JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ENHANCING CYBER SECURITY : A COMPREHENSIVE SURVEY OF MALWARE DETECTION TECHNIQUES

Mr. Trivedi Vishal Rajendrakumar
Research Scholar,
Department of Computer Science & Engineering
(Madhav University, Pindwara, Sirohi, Rajasthan)

Dr. Bhawesh Kumawat
Assistant Professor,
Department of Computer Science & Engineering
(Madhav University, Pindwara, Sirohi, Rajasthan)

Abstract : Network safety stays a basic worry in the computerized age, with malware representing a tenacious danger to the honesty and classification of data frameworks. Traditional approaches as well as cutting-edge technologies are covered in great detail in this abstract, which provides a comprehensive overview of current malware detection methods. The review starts by looking at signature-based detection, which has been around for a long time and relies on predefined patterns of known malware. It addresses the limitations of signature-based methods in the face of polymorphic and metamorphic variants of malware and discusses their efficacy. Consequently, the record investigates heuristic and behavior-based location components, clarifying how these procedures break down the way of behaving of possibly pernicious code to recognize dangers. Machine learning (ML) and artificial intelligence (AI) have received a lot of attention recently as effective malware detection tools. This abstract delves into various machine learning and artificial intelligence models, including techniques for supervised and unsupervised learning. It looks at these models' strengths and weaknesses, taking into account things like their ability to adapt to new threats and their ability to be interpreted. The survey also starts with a look at signature-based detection, focusing on its history and the problems that polymorphic and metamorphic malware have caused in the past. The role of heuristic and behavior-based analysis techniques in identifying malicious code based on observed patterns and behaviors is then discussed. The foundational method of signature-based detection, which relies on well-known malware patterns, is the focus of the investigation. This method works well against common threats, but it has problems with polymorphic and metamorphic malware variants, so it's important to know how it works in the current threat landscape. The subsequent analysis places an emphasis on the adaptive nature of heuristic and behavior-based detection methods in spotting potential threats based on observed patterns and behaviors. A dynamic approach to malware detection is provided by these techniques, which go beyond the limitations imposed by static signatures. The job of Threat Intelligence (TI) and sandboxing technologists analyzed as essential parts in proactive protection methodologies. Danger insight takes care of give continuous information on arising dangers, empowering fast reaction and moderation. In contrast, sandboxing creates isolated environments for suspicious file analysis, making it easier to identify malware behavior without jeopardizing the host system's integrity. The abstract delves into the paradigm shift in malware detection brought about by Machine Learning (ML) and AI. The effectiveness and adaptability of various machine learning (ML) models, such as supervised learning, unsupervised learning, and deep learning, are examined in light of changing cyber threats. The abstract acknowledges persistent difficulties in the field in spite of these advancements. Threat actors are looking to take advantage of vulnerabilities and alter detection mechanisms, so adversarial attacks on ML models represent a significant threat. The paramount ethical considerations are user privacy and data security, necessitating a delicate balance between effective detection and individual rights respect.

KEYWORDS

Cyber security, signature-based detection, heuristic analysis, behavior-based detection, machine learning, and artificial intelligence are all examples of malware detection.

INTRODUCTION

The Covid pandemic has emphatically expanded overall digital assaults, a phenomenon that has progressively been named the 'digital pandemic' [1] and is supposed to arrive at USD 10.5 trillion in yearly harm costs by 2025 [2]. In addition, the pandemic's new business model of "working from home" has significantly increased the threat exposure of most organizations [3]. Worse, prior to the pandemic,

an estimated 20% of cyber attackers used malware or attack methods that had never been seen before. Many of these attackers used machine learning models that adapt to their surroundings to go undetected. Throughout the pandemic, this percentage increased to 35% [4]. Cybercrime is expected to grow at an unprecedented rate over the next decade, affecting a significant portion of global businesses across all sectors. These trends, taken together, point in the wrong direction. In light of the rapid expansion of technology, data, and computing power, it is essential to employ more advanced tools in order to address contemporary issues. Humans cannot keep up with the increasing complexity on their own, so they must rely on AI. The market for artificial intelligence (AI) in cyber security is expected to grow from USD 3.92 billion in 2017 to USD 34.81 billion by 2025 [5]. In addition, according to a Capgemini Research Institute survey, 69% of businesses believe that AI is required to respond to cyber-attacks [6]. The public and private sectors are paying more attention to AI than ever before. In any case, its power will definitely fall into the hands of cybercriminals, making the up-and-coming age of man-made intelligence-controlled malware. Security professionals need to be able to quickly and accurately identify the types of malwares that have been detected due to the unprecedented rise in cybercriminal capabilities powered by AI. However, despite significant advancements in AI-driven malware detection techniques, the current rate of progress is insufficient, necessitating increased efforts to keep up with cybercriminals. There are many kinds of malware, which is short for "software," like viruses, worms, spyware, Trojans, ransom ware, and so on. but almost always aims to steal data and systems or hold a victim hostage for ransom. Traditionally, malware detection relied solely on comparing the signatures of known malware stored in a database to the continuous byte sequences of a suspected malware file. Signature-based detection became less effective over time as newer, "polymorphic" malware emerged, and it was eventually replaced by next-generation, heuristic-based, behavioral-based, and machine learning-based detection methods [7]. Machine learning algorithms underpin all of the leading malware detection solutions, also known as Endpoint Detection and Response, or EDR [8]. This literature audit aims to present and investigate the most recent efforts in creating novel, more effective ways to utilize man-made consciousness for malware location, fully intent on giving comprehensive guidance to ensuing examination. As a result, researchers may be able to gain a better understanding of malware detection, as well as the new developments and research directions being considered by the scientific community to address this challenging problem. Techniques for detecting malware based on AI are the subject of numerous papers. However, due to the fact that this study focuses on the most recent developments in AI-based malware detection, articles published prior to 2016 will not be included in the survey's scope. Methods for analyzing malware: The development of efficient methods for malware detection relies heavily on malware analysis, which serves as the foundation of malware detection. Malware detection would be impossible without malware analysis, which provides insight into the malicious file's classification and functionality. Static, dynamic, and hybrid approaches to malware analysis are described below [9]: Static analysis, which uses extracted low-level information like system calls, the control flow graph, and the dataflow graph to inspect and analyze a suspected malicious file without actually running it. There are few false positives with static analysis; Code obfuscation-based unknown malware, on the other hand, is missed by it. Dynamic analysis, in which a suspected malicious file is examined at runtime, typically within a sandbox (a separate virtual machine used for testing in which malware can be executed without affecting system resources). The benefit is that the malware can be executed and dissected. As a result, it is possible to successfully identify unknown malware. Nonetheless, unique examination is tedious and produces an elevated degree of false positives.

