DOI 10.7251/ZRSNG2525051C

A Systematic Review and Comparative Analysis of Malware Detection Techniques

Aleksandar Sandro Cvetković [1*], Saša Adamović [1], Marko Šarac [1]
[1] Faculty of Computing and Informatics, Sinergija University, Bijeljina, Bosnia and Herzegovina
[*]ascvetkovic@sinergija.edu.ba

Abstract – The increasing sophistication and volume of malware pose a persistent and evolving threat to cybersecurity. The research paper systematically examines and compares various malware detection techniques, including traditional methods such as signature, heuristic, and anomaly detection, and more advanced methods such as behavior analysis, sandboxing, and machine learning, including deep learning. The study examined the mechanisms, advantages, and disadvantages of each technique using recent empirical data from academic literature. A comprehensive table provides a parallel comparison of these methods based on key performance indicators, efficacy against evasive malware, and resource consumption. In addition, it discusses the current challenges of malware detection, such as the increasing complexity of malware, evasion tactics, and threats to machine learning models. Finally, it explores emerging trends and future directions in this field, including integration of artificial intelligence, cloud analysis, proactive defense mechanisms, and the growing role of large language models. This review underscores the need to continuously innovate and adapt malware detection strategies to effectively counter the evolving landscape of cyber threats**.**

Keywords – Malware Detection Techniques, Malware Analysis, Machine Learning, Deep Learning, Evasion Techniques.

Introduction

Digital landscapes are faced with constant attacks from malicious software, commonly known as malware. Due to the increasing size and complexity of these threats, robust and adaptive detection mechanisms must be used to protect individuals, organizations, and key infrastructures. Recent statistics have highlighted the gravity of this problem. In 2024, an astonishing 76% of companies reported being affected by malware attacks, with an average of 560,000 new malware cases per day detected, contributing to the already huge database of more than one billion existing malware programs [1]. This constant proliferation highlights the urgent need for effective cybersecurity strategies, where malware detection is the cornerstone. Furthermore, threats are not static, malware is becoming more and more sophisticated, and various technologies are being used to avoid traditional security measures [2]. The emergence of platform-specific threats, such as Mirai viruses that have affected Internet of Things (IoT) devices since 2021, further complicates the landscape, requiring not only effective but also multifunctional detection technologies across different computer environments [3].

Given the dynamic nature of malware and the ongoing advances in cyberattack methodology, an in-depth understanding of existing detection techniques is essential. While many studies have focused on malware detection on specific platforms and specific techniques, a holistic and comparative overview remains essential [1]. Existing literature sometimes lacks a unified view of cross-platform analysis and the complex relationship between analytical methods and the data types they use. This research seeks to eliminate these gaps by providing a systematic review and detailed comparative analysis of a wide range of malware detection techniques. This paper aims to provide valuable information to researchers and practitioners in the field of cyber security by studying their strengths, weaknesses, and performance characteristics.

The purpose of this study is multifaceted. First, it aims to provide a comprehensive review of the current state-of-the-art malware detection technology. Second, the objective is to compare the effectiveness of these techniques based on the empirical data reported in recent studies. Third, investigating the existing challenges that prevent the effective detection of malware. Finally, the paper will explore the emerging trends and future directions in this critical field. To achieve these objectives, the research paper is organized as follows: Chapter I provides an overview of the different types and techniques of malware detection. Chapter II presents a comparative analysis of these techniques in a table format. Chapter III discusses current challenges and future directions in malware detection. Finally, the paper provides a final summary of the research.

I. MALWARE DETECTION TECHNIQUES

Malware detection has evolved dramatically over time and has produced a variety of techniques, each with its approach to detecting malicious software. These techniques can be broadly classified into several important areas.

Signature-Based Detection

The foundational technique works by identifying malware based on unique identifiers associated with known threats, commonly referred to as signatures [4]. These signatures may be specific byte sequences, malicious file hash values, or other characteristics. Signature-based detection is recognized for its speed and accuracy in identifying known malware and usually has a low false positive rate [5]. If the signature of a file matches the entry of the known malware signature database, the file is likely malicious. However, this method has important limitations. It is ineffective against unknown variants and viruses mutated by the virus, such as polymorphism and metamorphism, because their signatures differ from those

stored in the database [5]. Polymorphic malware, for instance, changes its code structure with each infection, while metamorphic malware rewrites its code entirely [6]. The constant emergence of new malware necessitates frequent updates to the signature database, which poses a considerable challenge in maintaining its effectiveness against the latest threats [5]. While signature-based detection remains a rapid and efficient method for identifying established threats, its reactive nature and vulnerability to evolving malware diminish its efficacy as a standalone solution in the current threat landscape.

