Abstract

As digital technologies advance and become integral to various sectors, the security of these systems faces significant threats from malware and malware-based attacks. Malware, malicious software designed to disrupt, damage, or gain unauthorized access to computer systems, presents a multifaceted challenge due to its evolving nature and increasing sophistication. These malicious entities range from traditional viruses and worms to advanced ransomware and spyware, capable of executing complex attack vectors that can bypass conventional security measures. Malware-based attacks exploit system vulnerabilities to perform unauthorized activities, such as data theft, espionage, and operational disruption, leading to substantial financial and reputational damage.

Malware-based attacks exploit system vulnerabilities to execute unauthorized activities, which can include stealing sensitive information, deploying additional malicious payloads, or even controlling systems for botnet-based attacks. The detection of these threats is thus critical to preserving the integrity, confidentiality, and availability of computer systems. Traditional detection methods, which often rely on signature-based or heuristic approaches, struggle to keep pace with the rapid evolution of malware techniques, such as polymorphism and obfuscation. As a result, there is a pressing need for the development of more advanced, adaptive detection systems that can preemptively identify and neutralize these threats.

The main challenge in malware detection lies in designing methods that are not only effective in identifying known threats but are also capable of detecting novel or evolving malware with high accuracy. This necessity is underscored by the increasing sophistication of attack vectors, the diverse range of malware types, and the continuous development of evasion techniques by malicious actors. Emerging technologies such as machine learning, artificial intelligence, and deep learning present promising solutions to address these challenges. These cutting-edge technologies can analyze vast amounts of data, recognize intricate patterns, and adapt to new malware, providing a more robust and dynamic defense against malware. By harnessing the power of these advanced technologies, innovative, adaptive, and proactive defense mechanisms can be developed to safeguard digital assets and en-

sure the ongoing security of technological ecosystems, thereby staying ahead of the constantly evolving landscape of cyber threats.

This thesis work comprises several key contributions that address these challenges in malware detection. Firstly, it provides an extensive, in-depth study of malware, its taxonomy, detection approaches, and various analysis tools. It includes the design and development of a malware dataset generation pipeline, presenting two malware feature datasets on two different platforms to support the validation of the effectiveness of a malware detection method. Additionally, the thesis introduces a supervised filter-based feature selector based on rough set theory, which proposes a new criterion for identifying the most relevant features. It also presents a fast and reliable ransomware defense solution powered by an optimal feature selection method to discriminate the ransomware class as a whole and the eleven variants of the ransomware family from goodware instances. Furthermore, it develops an ensemble approach to identify key features that significantly contribute to malware detection. The work includes a parallel version of the KNN algorithm, with different proximity measures to enhance detection accuracy and efficiency. The thesis also explores a Convolutional Neural Network (CNN)-based malware defense solution, leveraging CNNs' strengths in pattern recognition to improve malware detection. Lastly, it investigates a Graph Neural Network (GNN)-based malware defense solution, utilizing GNNs' capabilities to analyze and interpret complex relationships within data, thus enhancing detection performance.

Keywords: Malware, Signature, Detection, Classification, Static Analysis, Dynamic Analysis, Honeynet, Anomaly, Feature, Ensemble, Markov Chain, Deep Learning, and Graph Neural Network