International Journal of Engineering Research and Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 22, Issue 4 (April 2026), PP 327-330

Survey on AI-based forensics analysis of self-healing malware in ethical hacking

Rakhi Jamdade, Rutuja Mohite, Aisha Patel, Jasmin Sahu, Kimaya Waghere
Department of Information Technology, K.C College of Engineering and Management Studies and Research, Thane, India

Abstract
Self-healing malware represents a sophisticated category of cyber threats characterized by its ability to restore lost components and dynamically adapt to detection mechanisms. Unlike conventional malware, it maintains persistence even after partial removal, posing significant challenges for forensic analysis. The traditional malware used to get completely deleted and wouldn't grow. Nowadays, the self-healing malware can restore its functionality even after partial removal, which makes forensic investigation difficult. This paper is a survey of Artificial Intelligence-based techniques used for the analysis of self-healing malware in ethical hacking. It studies the existing methods that use machine learning, behavioral analysis, and anomaly detection to identify and examine such threats.

The survey analyzes how AI models help in detecting hidden patterns, monitoring system behavior, and improving the efficiency of forensic investigation. It also highlights the drawbacks of traditional forensic tools and discusses the need for intelligent and automated solutions. The survey studies the ongoing challenges and proposes a solution. Overall, this paper will provide an overview of AI-driven forensic techniques and their role in identifying and eliminating malware threats.

Date of Submission: 11-04-2026 Date of acceptance: 23-04-2026

I. Introduction

The expansion of digital technologies and interconnected systems has contributed to a substantial rise in cyber threats and malicious activities across various domains. Among all the threats, the use of malware continues to increase. Malware is the most common and damaging tool to compromise systems, steal sensitive data, and disrupt services. An emerging type of malware is the self-healing malware, which is known for its ability to automatically recover and adapt after being partially removed or detected. It is a highly resilient and difficult to analyze using traditional techniques and forensic methods.

Digital forensics is essential for examining cyber incidents through the systematic collection, preservation, and analysis of digital evidence to support investigation processes. The dynamic and adaptive nature of self-healing malware gives rise to new challenges for forensic investigators. Traditional techniques such as signature-based detection and manual log analysis are often not sufficient to identify such complex threats, as the malware can continuously modify its behavior and evade detection mechanisms.

To address these challenges, AI is used as a powerful tool in cybersecurity and digital forensics. Machine learning and behavioral analysis enable the identification of hidden patterns and anomalies in data. This survey focuses on reviewing existing AI techniques for the forensic analysis of self-healing malware in ethical hacking. It highlights the role of intelligent systems in improving detection accuracy, reducing investigation time, and supporting proactive cybersecurity measures.

II. Literature Review

Recently, Artificial Intelligence has been widely adopted in cybersecurity due to an increase in malware and the need for malware detection and digital forensic analysis. Traditional methods, such as signature-based detection and rule-based systems, were effective earlier for identifying threats, but now, with the increase in modern malware, detection and elimination are not possible. Especially the self-healing variants that continuously modify their structure and behavior to avoid detection. Therefore, researchers have shifted towards intelligent and adaptive techniques.

Various machine learning techniques, including Decision Trees, Random Forests, and Support Vector Machines, are commonly applied to classify malware by analyzing features derived from system logs, file attributes, and network activity. These models are used to identify patterns associated with malicious activities, and they can detect previously unseen threats. In addition, Deep Learning approaches, including neural networks, have shown improved performance in analyzing complex and large-scale datasets.

Behavioral analysis has also gained importance in recent studies, where malware is detected based on its runtime activities rather than static signatures. Techniques such as anomaly detection and sequence modeling help in identifying unusual system behavior that may indicate the presence of self-healing malware. Furthermore, AI-based forensic tools have been developed to automate evidence collection, log analysis, and event correlation, thereby reducing manual effort and investigation time.

Despite these advancements, existing approaches still face limitations, including high computational requirements, reliance on high-quality datasets, and challenges in detecting highly adaptive malware. This highlights the need for more robust and efficient AI-driven forensic solutions.

