A Comprehensive Review on Malware Detection Techniques

Unnimaya M U

Dept.Computer Science and Engineering WYD23CSNS08
unnimayamu2001@gmail.com

Abstract

Malware programs pose a significant threat as they are designed to disrupt computer systems and propagate through networks or Internet connections. Researchers are actively de- veloping anti-malware systems aimed at effectively detecting and protecting against these threats. Two primary approaches have been proposed: signature-based detection, which relies on identifying known malware through specific patterns or signatures, and heuristic-based detection, which utilizes rules to identify potentially malicious behavior. However, signature- based techniques are often ineffective against unknown malware variants and sophisticated evasion techniques such as code obfuscation, packing, polymorphism, and metamorphism. This survey paper provides an overview of current methodologies for detecting and analyzing malicious code, including static analysis, dynamic analysis, and hybrid approaches.

Keywords: Malware detection, Security, Survey, Malicious programs, Static analysis, Dynamic analysis, Hybrid Analysis.

I. Introduction

Malware, a term derived from ”malicious software,” poses a significant threat to computer systems worldwide. It encom- passes a variety of hostile, intrusive, and often destructive programs designed to infiltrate computer systems without the owner’s knowledge or consent. Viruses, worms, Trojans horses, rootkits, and spyware represent different forms of malware, each with its own method of spread and intended purpose.

The impact of malware can range from inconvenience to severe financial implications for individuals and institutions. Ransomware, for instance, emerged as a particularly alarm- ing threat, encrypting files or rendering systems inaccessible until a ransom is paid. This type of malware has evolved significantly since its inception, with increasingly sophisticated encryption techniques and extortion tactics.

Beyond ransomware, malware can manifest in various forms, such as adware that bombards users with pop-up ads or Trojans that grant hackers control over compromised systems. These malicious programs can compromise sensitive informa- tion like bank details, credit card numbers, and personal data, posing a serious risk to users’ privacy and security.

Addressing the threat posed by malware requires effective detection and removal strategies. Detecting ransomware, poly- morphic programs, and other malicious software is crucial for safeguarding computer systems and preventing further harm. Therefore, the development of robust detection mechanisms is essential to mitigate the risks associated with malware infections and protect users from potential threats.

MALWARE DETECTION TECHNIQUES

A Blockchain represents a distinct type of database dis- tinguished by its unique storage method. Unlike traditional databases, Blockchains organize data into blocks that are interconnected. Whenever new data is entered, a new block is appended to the chain. Central to Blockchain Technology is its expansive network comprising individuals who can act as validators, reaching consensus on various matters such as transactions. This process undergoes rigorous verification through mathematical algorithms, ensuring network security and integrity.

A. *Static analysis detection technique:*

Before running the program and testing it, errors in the source code are fixed automatically. The application is split using various tools like disassemblers, decompilers, debuggers, and source code analyzers such as Ollydbg. These tools help in understanding malware and its structure. Static analysis helps in finding weaknesses in the source code to prevent security issues. Source code can be checked manually, but using machine tools is more efficient.

1) *Advantages of Static analysis::*

- Is fast and safe.
- It gathers the structure of the code of the program under inspection.
- If we can use static analysis to figure out how malware behaves in an application, we can then use that informa- tion.

2) *Disadvantages of Static analysis::*

- Analyzing unknown malware is hard because static anal- ysis doesn’t help much, and often we can’t see the source code of programs.
- To analyze code well, researchers need to know assem- bly language and understand how the operating system works.

B. *Dynamic analysis detection technique:*

Dynamic code analysis, part of debugging, examines how an application behaves while running. It helps test the program in various situations without needing to create fake scenarios, which could lead to unexpected errors. This process can uncover issues that weren’t apparent during the design phase and can’t anticipate every possible scenario. It’s a common practice because it saves time and money on testing and makes maintenance easier.

1) *Advantages of dynamic analysis::*

- It can find connections that static analysis can’t catch, like dynamic links through reflection, dependency injection, or polymorphism.

2) *Disadvantages of dynamic analysis::*

- May negatively impact the performance of the applica- tion.
- We can’t promise that all the source code is covered because it runs depending on user actions or automated tests.

II. MALWARE DETECTION METHODS

A. *Signature-based Detection :*

Signal-based detection uses a special code to find malware on your computer. This code is like a unique signature that tells the antivirus program what type of malware it is. The antivirus program keeps a big list of these codes in a cloud database. When a file tries to access your computer, the antivirus checks if it matches any of the codes in the list. If it finds a match, it knows the file is malicious and gets rid of it. The antivirus also takes apart the infected file to understand its pattern and figure out what type of malware it is. If it finds a new type of malware, it adds its code to the list. There are different ways to do this detection, like static, dynamic, or mixed techniques. Additionally, there are three main types of signatures used to detect worms: network payload signatures, file signatures, and log file analysis.

