Role of AI in Threat Detection and Zero-day Attacks

Kelly Morgan
Old Dominion University

Role of AI in Threat Detection and Zero-day Attacks

Kelly Morgan

Old Dominion University

COVA CCI Undergraduate Research Program

April 12, 2023

Abstract

Cybercrime and attack methods have been steadily increasing since the 2019 pandemic. In the years following 2019, the number of victims and attacks per hour rapidly increased as businesses and organizations transitioned to digital environments for business continuity amidst lockdowns. In most scenarios cybercriminals continued to use conventional attack methods and known vulnerabilities that would cause minimal damage to an organization with a robust cyber security posture. However, zero-day exploits have skyrocketed across all industries with an increasingly growing technological landscape encompassing internet of things (IoT), cloud hosting, and more advanced mobile technologies. Reports by Mandiant Threat Intelligence (2022) concluded that 2021 had the largest increase in zero-days accounting for at least 80% that had been exploited. State-sponsored actors led by Chinese groups were the primary attackers. Traditional methods of defense, which include antivirus software, patching, firewalls, and other cybersecurity controls are less effective against zero-days, which are unknown to vendors and organizations. Zero-days bypass the traditional signature and anomaly-based detections and antivirus software, which contain signatures information for known attacks. To deal with a changing and advanced threat landscape, techniques incorporating artificial intelligence such as machine learning and deep learning along with IDS have been implicated in detecting and preventing zero-day attacks.

AI Background

The concept of Artificial Intelligence, AI, first appeared around the 1940s but largely remained unobtainable until the rise of Big Data and advances in technology with greater processing power. AI is defined as the simulation of humanlike properties, namely intelligence, by a machine, which includes hardware and software. Businesses, governments, and people around the world apply AI capabilities in diverse areas including automobiles, computer applications, agriculture, medicine, and cybersecurity ranges and defense methods. There are several subsets of AI, which include machine learning, deep learning, expert systems, and neural networks, and different ways of honing AI, which include supervised and unsupervised learning. While a variety of AI based modalities have wide ranging implications in cybersecurity techniques, this paper will focus on machine learning, which includes deep learning.

Machine learning, ML, is a type of AI that enables systems to learn and evolve without explicit programming. ML algorithms enable systems to leverage mostly labeled structured data for building mathematical models that are used for making predictions or decisions without the need for intervention. Deep learning, DL, is a subset of ML that uses artificial neural networks comprised of several layers, which mimics human thinking, to process unstructured data, build more complex models, and requires less intervention than ML.

ML and DL algorithms can use diverse types of learning to train algorithms. This paper focuses on two types of learning, supervised and unsupervised. In supervised learning, labelled data sets are utilized for training models, which in turn is used for training subsequent data sets and outcome prediction. This method involves human intervention during the training phase and is useful for predicting outcomes and solving real world problems. During unsupervised learning, models are trained using raw, unlabeled data. Models process large amounts of data without human intervention and are effective for clustering large data sets and detecting anomalies.

A significant hindrance to the widespread adoption of AI-based interventions is the explainability of ML and DL. Due to the large amount of data stored and used by models for threat countermeasures serious implication exist if these systems are exploited, which may involve data poisoning, where the attacker alters the data used for learning, and inherent system vulnerabilities. Another challenge is the obscurity of AI systems, and the threat of advanced attacks in AI systems that are still inadequately understood. However, using ML and DL as part of a comprehensive detection and defense system with a symbiotic approach, involving human oversight, is effective for analyzing large data sets, establishing baselines, and detecting anomalous behavior without relying on known attack vectors or signatures.

Threat Landscape

Ponemon Institute (2020) estimated that 80% of successful incidents involving compromised data in 2019 were the result of zero-day attacks. Zero-day attacks are exploits of unknown vulnerabilities, or flaws in software and hardware. These vulnerabilities exist in the wild and are unknown to the vendors or the cybersecurity community. Therefore, traditional antivirus software, patching, or similar techniques that rely on malware signatures for threat detection are ineffective. Hackers and state-sponsored groups will leverage zero-days either by purchasing from other hackers or by selling them, which can reach prices as high as $1 million. Zero-day attacks can come in the form of polymorphic worms, viruses, trojans, and other types of malwares. Exploitation methods involving zero-days include phishing and spamming using malicious emails, embedding exploits in compromised sites or browsers, and software or hardware vulnerability hunting. Notable zero-day attacks include Stuxnet, a computer worm that targeted Iranian SCADA systems to sabotage their nuclear weapons program. The Sony zero-day attacks, which released vast amounts of the company's sensitive data to file-sharing sites. RSA attack, which exploited Adobe Flash zero-day using malicious phishing email attachments. The DNC hack, which involved the exploitation of a DNC server and release of confidential information. The challenge in dealing with zero-days is their ability to bypass signature-based detection methods, as there is no signature available for it yet. Another challenge is that zero-day attacks involve a diverse range of technologies with a multi-step approach involving methods beyond detection, such as social engineering and pursuing specific targets.