III. Problem Statement

The rapid evolution of malware techniques has introduced new challenges in the field of digital forensics, particularly with the growth of self-healing malware. Unlike traditional malware, self-healing variants can restore deleted components, modify their structure, and continue execution even after partial detection or removal. This type of nature makes them highly resistant to conventional detection mechanisms, and it complicates forensic investigations significantly.

Traditional approaches such as signature-based detection and manual log analysis are often ineffective against such dynamic threats, as they rely on predefined patterns that can be easily altered by self-healing mechanisms. Moreover, forensic investigators face difficulties in collecting reliable evidence, as the malware may manipulate or regenerate system data, leading to incomplete or misleading analysis.

Although Artificial Intelligence (AI) has been applied in cybersecurity, there is still a need to systematically analyze its role in handling self-healing malware from a forensic perspective. This survey aims to address this gap by reviewing existing AI-based techniques and identifying their effectiveness and limitations in improving malware detection and forensic analysis.

IV. Proposed Method

This survey proposes a generalized AI-based framework for the forensic analysis of self-healing malware by reviewing and combining existing approaches. The framework focuses on integrating data collection, intelligent analysis, and automated reporting to improve the efficiency of forensic investigations.

The first stage involves data collection, where relevant data such as system logs, network traffic, and suspicious files are gathered from infected systems. This is followed by preprocessing, where the collected data is cleaned, normalized, and transformed into a structured format suitable for analysis.

In the next stage, AI-based detection techniques are applied using machine learning models such as Random Forest, Support Vector Machines, or Neural Networks to classify and identify potential malware. This is complemented by behavioral analysis, which examines runtime activities to detect unusual patterns and identify self-healing behavior, such as repeated regeneration of malicious components.

The framework also includes a correlation stage, where relationships between different events and data sources are analyzed to trace the origin and the spread of the malware. Finally, an automated reporting module generates structured forensic reports, summarizing findings and supporting further investigation. This framework highlights how AI can enhance traditional forensic processes by introducing automation and intelligent analysis.

V. Technical Approach

The technical approach of AI-based forensic analysis for self-healing malware involves the application of machine learning and data analysis techniques to identify and understand malicious behavior. Initially, collected data such as system logs, file attributes, and network activities are processed to extract relevant features. These features may include file size variations, execution patterns, frequency of system calls, and unusual network connections.

Machine learning models are then trained on these features to distinguish between normal and malicious behavior. Supervised learning techniques can be used when labeled datasets are available, while unsupervised methods, such as anomaly detection, can help identify previously unknown threats. In the case of self-healing malware, behavioral analysis plays a crucial role, as the system continuously monitors for repeated or regenerated activities.

Additionally, sequence-based analysis techniques can be used to track changes over time, helping to identify patterns associated with self-repair mechanisms. The results from different analysis modules are combined to provide a comprehensive understanding of the malware's behavior. This approach enables forensic

investigators to detect hidden threats more effectively and supports ethical hacking practices by providing deeper insights into system vulnerabilities.

VI. Results & Analysis

As this paper presents a survey of existing approaches rather than an experimental study, the results are based on observations and findings from previously published research. Various studies have demonstrated that AI-based techniques significantly improve the detection and analysis of malware compared to traditional methods. Machine learning models have shown higher accuracy in identifying both known and unknown threats by learning complex patterns from data.

Behavioral analysis techniques have proven particularly effective in detecting self-healing malware, as they focus on runtime activities rather than static signatures. Research also indicates that AI-driven systems can reduce the time required for forensic investigations by automating tasks such as data analysis, pattern recognition, and report generation. This leads to improved efficiency and allows investigators to focus on critical aspects of the analysis.

However, the effectiveness of these techniques depends on factors such as the quality of training data, the choice of algorithms, and the ability of the system to adapt to evolving threats. Some studies also highlight challenges such as false positives and the difficulty of interpreting complex AI models. Overall, the analysis suggests that while AI-based approaches offer significant improvements, further research is required to enhance their reliability and applicability in real-world forensic scenarios.