1) *Advantages of the signature-based detection::*

- Broadly accessible.
- Easy to run.
- Fast identification.
- Finding comprehensive malware information.

2) *Disadvanatages of the signature-based detection::*

- Failing to detect polymorphic malware. vast database.

B. *Heuristic -based Detection:*

This approach helps detect and tell apart normal and abnor- mal behavior in a system to spot both known and unknown malware attacks. It uses weight-based rules to decide how risky a program's actions might be. If these rules cross a set limit, the system takes action to prevent harm, like isolating and deleting a file or sending a notification to the server administrator for alert.

III. SURVEY OF EXISTING WORKS

A. *Malware Detection Using Blockchain Technology*

In this paper, we propose a comprehensive solution to ad- dress the limitations inherent in individual malware detection methods. Recognizing the imperfections of existing detection techniques, particularly those utilizing blockchain technology, we advocate for a holistic approach that combines multiple methods for enhanced efficacy. Our proposed system operates in three main stages: firstly, when a user intends to download a file, its signature is cross-referenced in the blockchain to ascertain its presence (signature-based detection). If the signature is found, the associated details are retrieved, enabling users to assess its behavior in an isolated environment and vote on its maliciousness (behavior-based detection). Alternatively, if the signature is not present, the file undergoes testing, with the resulting signature added as a new block to the blockchain. Each blockchain block contains critical information, including the virus signature, the number of nodes voting for malicious or benign behavior, and the addresses of participating nodes. This integrated approach aims to mitigate the drawbacks of individual detection methods, thereby enhancing overall malware detection effectiveness.

B. *Malware Detection Using Machine Learning*

We propose a robust framework designed to effectively dif- ferentiate between malware files and clean files using various machine learning algorithms, with a focus on minimizing false positives. This paper presents the conceptual foundation of our framework, demonstrating its efficacy through two primary methodologies: cascade one-sided perceptrons and cascade kernelized one-sided perceptrons.

Initially, our framework was evaluated using medium-sized datasets comprising both malware and clean files. Through rigorous testing, we validated its effectiveness in accurately distinguishing between the two categories while minimizing the occurrence of false positives. Subsequently, we undertook a scaling-up process to enhance the framework's capabilities, enabling it to handle significantly larger datasets of malware and clean files.

Furthermore, our framework emphasizes the importance of adaptability and scalability, crucial factors in addressing the growing volume and sophistication of malware threats. Through continual refinement and optimization, we strive to ensure that our framework remains at the forefront of mal- ware detection technology, capable of effectively combating emerging threats across various scales and contexts.

C. *Malware Mobile Application Detection Using Blockchain and Machine Learning*

This paper introduces a novel approach for detecting malicious mobile applications through the integration of blockchain technology and machine learning algorithms. The proposed system utilizes a tailored blockchain architecture in conjunction with feature extraction techniques to enhance the detection of malicious applications. By employing both internal and external permissioned blockchains, the system can efficiently identify and prevent the distribution of harmful applications to users. Through the combination of static and dynamic analysis methods, the system achieves comprehensive malware detection capabilities. Moreover, the system's flexi- bility allows for the integration of various machine learning algorithms, ensuring adaptability to evolving threats.

S. N	Analysis Algorithm	Analysis/detection techniques	Challenges article	Computation cost	Future Works
1	a combination of behavioral analysis, signature-based detection, and heuristic- based detection, complemented by deep learning and machine learning algorithms [1]	Behavioral analysis,Signature-based detection,Heuristic-based detection,	Ensuring scalability and efficiency of the system as it grows in size and complexity. Addressing the computational cost associated with deep learning and machine learning algorithms. Mitigating the risk of false negatives and false positives in malware detection	minimize computational overhead	future work involves expanding the system's capabilities by incorporating more advanced heuristic-based detection methods, further

			. Overcoming potential regulatory and privacy concerns associated with blockchain technology and data sharing.		enhancing security by leveraging deep learning and machine learning algorithms, and refining the blockchain integration for improved accuracy and efficiency.
2	Feature Extraction using K-means Algorithm,Malware Detection with KNN [2]	Static and Dynamic Feature Extraction,Blockchain Integration,Collaborative Detection,Open System Integration	Interoperability	High	Application Verification, Feature-Based Filtering, Antimalware Solutions
3	cascade one-sided perceptron (COS-P),(COS- PMap) [3]	machine learning-based detection	Overcoming speed and memory limitations in commercial antivirus products	May be high	Integration of more classification algorithms such as large margin perceptrons and Support Vector Machines is planned to further enhance the framework’s detection capabilities and reliability.
4	machine learning algorithms [4]	online analysis of memory access patterns, employing a system/function-call epoch based memory access	accurate detection of both kernel and user-level threats,managing the computational overhead of online memory data collection,adapting the framework to address evolving	minimal computational overhead	improving scalability for real-time detection in larger systems,exploring methods for mitigating false positives further,refining the