Proposed Strategy

To detect known and novel forms of malware an AI-based framework combining supervised and unsupervised machine learning alongside traditional defense methods is proposed. This framework leverages the precision of supervised classification for detecting known classes of malware and the flexibility of unsupervised learning for detecting new classes of threats.

The two most widely used techniques for malware detection can be categorized as signature-based and anomaly-based detection. Signature-based detection utilizes predefined signatures of known malware and attacks to inspect programs and activities for malicious code or patterns. A major drawback to using the signature-based approach lies in its failure to detect zero-days for which it has no signature until an exploit has occurred. Another problem is its failure to detect evolving threats, such as polymorphic and metamorphic malwares, that reprogram themselves after distribution and propagation and use encryption to evade detection. Anomaly-based detection operates on the established baseline of normal or good behavior and activities and utilizes this knowledge to detect patterns of deviations or anomalous activity.

While this technique is effective in detecting zero-day attacks, its susceptibility to a high false alarm rate makes it challenging to use in certain environments.

With supervised machine learning, the large volumes of data generated across a system or network can be rapidly analyzed, which is a major problem faced by human analysts in the field. ML models can train and learn from their own evolving sample sets, which include labeled datasets. ML is capable of automating tasks that fall into the categories of threat detection, response, and classifying new threat patterns, with malware classification being one of the most common applications of ML classifiers. Supervised ML methods improve the accuracy of threat detection overall due to the breadth of computational analysis it can manage. Despite the benefits of Supervised ML, challenges exist when detecting novel malware and building precise classification models.

Unsupervised learning utilizes unlabeled data sets, without the need for human intervention, to uncover unknown relationships, trends, and locate key patterns within data. Unsupervised learning is capable of making predictions beyond that which is known by analysts or supervised learning models, to reveal previously unknown insights and relationships. Data provided to this model is used to make vast amounts of probability-based calculations, which makes it effective in detecting new threat patterns and malware that does not rely solely on past threat information. From this data, unsupervised learning models can establish normal behaviors, deviations, unseen threats or zero-days that other detection methods would have missed.

Summary

The threat landscape has grown rapidly since the pandemic as more technologies are introduced and businesses continue to go digital. With this change in how business is conducted, new attack vectors are growing quickly. Zero-days present a significant challenge for cybersecurity professionals, vendors, and businesses because they are unknown to them. Zero-days are often exploited by higher-level hackers with financial and political goals. Legacy techniques for detecting zero-day attacks are not as effective, because these attacks have yet to be documented with methods of handling them. Supervised and Unsupervised ML along with current practices like IDS, and antivirus software are a more effective method of detecting zero-day attacks.

References

SADOWSKI, J. (2022, April 21). *Zero tolerance: More zero-days exploited in 2021 than ever before*. Mandiant. Retrieved April 12, 2023, from https://www.mandiant.com/resources/blog/zero-days-exploited-2021#:~:text=Zero%2DDay%20Exploitation%20Reaches%20All%2DTime%20High%20in%202021&text=By%20the%20end%20of%202021,record%20of%2032%20in%202019

Ahmad, R., Alsmadi, I., Alhamdani, W., & Tawalbeh, L. (2023, February 27). *Zero-Day attack detection: A systematic literature review*. SpringerLink. Retrieved April 12, 2023, from https://link.springer.com/article/10.1007/s10462-023-10437-z

Ali, S., Rehman, S. U., Imran, A., Adeem, G., Iqbal, Z., & Kim, K.-I. (2022, November 28). *Comparative evaluation of AI-based techniques for Zero-day attacks detection*. MDPI. Retrieved April 12, 2023, from https://www.mdpi.com/2079-9292/11/23/3934

Ananth, A. N. (2022, August 9). *Human threat hunters are essential to thwarting zero-day attacks*. Dark Reading. Retrieved April 12, 2023, from https://www.darkreading.com/attacks-breaches/human-threat-hunters-are-essential-to-thwarting-zero-day-attacks

Guo, Y. (2022, November 25). *A review of machine learning-based zero-day attack detection: Challenges and future directions*. Computer Communications. Retrieved April 12, 2023, from https://www.sciencedirect.com/science/article/abs/pii/S0140366422004248?via%3Dihub

Parrend, P., Navarro, J., Guigou, F., Deruyver, A., & Collet, P. (2018, April 24). *Foundations and applications of Artificial Intelligence for Zero-day and multi-step Attack Detection*. SpringerLink. Retrieved April 12, 2023, from https://link.springer.com/article/10.1186/s13635-018-0074-y

Zero-day (0day) exploit. imperva. (n.d.). Retrieved April 12, 2023, from https://www.imperva.com/learn/application-security/zero-day-exploit/