected ransomware by analyzing bitstreams in memory, focusing on behavioral patterns rather than code structure. The literature [17] adopts a structural analysis method based on bytecode, [18] detects through static API characteristics and supervised learning, and [19] focusses on dynamic behavior analysis in combination with deep learning models. Overall, [17] and [18] perform well in detecting known ransomware, but have limited identification ability of new or variant ransomware; in contrast, [19] based on behavioral detection and deep learning, it has a stronger generalization ability and can more effectively cope with the evolving ransomware. Coerce.

Kharaz et al. [20]This paper proposes a Windows-based detection system named UNVEIL, which uses dynamic analysis method to detect ransomware. The system identifies ransomware activity by monitoring changes in the similarity of desktop screenshots. UNVEIL simulates real systems in a sandbox environment and observes the malware's interaction with the environment for detection. In addition, the system was able to detect previously unknown zero-day ransomware.

UNCOVER detects both real-time data flow and real-world data sources, and focuses on monitoring file write and delete requests by hooking up to file system drivers. To improve detection efficiency, the system operates at the kernel level and is specifically designed to detect file-locked and screen-locked ransomware.

Recent studies [21-24]have proposed various ransomware detection approaches, including dynamic analysis using honeypots, hybrid user-mode and kernel-mode monitoring, user-interactive defense mechanisms, and machine learning-based methods tailored for Android platforms. These systems focus on identifying malicious behaviors, enhancing data protection, controlling suspicious network connections, and optimizing feature selection for imbalanced datasets, aiming to improve the detection accuracy and resilience against ransomware attacks.

4.2. Threat defense in the cloud environment

DOI: 10.54254/2755-2721/2025.TJ23838

The advancement in the digital world has brought unique threats to Information Technology Security; current sophisticated organizations face innovative and continuous cyber threats. Conventional security models, which rely on established perimeters and boundaries, must be better suited to contend with such threats. Recently years, the Zero Trust Architecture (ZTA) concept has changed the traditional approach to cybersecurity by indicating that any entity, internal or external, must only be trusted once it is validated. This architectural style focuses strongly on controlling access, constant systematical checks, and dynamic authentication to reduce the risk as much as possible[25].

ZTA differs from traditional perimeter security and revolves around the concept of 'Never Trust' from request origin, implied identity, device, or application requesting access to a given network or data asset. ZTA is based on the assumption that breaches will occur and must be controlled, unlike traditional models that assume identities based on geographical position or pre-verified authentication. Since the initial conceptualization of ZTA, this model has quickly translated to industries and governmental organizations because of its capability to adapt to the risks in new decentralized and hybrid computing system schemes. Essentials of ZTA include identity and access control, segmentation, and monitoring, all of which reflect enterprises' current changing and growing demands in handling cloud systems, increased remote employees, and IoT ways of connecting to systems [25]. Technical Challenges : Contradiction between GDPR compliance and model training

In the process of training artificial intelligence models, the issue of GDPR compliance is increasingly prominent. First of all, GDPR requires that the explicit consent of the user must be obtained when collecting and processing personal data, and the purpose should be limited to a specific purpose. However, large-scale language models usually capture data from the Internet on a large scale through automated means. In the process, it is often difficult to ensure the legal source and authorization of each piece of data, and there is a risk of infringing on the rights of the data subject.

Secondly, GDPR gives individuals the "right to be forgotten," that is, the right to request the deletion of personal information related to themselves. However, in deep learning, once the training model is completed, the original data has been integrated into the model parameters. It is difficult to trace and delete specific data, making it challenging to comply with the "right to be forgotten" legal obligation.

In addition, GDPR emphasizes the transparency and interpretability of automated processing processes, but the existing deep neural network models generally have "black box" characteristics, and it is difficult to clearly explain the relationship between output results and training data, which also poses a challenge to compliance.

In summary, there is a fundamental institutional tension between the strict provisions of GDPR on the protection of data rights and the dependence on large-scale data, irreversible processing, and model complexity in AI model training.

5. Conclusion

The rapid evolution of cyber threats requires the exploration of advanced solutions beyond traditional Defense mechanisms. This article deeply analyzes the role of artificial intelligence in enhancing network security, focusing on technologies such as supervised learning, unsupervised clustering, and deep learning models, including long-term and short-term memory networks (LSTM) and graph neural networks, which are particularly suitable for time analysis and attack path

DOI: 10.54254/2755-2721/2025.TJ23838

modeling. Integrating artificial intelligence into automatic threat detection, response scheduling, and predictive analysis provides a promising method for dealing with known and emerging threats.

However, the implementation of artificial intelligence in network security is not without challenges. The fundamental tension between the data-intensive nature of the artificial intelligence model and the regulatory framework, such as GDPR poses a major obstacle to compliance, especially in the field related to data privacy and the "right to be forgotten." In addition, the inherent opacity of the deep learning model brings difficulties in meeting the requirements of transparency and interpretability.

Future research should focus on overcoming these obstacles by exploring distributed frameworks (such as federal learning), which allow collaborative Defence without sacrificing data privacy and prevent emerging threats from quantum computing through integrated quantum cryptography technology. By solving these challenges, artificial intelligence-driven security systems can become stronger, more expandable, and compliant with privacy regulations, providing a security foundation for the future of digital infrastructure.

References

[1]C. Douligeris and A. Mitrokotsa, “DDoS attacks and defense mechanisms: classification and state-of-the-art,” Computer Networks, vol. 44, no. 5, pp. 643–666, 2004.

[2]G. Somani, M. S. Gaur, D. Sanghi, M. Conti, and R. Buyya, “DDoS attacks in cloud computing: Issues, taxonomy, and future directions,” Computer Communications, vol. 107, pp. 30–48, 2017.