		summary to identify malware-infected runs of known applications	malwar e techniques		framework to handle a broader range of malware threats
5	The proposed static model utilize s permissions analysis in a blockchain environment, ranking permissions based on Information Value to improve detection accuracy [5]	applying machine learning and deep learning algorithms to the top-ranked permissions	managing the complexity of analyzing numerous Android permissions, ensuring the relevance and accuracy of the ranked permissions, and address- ing the evolving nature of Android malware.	NA	Future work aims to incorporate additional static and dynamic features such as other manifest file components, system calls, network traffic, CPU and memory usage, to enhance the model's detection capabilities.
6	deep learning algorithms [6]	machine learning techniques	ensuring the security and accuracy of malware de- tection in a distributed system, as well as op- timizing the division of participants into functional groups	Highly effective	Future work will focus on optimizing partic- ipant grouping methods and network topology for enhanced efficiency and scalability in malware de- tection.

IV. Conclusion

In the contemporary digital landscape, malware has emerged as a pervasive global menace, inflicting significant harm on individuals, organizations, and entire economies. Research underscores an alarming trend: the evolving nature of malware is exacerbating its impact, posing formidable challenges to cybersecurity efforts worldwide. In response, the forefront of defense against malware comprises sophisticated malware detection devices, which stand as bulwarks shielding against malicious incursions into networks and systems.

The efficacy of these detection devices hinges crucially on the methods they employ. Therefore, understanding the intri- cacies of malware detection techniques becomes paramount. This paper has undertaken a comprehensive exploration of the technical landscape surrounding malware, illuminating the diverse array of malware detection methodologies. By dissecting these techniques, we have discerned their strengths and limitations, providing invaluable insights for cybersecurity practitioners and researchers alike.

The review presented herein encapsulates an exhaustive examination of various malware detection systems. Despite the relentless proliferation of malware and the rapid evolution of detection technologies, this study serves as a foundational resource for developers and stakeholders in the cybersecurity domain. By distilling complex technical concepts and offering a nuanced understanding of malware detection mechanisms, this research endeavors to equip practitioners with the knowl- edge needed to fortify their defenses against the ever-evolving threat of malware.

In conclusion, while the arms race between cybercriminals and defenders escalates unabated, endeavors such as this serve as beacons of knowledge, guiding the development of robust and effective countermeasures against the scourge of malware. By fostering a deeper understanding of malware detection techniques, we empower the cybersecurity community to stay one step ahead in the ongoing battle for digital security.

.

REFERENCES

1. Swarna Madhuri Pichikala,Rachana G,Sanjanapatel H,Saumya Shanu,Nandhini Vineeth.’Malware Detection using Blockchain Technology’,*2021 2nd International Conference for Emerging Tech- nology (INCET) — 978-1-7281-7029-9/20/*31.002021*IEEE|DOI* : 10.1109/*INCET* 51464.2021.945616
2. Naman Aneja,Sandeep Suri ,Sachin Papneja,Nikhil Khurana. ’Mal- ware Mobile Application Detection Using Blockchain and Machine Learning’,*2021 2nd Global Conference for Advancement in Tech- nology (GCAT) — 978-1-6654-1836-2/21/*31.002021*IEEE|DOI* : 10.1109/*GCAT* 52182.2021.9587880.
3. Dragos¸ Gavrilut, Mihai Cimpoes u, Dan Anton, Liviu Ciortuz. ’Mal- ware Detection Using Machine Learning’, *Proceedings of the Interna- tional Multiconference on Computer Science and Information Technol- ogy pp. 735–741,978-83-60810-22-4/09/*25.00c2009*IEEE.*
4. Zhixing Xu, Sayak Ray, Pramod Subramanyan, Sharad Malik. ’Malware Detection using Machine Learning Based Analysis of Virtual Mem- ory Access Patterns’, *2017 Design, Automation and Test in Europe (DATE),978-3-9815370-8-6/17/*31.00c2017*IEEE*
5. Siddhant Gupta, Siddharth Sethi, Srishti Chaudhary, Anshul Arora. ’Blockchain Based Detection of Android Malware using Ranked Permis- sions’, *International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958 (Online), Volume-10 Issue-5, June 2021*
6. [6] Dmytro Denysiuka , Olena Geidarovaa , Mariia Kapustiana , Sergii Lysenkoa , Anatoliy Sachenkob,c. ’Blockchain-based Deep Learning Algorithm for Detecting Malware’, *IntelITSIS’2023: 4th International Workshop on Intelligent Information Technologies and Systems of Infor- mation Security, March 22–24, 2023*