• Hybrid analysis, in which the difficulties of both static and dynamic analysis are combined to overcome one another. In order to explain how cutting-edge machine learning-based malware detection systems are built, this paper will refer to these three types of analysis. 3. Malware Detection Techniques There are three broad categories of malware detection techniques, as previously mentioned: based on behavior, heuristics, and signatures. These methods are based on malware analysis results, and each method has its own advantages and disadvantages [9]. Signature-based detection makes use of a well-established list of indicators of compromise (IOCs), such as particular byte sequences, API calls, file hashes, malicious domains, or network attack patterns. However, signature-based detection does not require machine learning models and cannot identify previously unknown or encrypted malware.

• Behavioral-based detection involves collecting all exhibited behaviors from a suspected executable file and monitoring it in an isolated environment. After that, methods of extracting useful features are used to help a machine learning model classify the malicious behavior. Heuristic-based discovery - this technique relies on creating rules in light of the aftereffects of the static/unique examination to direct the review of the removed information to help the proposed malware identification model. Such principles can either be generated manually (depending on the ability of the security experts) or naturally, utilizing AI or devices like YARA. Feature Extraction and Selection Throughout this paper, various machine learning algorithms will be referred to as "features." The process of converting raw data into numerical features that the machine learning algorithm can "understand" is known as feature extraction or feature engineering in machine learning. Because applying machine learning directly to raw data is typically ineffective, feature extraction is required to enhance the model's effectiveness. Then again, include choice is the process of eliminating superfluous elements to help with fostering a prescient model. Classification Methods Based on Shallow Learning for Malware Detection The majority of machine learning models proposed prior to 2006, specifically those that are not classified as deep learning, fall under the category of shallow learning (SL). SL approaches customarily rely vigorously upon highlights physically intended to settle a given errand. However, shallow learning algorithms are still widely used, particularly in malware detection and cyber security as a whole. Fig. The most widely used SL-based classification techniques are listed in Table 1 below. Fig. 1 Common algorithms for shallow learning.

I collected all the information from the different types of books and online. I inspired through the content during the research about malware. During the survey of malware detection, I Found that 20% of cyber attackers used malware and throughout the pandemic, this percentage increased to 35%. By collecting all this information, I will publish my all books and papers based on this collecting information in future.

INFERENCE

In this section, the conclusions derived from this extensive evaluation are presented.

• Malware disclosure methodologies with 100% capability concerning precision, TPR, FPR, exactness, and survey are a terrible dream for engineers in light of the fact that trendy malware is coded in a jumbled manner and undeniable level evasion techniques are used against conspicuous security frameworks like firewalls and unfriendly to diseases.

• Malware that is integrated with confusion methods is beyond the scope of static investigation techniques. Yes, it does.

CONCLUSION

One important example that security provides is being proactive rather than receptive. Developing a framework for malware discovery is currently challenging; especially when dealing with cutting-edge malware. New generations of malware have emerged thanks to high-level avoidance systems, which have had extremely significant effects. Nevertheless, the malware discovery innovation based on profound learning minimizes the drawbacks of conventional and traditional methods. This paper presents a deliberate overview of malware revelation.

REFERENCES

[1] M. Fichtenkamm, G. Burch, J. Burch, "Cyber security in a COVID-19 World: Insights on How Decisions Are Made."

[2] S. Morgan, "Cybercrime to Cost the World $10.5 Trillion Annually by 2025."

[3] Deloitte, "Cyber-crime – the risks of working from home." deloitte.com

[4] C. Nabe, "Impact of COVID-19 on Cyber security." deloitte.com

[5] Markets and Markets, "AI in Cyber security Market by Offering (Hardware, Software, Service), Technology (Machine Learning, Context Awareness, NLP), Deployment Type, Security Type, Security Solution, End-user, and Geography – Global Forecast to 2025."

[6] G. Belani, "The Use of Artificial Intelligence in Cyber security: A Review." computer.org

[7] A. Ray, Cyber security for Connected Medical Devices, United States of America: Academic Press, 2021, pp. 217-262.

[8] Forrester, "The Forrester Wave: Endpoint Detection and Response Providers, Q2 2022."

[9] F. A. Aboaoja, A. Zainal, F. A. Ghaleb, B. A. S. Al-rimy, T. A. E. Eisa, and A. A. H. Elnour, "Malware Detection Issues, Challenges, and Future Directions: A Survey," Applied Sciences, vol. 12, no. 17, p. 8482, 2022.

[10] Y. Xu, Y. Zhou, P. Sekula, and L. Ding, "Machine learning in construction: From shallow to deep learning," Developments in the built environment, vol. 6, p. 100045, 2021.