VII. Advantages & Limitations

Advantages

It enables faster detection by automatically analyzing large volumes of data, which reduces the time required for manual investigation. The machine learning models can identify hidden patterns and anomalies that are difficult to detect using traditional methods. This improves the overall accuracy of malware detection, especially for unknown or evolving threats. Additionally, AI-based approaches support behavioral analysis, allowing investigators to monitor how malware operates in real-time rather than relying only on static signatures. This is particularly useful in identifying self-healing behavior where malware repeatedly restores itself. Furthermore, automation in forensic processes helps reduce human effort and minimizes errors, making investigations more efficient and reliable.

Limitations

Despite its advantages, AI-based forensic analysis also has certain limitations. One major challenge is the dependency on high-quality training data. Inaccurate or insufficient data can affect the performance of machine learning models. Another limitation is the high computational cost associated with training and deploying advanced AI models. Additionally, highly sophisticated self-healing malware may still evade detection by adapting to AI-based systems. There are also concerns related to interpretability, as some AI models function as "black boxes," making it difficult to explain their decisions during forensic investigations. These limitations highlight the need for continuous improvement in AI techniques.

VIII. Future Scope

The future of AI-based forensic analysis in handling self-healing malware is promising, with several areas for improvement and development. One potential direction is the use of advanced deep learning techniques that can better understand complex malware behaviors and adapt to evolving threats. Integration of real-time monitoring systems can further enhance the ability to detect and respond to malware attacks instantly.

The use of cloud-based forensic platforms can enable scalable and efficient analysis of large datasets. Collaborative approaches may also allow organizations to share knowledge without compromising sensitive data. Future research can focus on improving dataset availability and creating standardized benchmarks for evaluating AI models in malware forensics. Overall, continuous advancements in AI will play a crucial role in strengthening cybersecurity and supporting ethical hacking practices.

IX. Conclusion

Self-healing malware represents a significant challenge in modern cybersecurity due to its ability to adapt, regenerate, and evade traditional detection techniques. This survey examined the role of Artificial Intelligence in enhancing the forensic analysis of such advanced threats within the domain of ethical hacking. It highlighted how AI-based approaches, including machine learning and behavioral analysis, can improve the detection and investigation of complex malware activities.

The study also discussed the limitations of conventional forensic methods and emphasized the need for intelligent and automated solutions. While AI offers improved accuracy and efficiency, challenges such as data

dependency, computational cost, and model interpretability still need to be addressed. Despite these limitations, AI-driven forensic techniques show great potential in strengthening cybersecurity defenses.

In conclusion, the integration of AI in digital forensics provides a promising direction for detecting and analyzing self-healing malware, thereby supporting ethical hackers and security professionals in combating evolving cyber threats.

References

[1]. H. Farid, "Image Forgery Detection," *IEEE Signal Processing Magazine*, 2009.
[2]. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
[3]. S. Axelsson, "Intrusion Detection Systems: A Survey," *Technical Report*, 2000.
[4]. M. Sikorski and A. Honig, *Practical Malware Analysis*, No Starch Press, 2012.
[5]. U. Bayer et al., "Scalable, Behavior-Based Malware Clustering," *NDSS*, 2009.
[6]. A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cybersecurity Intrusion Detection," *IEEE Communications Surveys & Tutorials*, 2016.
[7]. K. Rieck et al., "Automatic Analysis of Malware Behavior using Machine Learning," *Journal of Computer Security*, 2011.
[8]. S. Garfinkel, "Digital Forensics Research: The Next 10 Years," *Digital Investigation*, 2010.
[9]. S. Hou, A. Saas, Y. Ye, and L. Chen, "Deep4Malware: Deep learning-based malware detection," in *Proc. IEEE Int. Conf. Smart Computing (SMARTCOMP)*, 2016, pp. 1–8.
[10]. D. Ucci, L. Aniello, and R. Baldoni, "Survey of machine learning techniques for malware analysis," *Computers & Security*, vol. 81, pp. 123–147, 2019.