Chapter 18
Adversarial Machine Learning in Industrial Cybersecurity: Challenges and Solutions

Binastya Anggara Sekti
https://orcid.org/0000-0001-5489-4888
Universitas Esa Unggul, Indonesia

ABSTRACT

As industrial systems increasingly integrate machine learning (ML) for cybersecurity, they face a growing threat from adversarial machine learning (AML) attacks. AML techniques, such as evasion, poisoning, and model extraction, exploit vulnerabilities in ML models to manipulate security defenses, leading to misclassifications, false positives, and system failures. These threats pose severe risks to industrial environments, including operational disruptions, financial losses, and compromised safety. We will explores the challenges of AML in industrial cybersecurity, focusing on the lack of robustness in ML models, limited availability of high-quality industrial datasets, high computational costs, and the evolving nature of adversarial attacks. Additionally, future research directions, such as secure federated learning and AI-driven attack response mechanisms, are discussed. By strengthening ML-based security frameworks, industrial organizations can enhance resilience against adversarial threats and protect critical infrastructure from evolving cyber risks.

1. INTRODUCTION

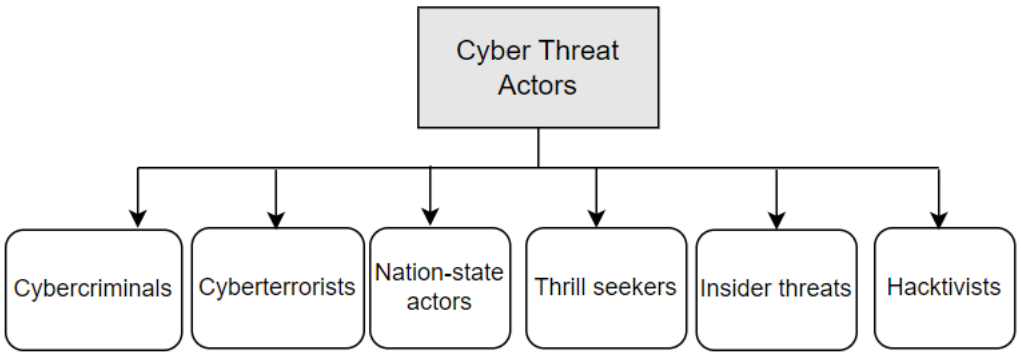
Cybercrime has occurred a lot in the lives of everyone in the world, this crime is starting to be unsettling because it can be disturbing to harm in large numbers. This then creates a security system that can ward off cyber attacks. Today's cybercrime continues to get more sophisticated so traditional intrusion detection systems (IDS) often face challenges in identifying unknown attacks and tend to have high rates of false positives (Dhanvijay, 2025). In handling various cyber attacks that are increasingly sophisticated and massive, cyber security technology is also constantly improving its capabilities using AI, machine learning, deep learning, and others (Ortiz-Ruiz, 2024). In general, cybercriminal actors continue to attempt to exploit vulnerabilities in web systems that aim to spread malware, carry out phishing attacks, steal sensitive information, and commit various forms of cybercrime. Traditional signature-based methods of

DOI: 10.4018/979-8-3373-3241-3.ch018

detecting malicious web pages often struggle to keep pace with the rapid evolution of malware and cyber threats. As a result, there is a growing demand for more sophisticated and proactive approaches that can effectively identify malicious web content based on its characteristics and behavior (Liaquathali, 2025).

Cybersecurity threats can be said to be all kinds of potential dangers that can disrupt, damage, or exploit data, information systems, and computer network systems. Threats can come from a variety of sources, such as individual hackers, a group of cybercriminals, a criminal organization in a country, or internal user errors. Cybersecurity threats include various attacks such as phishing, malware, denial-of-service (DoS) attacks, software vulnerability exploits, and attacks on artificial intelligence (AI)-based systems. A cybersecurity threat is any type or action that aims to access, damage, or disrupt an information system in an unauthorized way, carried out by means of exploiting technological weaknesses or user manipulation. Increasingly dependent on information technology, food will face increasingly complex and sophisticated threats from cybercriminals. (Darem, 2023; Razaque, 2021). Figure 1 is a set of cyber threat actors.

Figure 1. Cyber threat actors



The rapid growth of the Internet of Things (IoT) and its widespread use in many regions, such as smart homes, smart communities, and vehicles, has made IoT security even more important. Ransomware is an advanced and customizable threat that affects users globally, restricting access to their data or systems through models such as file encryption or screen locking. Traditional ransomware detection methods are losing the ability to successfully combat these threats (Alzakari, 2025).

Cybersecurity threats can also be defined as a condition that can have a negative impact on information systems, stemming from cybercrime, technical glitches, or human error. (Malatji, 2022). Figure 2 shows the types of cyber attacks.