[3]S. S. C. Silva, R. M. P. Silva, R. C. G. Pinto, and R. M. Salles, “Botnets: A survey,” Computer Networks, vol. 57, no. 2, pp. 378–403, 2013.

[4]L. Yang et al. Defense of DDoS attack for cloud computing Computer Science and Automation Engineering (CSAE), 2012 IEEE International Conference on 2012

[5]J. Zhang, Y.-W. Zhang, J.-B. He, and L. Zhou, “A robust and efficient detection model of DDoS attack for cloud services,” in Proc. 15th Int. Conf. Algorithms and Architectures for Parallel Processing (ICA3PP), Zhangjiajie, China, Nov. 18–20, 2015, pp. 611–624.

[6]Q. Yan, F. R. Yu, Q. Gong, and J. Li, “Distributed denial-of-service attacks in software-defined networking with cloud computing,” IEEE Commun. Mag., vol. 53, no. 4, pp. 52–59, Apr. 2015.

[7]Tankard C. Advanced persistent threats and how to monitor and deter them[J]. Network security, 2011, 2011(8): 16-19.

[8]Chen P, Desmet L, Huygens C. A study on advanced persistent threats[C]//Communications and Multimedia Security: 15th IFIP TC 6/TC 11 International Conference, CMS 2014, Aveiro, Portugal, September 25-26, 2014. Proceedings 15. Springer Berlin Heidelberg, 2014: 63-72.

[9]R. C. Staudemeyer, “Applying long short-term memory recurrent neural networks to intrusion detection,” South African Computer Journal, vol. 56, pp. 136–154, 2015.

[10]R. C. Staudemeyer and C. W. Omlin, “Evaluating performance of long short-term memory recurrent neural networks on intrusion detection data,” in Proc. South African Inst. Comput. Sci. Inform. Technol. Conf., New York, NY, USA, 2013, pp. 218–224.

[11]H. Hindy, D. Brosset, E. Bayne, A. Seeam, C. Tachtatzis, R. Atkinson, and X. Bellekens, “A taxonomy and survey of intrusion detection system design techniques, network threats, and datasets,” arXiv, 2018, arXiv:1806.03517.

DOI: 10.54254/2755-2721/2025.TJ23838

[12]Evolutionary Tools. Available online: https://deap.readthedocs.io/en/master/api/tools.html (accessed on 29 April 2020).

[13]Muhuri P S, Chatterjee P, Yuan X, et al. Using a long short-term memory recurrent neural network (LSTM-RNN) to classify network attacks[J].

[14]R. A. Bridges, A. E. Rice, S. Oesch, and J. D. Smith, “Testing SOAR tools in use,” Computers & Security, vol. 129, p. 103201, 2023.

[15]B. A. S. Al-rimy, M. A. Maarof, and S. Z. M. Shaid, “Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions,” Computers & Security, vol. 74, pp. 144–166, 2018.

[16]J. A. Herrera Silva, L. I. Barona López, Á. L. Valdivieso Caraguay, and M. Hernández-Álvarez, “A survey on situational awareness of ransomware attacks—detection and prevention parameters,” Remote Sensing, vol. 11, no. 5, p. 1168, 2019.

[17]F. Mercaldo, V. Nardone, A. Santone, and C. A. Visaggio, “Ransomware steals your phone. Formal methods rescue it,” in Proc. Int. Conf. Formal Techn. Distrib. Objects, Components, and Syst., Heraklion, Crete, Greece, Jun. 6–9, 2016, pp. 212–221.

[18]D. Maiorca, F. Mercaldo, G. Giacinto, C. A. Visaggio, and F. Martinelli, “R-PackDroid: API package-based characterization and detection of mobile ransomware,” in Proc. Symp. Appl. Comput., Marrakech, Morocco, Apr. 3–7, 2017, pp. 1718–1723.

[19]K. Alrawashdeh and C. Purdy, “Ransomware detection using limited precision deep learning structure in FPGA,” in Proc. NAECON 2018 IEEE Nat. Aerospace Electron. Conf., Dayton, OH, USA, Jul. 23–26, 2018, pp. 152–157.

[20]A. Kharaz, S. Arshad, C. Mulliner, W. Robertson, and E. Kirda, “UNVEIL: A large-scale, automated approach to detecting ransomware,” in Proc. 25th USENIX Security Symp. (USENIX Security 16), Austin, TX, USA, Aug. 10–12, 2016, pp. 757–772.

[21]K. Cabaj, P. Gawkowski, K. Grochowski, and D. Osojca, “Network activity analysis of CryptoWall ransomware,” Prz. Elektrotech., vol. 91, pp. 201–204, 2015.

[22]M. Shukla, S. Mondal, and S. Lodha, “Poster: Locally virtualized environment for mitigating ransomware threat,” in Proc. 2016 ACM SIGSAC Conf. Comput. Commun. Secur., New York, NY, USA, Oct. 24–28, 2016, pp. 1784–1786.

[23]G. Cusack, O. Michel, and E. Keller, “Machine learning-based detection of ransomware using SDN,” in Proc. 2018 ACM Int. Workshop Secur. Softw. Defined Networks & Netw. Function Virtualization, Tempe, AZ, USA, Mar. 21, 2018, pp. 1–6.

[24]I. Almomani, R. Qaddoura, M. Habib, S. Alsoghyer, A. Al Khayer, I. Aljarah, and H. Faris, “Android ransomware detection based on a hybrid evolutionary approach in the context of highly imbalanced data,” IEEE Access, vol. 9, pp. 57674–57691, 2021.

[25]S. Tiwari, W. Sarma, and A. Srivastava, “Integrating artificial intelligence with zero trust architecture: Enhancing adaptive security in modern cyber threat landscape,” Int. J. Res. Anal. Rev., vol. 9, pp. 712–728, 2022.