Heuristic-Based Detection

To address the limitations of signature-based methods, heuristic detection uses artificial intelligence to analyze program behavior and identify suspicious code patterns that can indicate malicious intentions [4]. Heuristic analysis does not rely on pre-defined signatures but instead uses algorithms that detect behavior and characteristics associated with malware that can change system files, access sensitive data, or establish unauthorized network connections [7]. Heuristic analysis is advantageous in detecting new and unknown malware, including polymorphic and metamorphic variants, by recognizing malicious behaviors even when the specific code signature is not known [4]. However, a significant drawback of heuristic-based detection is its tendency to generate a higher rate of false positives [5]. Legitimate programs may sometimes exhibit behaviors that are similar to those of malware, leading to incorrect classifications and potential disruptions [7]. Balancing the sensitivity of heuristic rules to detect novel threats while minimizing false alarms remains a critical challenge in this approach.

Anomaly-Based Detection

This technique helps find harmful activities on systems or networks by first learning what normal behavior looks like. Once it understands normal activity, it can spot anything unusual that might be harmful [8]. This method is useful because it doesn't need to know about specific malware signatures beforehand. It can detect new and unknown threats, even ones called zero-day attacks [8]. This detection approach uses smart methods such as heuristics, artificial intelligence, and data mining to define what normal activity is [8]. While it is great at discovering new threats, it often triggers too many false alarms [8]. This happens when regular but odd activities are mistakenly flagged as threats, leading to unnecessary concerns and checks [9]. Another challenge is determining what normal behavior is, especially in systems that are complex and always changing. The success of this method relies heavily on the quality and completeness of the data used to establish what normal behavior is [8]. Continually learning and adjusting what is considered normal is important to reduce false alarms and make the method more accurate over time.

Behavior Analysis and Sandboxing

Behavior analysis means watching a program's actions in a fake, safe setup. This helps to see if the program is harmful [4]. Sandboxing is a method used for testing. In sandboxing, a file that looks dangerous is run in a separate space that copies a real computer system [10]. Experts use this to watch how malware behaves with files, the registry, the internet, and other programs. This helps them understand how it works and what damage it might do [10]. Sandboxing is good for finding new and tricky malware that can fool regular detection methods [10]. It allows experts to study the malware closely without risking harm to the real computer system [10]. But some smart malware can tell when it's in a sandbox and change its actions to hide better. Also, running many such checks can take a lot of computer power and time [10]. The challenge remains to make sure the sandbox acts like a real system and can catch smart tricks used by malware.

Machine Learning-Based Detection

Machine learning has significantly changed how we detect malware, allowing us to analyze large amounts of data and identify complex patterns that might indicate harmful behavior [1]. Various machine learning methods, like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees (DT), and ensemble methods such as Random Forest (RF) and gradient boosting, are used to decide if the software is safe or harmful based on specific features [1]. These features can be derived from static analysis (e.g., file headers, opcode sequences, metadata), dynamic analysis (e.g., system calls, API invocations, network activities), or memory analysis (e.g., memory allocation patterns) [1]. Machine learning can find hidden malware and handle the complexity and size of new threats well [11]. However, the models' effectiveness depends a lot on the quality and amount of training data, and they can be fooled by attackers creating inputs to escape detection [12]. Additionally, as malware evolves, we face the challenge of concept drift, which means malware characteristics change over time, requiring frequent model retraining and updates [13].

Deep learning-Based Detection

Deep learning is a strong method for detecting malware and is part of machine learning. It is very effective when looking at large and complicated data sets [3]. This approach uses different models, like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformers. These models have many layers of artificial neurons that automatically learn from raw data. This means we don't have to select features manually [1]. Deep learning models often perform better than traditional machine learning methods for detecting malware [3]. CNNs are excellent at examining binary code to find harmful patterns. On the other hand, RNNs and LSTMs are good at analyzing sequences, such as system calls and API calls, to find unusual behavior [3]. However, these models have some drawbacks. They need a lot of computing power for training. Also, the decisions they make can be difficult to understand because the models are complex [12]. Like older machine learning models, deep learning models can be attacked and tricked by adversarial techniques. They also need regular updates to stay effective against new types of malware [12].

II. COMPARISON OF MALWARE DETECTION TECHNIQUES

The following Table 1. provides a comparative analysis of the malware detection techniques discussed in the previous section. These include signature-based, heuristic-based,

anomaly-based, behavior analysis/sandboxing, machine learning-based, and deep learning-based techniques. The comparison focuses on key aspects like how each technique works to find malware, their main advantages and disadvantages, their performance regarding accuracy and false positive rate (FPR), how well they handle malware that tries to avoid detection, and the amount of computer resources each method uses.

