Malware Detection in Network Flows With Self-Supervised Deep Learning

Thomas Alan Woolman
On Target Technologies, Inc., USA

Philip Lunsford
East Carolina University, USA

INTRODUCTION

With ever-increasing network complexities and threat actor sophistication, the vulnerabilities of critical network infrastructure and host systems are potentially greater than ever before. The ability of targeted organizations and government entities to defend their network perimeters utilizing traditional threat detection systems provides only a limited set of tools that are traditionally based on simple statistical tests of network activities and known threat signatures. These threat signatures generally rely on predefined malware detection rules based on known, previously encountered network intrusion attack types. As a result, sensitive information and critical resource applications can potentially be highly vulnerable to novel sophisticated and evolving network intrusion types, potentially putting commercial and public sector resources and information in mounting jeopardy.

The ability to detect legacy cyber threats through a multilayered defense approach is based on research pioneered by Chess and White (1987), initially based on permutations of signature detection methods proposed by Cohen (1987). While signature-based network malware and intrusion detection are still among the most heavily used techniques, heuristic approaches that are able to discern multiple, related threats from a single definition source have been increasingly common, as defined by Kaspersky Lab ZAO (2013). However, novel threats as well as more advanced cyber malware and intrusion events that are explicitly designed to avoid detection by the more commonly used available tools and techniques are becoming increasingly common. By being able to bypass the network security perimeter, intrusions and malware can quickly propagate throughout the network and operate undetected for substantial lengths of time. In many cases, these network intrusions can access restricted information while remaining undetected, masquerading their traffic signatures as legitimate, benign activities.

As the capability to resist successful classification is increasing with the latest generation of network intrusion technologies, continuous improvement in the multilayered network defense approach first proposed in 1987 becomes increasingly necessary. One example of this emerging malware threat class is a sophisticated modular malware known as Flame, first discovered in 2012 on networked devices running the Microsoft Windows operating system (ICIRT, 2012). Flame, also known as Skywiper, is believed to likely have been developed by a state actor as a cyber-weapon that was deployed for espionage purposes for one or more targets in the Middle East (Kaspersky Lab ZAO, 2013).

DOI: 10.4018/978-1-7998-9220-5.ch139

First detected inadvertently in 2012, Flame is now generally regarded as an unusually robust backdoor attack toolkit, with worm-like features and Trojan capabilities, with the ability to replicate both within a targeted network as well as on removable media upon receipt of commands to do so by a remote threat actor's command and control server. Although the exact method of entry into a network has not yet been determined, Flame's ability to take on different roles through a wide range of add-on functional libraries allows it to be both extraordinarily adaptable and difficult to analyze by traditional mitigation and detection methods, utilizing the novel technique of concealment through an unusually large and variable codebase compared to most other network malware threats.

Flame is capable of harvesting sensitive data in a variety of ways, including robust SQL database query insertions, compressed digital audio microphone recording, Bluetooth wireless connectivity attacks from inside the network, as well as file and network traffic ingestion and analysis. Furthermore, Flame can also take recurring screenshot images from infected devices. Flame is capable of reporting back to an external command and control server from within the targeted network via a covert SSL data channel, as well as turning other host devices within the network into beacons that are discoverable via Bluetooth connections, according to Kaspersky Lab ZAO (2013).

Advances in this and similar emerging, novel threat categories of malware and network intrusion capabilities thus require adaptable learning technologies for detection that are not based on predefined statistical patterns, heuristics, or rule-based detection methods. One increasingly utilized form of malware and network intrusion detection technology is anomaly-based detection methods, often utilizing data mining technologies including machine learning and deep learning. Deep learning anomaly detection utilizing network traffic analysis such as packet capture and network flow data is one such emerging advance in this field, one which does not require the use of predefined, human-labeled training data that could be obsolete when faced with novel malware threats. Thus, unsupervised deep learning algorithms for anomaly detection in network connection datasets represent a possible significant component of a multi-layer network defense strategy designed to detect advanced novel malware threats across digital networks, because these systems do not rely on pre-programmed malware threat signatures.

An advantage of utilizing packet capture data sets for anomaly-based network intrusion detection systems (NIDS) is that full packet capture allows for a mirror image of the entirety of the network traffic for a given period of time, allowing robust deep packet inspection (DPI). The DPI data set allows for a complete forensic analysis of all available features including protocols, payloads, and source and origins for each packet, as well as a variety of measurements related to packet transmission speeds and delays.

One significant disadvantage of packet capture data set forensics for NIDS is that DPI imposes a significant burden on routers, switches, and network infrastructure in general during this mirroring cycle to capture and store the vast amount of network traffic. Furthermore, the data storage, processing, and analysis of these often quite deep (often multi-terabytes per day within enterprise networks), wide (typically on the order of dozens of independent variables per observation), and complex data sets often requires the use of more cumbersome "Big Data" analytical cluster computer environments. These specialized analytical frameworks thus necessitate an increase in the scope and complexity of these projects. Robust encryption methods of packet payload data further increase the signature detection complexity of DPI analysis (Woolman & Lee, 2021).

Conversely, network flow data sets represent a more "high level" metadata scope of network traffic within the enterprise, providing summarized level data between the source (IP address and port) and destination (IP address and port) per protocol. Rather than recording the actual packet payload of each component of network traffic, the network flow data set typically records only the information about