Chapter 7
Object Detection in Cybersecurity:
A Review of Automation of Malware Detection

Stones Dalitso Chindipha
Rhodes University, South Africa

ABSTRACT

With the increase in malware attacks, the need for automated malware detection in cybersecurity has become more important. Traditional methods of malware detection, such as signature-based detection and heuristic analysis, are becoming less effective in detecting advanced and evasive malware. It has the potential to drastically improve the detection of malware, as well as reduce the manual efforts required in scanning and flagging malicious activity. This chapter also examines the advantages and limitations and the challenges associated with deploying object detection in cybersecurity, such as its reliance on labeled data, false positive rates, and its potential for evasion. Finally, the review presents the potential of object detection in cybersecurity, as well as the future research directions needed to make the technique more reliable and useful for cybersecurity professionals. It provides a comparison of the results obtained by these techniques with traditional methods, emphasizing the potential of object detection in detecting advanced and evasive malware.

DOI: 10.4018/978-1-6684-8127-1.ch007

INTRODUCTION

The use of machine learning techniques in cybersecurity has been active for at least two decades and it has produced actionable results that others have built on. For instance, as early as 1998, Machine Learning (ML) techniques were applied to identify the discriminant features of malicious network traffic and classify legitimate network traffic from malicious traffic (Shafiq et al., 2020; Kundu et al., 2022).

It is the same machine-learning technique that emails use to filter out spam. Despite these ML techniques being used, the rapid change in the format of spam emails eludes some of these algorithms, and thus users find these spam emails in their inboxes. Another example of targeted cyber security attacks on machine learning techniques is presented by (Xi, 2020) which involves adversarial attacks against Deep Neural Networks (DNN). DNN has drawn significant attention because of how it is now applied in critical tasks, such as autonomous driving systems and partly automated vehicles. Thus, though the use of ML techniques is good for improving malware detection, some ML models are being targeted by vicious attacks. These too have to be looked at irrespective of the job they do in detecting malware.

While some autonomous malware detection strategies have worked, others have not worked. This chapter reviews and evaluates the peer-reviewed work on autonomous malware detection and looks at the strengths and weaknesses of each technique that has been used thus far and how some of these ML techniques have opened loopholes in systems. For instance, deep neural networks fail to correctly classify adversarial images (Xi, 2020), more work needs to be put into understanding why such is the case by looking at what other researchers have done and reporting those with higher success and what they did differently than the preceding work to achieve greater success. Even ML techniques that work in cyberspace have their attacks designed specifically for them to avoid detection. This includes but is not limited to poisoning attacks and evasion attacks (Xi, 2020).

Poisoning attacks have targeted machine learning techniques for some time. They work by contaminating the training dataset before training which in turn causes a learning model to make costly mistakes (Tian et al., 2022). These poisonous attacks can further be split into targeted attacks or non-target attacks. With targeted attacks, a threat actor works with tailor-made malware that would affect a specific organization while non-targeted attacks aim at reducing the overall accuracy of a learning model thus resulting in a majority of the malware being undetected and damaging systems. A major problem here is that a majority of publicly available datasets are outdated and may not be sufficient in identifying the undocumented behavioural patterns of various cyber-attacks (Xi, 2020). That means due to a lack of freely accessible data, this same dataset can be shared by many people which in turn means that if the data was poisoned at the collection point, then every other finding from such a