Table 1. Comparative analysis of the malware detection techniques

Detection Technique	**Detection Mechanism**	**Advantages**	**Disadvantages**	**Typical Performance Metrics (Accuracy/FPR)**	**Efficacy Against Evasive Malware**	**Resource Consumption**
Signature-Based	Matches unique identifiers of known malware against a database.	Fast, efficient for known threats, low false positives.	Ineffective against unknown and polymorphic malware, requires frequent updates.	High accuracy for known malware, 0% FPR (ideally) [5]	Low	Low
Heuristic-Based	Analyzes program behavior and code patterns for suspicious characteristics using AI.	Detects unknown and polymorphic malware.	High false positive rate.	Accuracy varies, high FPR [5]	Medium	Medium
Anomaly-Based	Establishes a baseline of normal behavior and flags deviations.	Detects new and unknown threats (zero-day).	High false positive rate, requires accurate baseline, high resource consumption.	Accuracy varies, high FPR [8]	Medium	High
Behavior Analysis/Sandboxing	Executes malware in an isolated environment to observe runtime behavior.	Detects zero-day threats, provides detailed examination of malware functionality.	Can be evaded by sophisticated malware, resource-intensive, time-consuming.	Accuracy depends on the analysis and environment [14]	High	High
Machine Learning-Based	Uses various ML algorithms to classify software based on extracted features.	Detects obfuscated malware, handles complexity and scale, adapts to new threats.	Performance depends on data quality, susceptible to adversarial attacks, concept drift.	Accuracy varies widely [2], 98% [15], FPR varies	Medium to High	Medium to High
Deep Learning-Based	Employs deep neural networks for automatic feature extraction and pattern recognition.	High accuracy, automatic feature learning, handles complex data.	High computational cost, lack of interpretability, susceptible to adversarial attacks.	Accuracy often > 95% [16], DL Model: 98% [17], FPR varies	High	High

The Table 1. illustrates a balance between the complexity of malware detection methods and their effectiveness in finding new and elusive malware.

Basic methods like Signature-Based detection are very good at detecting known threats with few errors, but they miss new types of malware. This makes them important but insufficient on their own in today's cybersecurity.

More sophisticated methods like Heuristic-Based and Anomaly-Based detection aim to find unknown threats but often have more false positive rates, which can lead to too many alerts and the risk of ignoring real threats.

Behavior Analysis, also known as Sandboxing, is a powerful way to detect unknown malware and understand how it acts, but it requires a lot of resources and can be evaded by advanced malware.

Newer techniques such as Machine Learning and Deep Learning-Based detection are at the forefront of malware detection. They promise high accuracy for both known and unknown, especially stealthy, threats. However, these methods require large, high-quality data sets, and substantial computing power, and are vulnerable to tricks designed to fool them. Moreover, it's hard to understand precisely why a deep learning model labels something as a threat because it functions as a "black box."

The Table 1. supports the idea that a multi-layered approach to malware detection is the best strategy. Relying on only one method could leave systems vulnerable. A mix of signature-based detection to efficiently handle known threats, combined with more advanced techniques like behavior analysis and machine learning, offers complete protection against new and elusive malware.

Deciding which techniques to emphasize depends on specific factors like the sensitivity of the data, available resources, and how many false positives are acceptable. As malware continues to evolve, working on and improving machine learning and deep learning detection methods will be crucial in the ongoing battle against cyber threats.

III. Challenges and Future Directions

The world of malware detection is always evolving because cybercriminals are becoming smarter with their tactics. There are many important challenges that require ongoing research and creative solutions.

One big issue is how advanced malware has become. New types of malware use various tricks to avoid detection. For instance, polymorphic malware changes its code every time it spreads, making traditional detection methods less effective [6]. Metamorphic malware goes even further by completely rewriting its code each time [18]. Fileless malware is another threat, it operates in a computer’s memory without leaving traces on the hard drive, bypassing traditional scanning methods [12]. These problems require new ways to detect malware, focusing more on behavior and memory than just code.

Malware authors use clever tactics to dodge security systems. They might hide the true nature of malware using code packing and encryption [19]. Additionally, advanced malware can trick sandbox tests (protective testing environments) by pretending to be harmless during analysis [10]. Techniques like changing the code of legitimate software and using anti-debugging methods make it even harder for experts to detect and understand malware [20].

