Opanovych Maksym, PhD student

National University ‘Lviv Polytechnic’, Lviv

ORCID:https://orcid.org/0000–0002–2748–2965

Andrian Piskozub, Associate Professor

Lviv Polytechnic National University, Lviv

ORCID:https://orcid.org/0000–0002–3582–2835

Challenges of Using Machine Learning to Detect Cyber Attacks

**Ключові слова: Машинне навчання, кібербезпека, APT**

**Keywords: Machine learning, cybersecurity, APT**

**Introduction**

As cyber threats grow in complexity and sophistication, the reliance on traditional signature-based detection techniques has proven inadequate. Machine learning (ML) methods have emerged as an innovative alternative, leveraging large volumes of data to uncover patterns and predict potential attacks in real-time. ML's adaptability makes it suitable for identifying both known and novel attack vectors across diverse network environments. This adaptability is particularly valuable in a landscape where Advanced Persistent Threats (APTs), fileless malware, and zero-day vulnerabilities continually evolve to bypass traditional defenses. Nonetheless, integrating ML models into cyber defense systems presents distinct challenges. These challenges stem from the inherent complexity of cyber data, adversarial tactics, and the practical deployment and maintenance of ML systems within high-stakes cybersecurity frameworks [1, 2].

This paper explores the primary challenges associated with employing machine learning for cyber attack detection, outlining issues related to data quality and labeling, model performance, real-time constraints, adversarial attacks, and the adaptability of ML-based systems.

**Challenges of Using Machine Learning to Detect Cyber Attacks**

1. Data Quality and Labeling

Machine learning models require high-quality, accurately labeled data to function effectively. In cybersecurity, data typically consists of network traffic logs, system event data, and threat intelligence feeds. However, obtaining labeled datasets is often challenging due to the scarcity of labeled malicious events and the variability of attack signatures. Additionally, cybersecurity data is often imbalanced, with benign events vastly outnumbering malicious ones. This imbalance can cause ML models to develop a bias toward benign classifications, thereby diminishing their ability to detect rare, sophisticated attacks. Annotations by experts can help, but they are costly and time-consuming, leading to potential delays in ML system deployment.

2. Model Interpretability and Complexity

The interpretability of machine learning models is critical in cybersecurity. While traditional methods (such as rule-based systems) provide clear, actionable alerts, ML-based detection models—particularly deep learning models—often function as "black boxes" with complex internal mechanisms. Security analysts need to understand why a model flagged certain events as suspicious to assess its accuracy and validate potential alerts effectively. This complexity can make it difficult for ML systems to integrate into existing security workflows, where clear and interpretable outputs are essential for prompt incident response [3].

3. Adversarial Attacks

One of the most significant challenges in applying ML to cyber defense is vulnerability to adversarial attacks. Adversaries can exploit ML algorithms by subtly manipulating input data, causing models to misclassify malicious activities as benign or evade detection entirely [4]. For example, slight alterations to network packet structures or injecting noise into data patterns can lead to incorrect predictions, allowing attackers to evade detection. Crafting ML models that are resilient against adversarial tactics requires complex adversarial training and constant adaptation to evolving attack methods, which increases the complexity of model maintenance [5].

4. Real-time Detection and Scalability

Cyber defense systems operate in real-time, requiring rapid analysis and response to evolving threats. However, ML algorithms, especially deep learning models, are computationally intensive, which can introduce latency issues. Real-time environments, particularly in large-scale networks, demand high-speed processing and low latency to prevent security breaches. Scaling these models to handle massive data flows, especially under peak load conditions, is challenging and often requires extensive computational resources [6]. Balancing detection accuracy with computational efficiency remains a significant obstacle, as does the need to minimize false positives, which can overwhelm security operations teams.

5. Adaptability and Model Drift

Attack techniques and tactics in cybersecurity are constantly evolving, leading to the phenomenon known as “model drift,” where ML models lose accuracy over time due to shifts in data patterns. An ML model trained on historical data might fail to detect a new, modified variant of a known threat, as it has not been exposed to the updated behavior. Cyber defense systems must, therefore, undergo regular retraining and adaptation to keep pace with the evolving threat landscape. This process is both time-intensive and resource-consuming, with the added complexity of ensuring that retrained models continue to operate effectively within the existing detection infrastructure.

6. Ethical and Privacy Concerns

Cybersecurity data often includes sensitive and personal information, raising ethical and privacy concerns regarding its use in ML models. Ensuring data anonymization while maintaining the dataset's utility for ML training purposes is challenging, as sensitive information might inadvertently be exposed or used improperly. Legal and regulatory frameworks, such as the GDPR, impose strict requirements for data handling and processing, necessitating careful consideration when designing ML-based security solutions

7. Integration with Existing Systems and Legacy Infrastructure

Many organizations rely on legacy systems and infrastructure that were not designed to support modern ML applications. Integrating ML-based cyber defense solutions with these systems requires extensive customization, which can lead to increased costs and complexity. Compatibility issues and data siloing often arise, making it difficult for ML models to access the full scope of data needed for accurate threat detection. Moreover, integrating ML with existing systems can introduce security risks if interoperability is not properly managed, as mismatches between legacy and new systems can create exploitable vulnerabilities.

References

1. Buczak, A. L., & Guven, E. (2016). "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection." IEEE Communications Surveys & Tutorials, 18(2), 1153–1176.
2. Sommer, R., & Paxson, V. (2010). "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection." IEEE Symposium on Security and Privacy (SP), 305–316.
3. Doshi, R., Apthorpe, N., & Feamster, N. (2018). "Machine Learning DDoS Detection for Consumer Internet of Things." IEEE Security and Privacy Workshops (SPW), 29-35.
4. Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B. I. P., & Tygar, J. D. (2011). "Adversarial Machine Learning." Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence, 43-58.
5. Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2016). "Practical Black-Box Attacks against Machine Learning." Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, 506-519.
6. Ahmad, I., Basheri, M., Iqbal, M. J., & Rahim, A. (2018). "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection." IEEE Access, 6, 33789–33795.