Machine learning is another tool in malware detection, but it has its own challenges. These systems can be fooled by special inputs, making them incorrectly identify harmful software as safe [12]. It’s essential to strengthen these models against such attacks. Another problem is the imbalance in data, there are usually more safe samples than harmful ones, which affects the performance of deep learning models [12]. Moreover, as malware evolves, earlier trained models may become less effective due to changes in malware patterns, known as concept drift [13].

Looking ahead, there are promising areas for enhancing malware detection. AI detection is expected to play a bigger role. By using machine learning and deep learning, these systems can process large amounts of data, find subtle patterns, and quickly adapt to new threats [21]. Future research will likely aim to develop smarter AI, improve accuracy, and enhance defense against attacks while being able to spot new malware types [21].

Cloud-based analysis provides the power and scalability needed to process big data involved in malware detection [22]. Research might focus on building better cloud platforms to enable faster and more detailed threat information gathering [22].

There's also a focus on prevention strategies to stop attacks before they can cause harm [22]. This includes using threat intelligence in security systems, strong vulnerability management, and innovative approaches like Automated Moving Target Defense (AMTD) [23].

Lastly, Large Language Models (LLMs) are being explored in this field. They can analyze written data like audit logs and threat reports to find harmful patterns [1]. They can also be used to create fake examples for testing detection models and studying malware-hiding techniques, opening up new possibilities for research [24].

Conclusion

The ongoing battle against malware requires continuous development of detection techniques. Traditional methods such as signature-based detection continue to play a role, but their effectiveness is increasingly limited by the sophistication of modern threats. Machine learning, especially deep learning, has emerged as a powerful paradigm for the detection of malware that provides the ability to analyze large datasets and adapt to changing threats. However, there is still a significant challenge, including the increasing complexity of malware, the effectiveness of evasion techniques, and the vulnerability of machine learning models to adversarial attacks. The future of malware detection is likely to be shaped by advancements in artificial intelligence, the scalability of cloud-based analysis, proactive defense strategies, and innovative applications of emerging technologies such as Large Language Models. In these areas, continuous research and development are crucial to staying ahead of cybercriminals and effectively protecting digital landscapes from constantly evolving malware threats.

References

[1] J. Ferdous, R. Islam, A. Mahboubi, and M. Z. Islam, “A Survey on ML Techniques for Multi-Platform Malware Detection: Securing PC, Mobile Devices, IoT, and Cloud Environments,” *Sensors*, vol. 25, no. 4, Feb. 2025, doi: 10.3390/s25041153.

[2] M. Azeem, D. Khan, S. Iftikhar, S. Bawazeer, and M. Alzahrani, “Analyzing and comparing the effectiveness of malware detection: A study of machine learning approaches,” *Heliyon*, vol. 10, no. 1, Jan. 2024, doi: 10.1016/j.heliyon.2023.e23574.

[3] A. A. Almazroi and N. Ayub, "Deep learning hybridization for improved malware detection in smart Internet of Things," *Sci Rep*, vol. 14, no. 1, Apr. 2024, doi: 10.1038/s41598-024-57864-8.

[4] T. Alsmadi and N. Alqudah, "A Survey on malware detection techniques," in *2021 International Conference on Information Technology, ICIT 2021 - Proceedings*, Institute of Electrical and Electronics Engineers, Jul. 2021, pp. 371–376. doi: 10.1109/ICIT52682.2021.9491765.

[5] J. Odii, J. Anenechukwu, and C. Hampo, "COMPARATIVE ANALYSIS OF MALWARE DETECTION TECHNIQUES USING SIGNATURE, BEHAVIOUR AND HEURISTICS," *Article in International Journal of Computer Science and Information Security*, vol. 17, no. 7, pp. 33–50, Jul. 2019, [Online]. Available: https://www.researchgate.net/publication/350017172_COMPARATIVE_ANALYSIS_OF_MALWARE_DETECTION_TECHNIQUES_USING_SIGNATURE_BEHAVIOUR_AND_HEURISTICS

[6] M. Satish Avhankar, J. Pawar, and V. Kumbhar, "A Comprehensive Survey on Polymorphic Malware Analysis: Challenges, Techniques, and Future Directions," *Communications on Applied Nonlinear Analysis*, vol. 32, no. 9, pp. 2765–2776, Mar. 2025, doi: 10.52783/cana.v32.4554.

[7] T. Dube, R. Raines, G. Peterson, K. Bauer, M. Grimaila, and S. Rogers, "Malware target recognition via static heuristics," *Comput Secur*, vol. 31, no. 1, pp. 137–147, Feb. 2012, doi: 10.1016/j.cose.2011.09.002.

[8] L. Diana, P. Dini, and D. Paolini, "Overview on Intrusion Detection Systems for Computers Networking Security," *Computers*, vol. 14, no. 3, Mar. 2025, doi: 10.3390/computers14030087.

[9] K. Wang, "Leveraging Deep Learning for Enhanced Information Security: A Comprehensive Approach to Threat Detection and Mitigation," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 15, no. 12, 2024, doi: 10.14569/IJACSA.2024.0151296.

[10] A. Afianian, S. Niksefat, B. Sadeghiyan, and D. Baptiste, "Malware Dynamic Analysis Evasion Techniques: A Survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 6, pp. 1–28, Nov. 2019, doi: 10.1145/3365001.

[11] M. Sewak, Sahay S. K., and H. Rathore, "Comparison of Deep Learning and the Classical Machine Learning Algorithm for the Malware Detection," in *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Busan, Korea (South): Institute of Electrical and Electronics Engineers, Aug. 2018, pp. 293–296. doi: 10.1109/SNPD.2018.8441123.

[12] A. Redhu, P. Choudhary, K. Srinivasan, and T. K. Das, "Deep learning-powered malware detection in cyberspace: a contemporary review," *Front Phys*, vol. 12, Mar. 2024, doi: 10.3389/fphy.2024.1349463.

[13] L. K. Shar, T. N. B. Duong, and D. Lo, "Empirical Evaluation of Minority Oversampling Techniques in the Context of Android Malware Detection," in *2021 28th Asia-Pacific Software Engineering Conference (APSEC)*, Taipei, Taiwan: Institute of Electrical and Electronics Engineers, Dec. 2021, pp. 349–359. doi: 10.1109/APSEC53868.2021.00042.

[14] Shijo P. V. and A. Salim, "Integrated Static and Dynamic Analysis for Malware Detection," *Procedia Comput Sci*, vol. 46, pp. 804–811, 2015, doi: 10.1016/j.procs.2015.02.149.

[15] A. Dixit and S. Singh, "Malware Detection Using Random Forest," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India: Institute of Electrical and Electronics Engineers, Nov. 2023, pp. 1–6. doi: 10.1109/ICCCNT56998.2023.10306628.

[16] Ali M. L., K. Thakur, S. Schmeelk, J. Debello, and D. Dragos, "Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study," *Applied Sciences (Switzerland)*, vol. 15, no. 4, Feb. 2025, doi: 10.3390/app15041903.

[17] B. Bokolo, R. Jinad, and Q. Liu, "A Comparison Study to Detect Malware using Deep Learning and Machine learning Techniques," in *2023 6th International Conference on Big Data and Artificial Intelligence, BDAI 2023*, Jiaxing, China: Institute of Electrical and Electronics Engineers Inc., Sep. 2023, pp. 1–6. doi: 10.1109/BDAI59165.2023.10256957.

[18] J. Alrzini and D. Pennington, "A Review of Polymorphic Malware Detection Techniques," *International Journal of Advanced Research in Engineering and Technology*, vol. 11, no. 12, pp. 1238–1247, Oct. 2020, doi: 10.34218/IJARET.11.12.2020.119.

[19] 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 (Switzerland)*, vol. 12, no. 17, Aug. 2022, doi: 10.3390/app12178482.

[20] M. Y. Wong, M. Landen, F. Li, F. Monrose, and M. Ahamad, *Comparing Malware Evasion Theory with Practice: Results from Interviews with Expert Analysts*, SOUPS 2024. Philadelphia, PA, USA: USENIX Association, 2024. [Online]. Available: https://www.usenix.org/conference/soups2024/presentation/yong-wong

[21] B. Singh and S. Singh Cheema, "Emerging Trends in AI-Powered Malware Detection: A Review of Real-Time and Adversarially Resilient Techniques," *Tuijin Jishu/Journal of Propulsion Technology*, vol. 45, no. 4, pp. 1841–1869, Nov. 2024, [Online]. Available: https://www.researchgate.net/publication/388405244

[22] A. Luz, S. Iseal, and G. Olaoye, "AI-Powered Malware Detection in Cloud Storage and Networks," 2025. [Online]. Available: https://www.researchgate.net/publication/390052712

[23] Y. Zhou, G. Cheng, K. Du, and Z. Chen, "Toward Intelligent and Secure Cloud: Large Language Model Empowered Proactive Defense," *ArXiv*, Dec. 2024, doi: 10.48550/arXiv.2412.21051.

[24] T. Akinsowon and H. Jiang, "Leveraging Large Language Models for Behavior-Based Malware Detection Using Deep Learning," Oct. 2024. doi: 10.17605/OSF.IO/YG62X.