Recent Submissions

ConFoc: Content-Focus Protection Against Trojan Attacks on Neural Networks

Deep Neural Networks (DNNs) have been applied successfully in several hallmark tasks of computer vision. Despite their success, the wide adoption of DNNs in image-related applications is threatened by their vulnerability to adversarial settings, including those classified as trojan attacks. Trojan attacks are characterized by the execution of slight changes to the models during training to insert an undesired behavior that is later exploited at inference or testing time. This paper presents an analysis of the attack vector exploited by adversaries with regard to the marks added to input images to insert and trigger this misbehavior. We identify that triggers have two different exploited features that need to be together in training and testing times for these attacks to be effective: content or semantic information, and style or texture information. We then introduce a novel defensive technique, in which DNNs are taught to focus on the content of inputs (disregarding their styles) so as to mitigate the effect of triggers on the attack. The generic applicability of our method was demonstrated in the context of a traffic sign and a face recognition application, each exposed to a different attack. Results show our content-focus approach reduces the attack success rate to or close to 0.0% on each application respectively, keeping and sometimes improving the accuracy of the models.

Submitted to: ACM CCS2020

LSTM-based Anomaly Detection In High Dimensional Sequential Data

Modern evasion attacks such as multistage persistent and mimicry-based attacks stand out by their sophistication in stealthiness and spanning for long periods of time. A common approach followed by adversaries is inserting dummy events to the malicious sequences to circumvent detection of traditional mechanisms that work based on the likelihood of fixed-length sequences. This brings the necessity for system behavior modeling that allows learning the expected event patterns in a system to identify anomalous sequences even when the intrusion spans for a long time. We introduce a deep learning based approach for the detection of such anomalies. More specifically, the approach is based on Long Short Term Memory (LSTM) models and answers the anomaly detection problem of given a sequence of events $e_1. . . e_{n-1}$, whether or not the sequence $e_1. . . e_{n-1}\ e_n$ should occur. The work includes a detailed analysis of the properties of these models, their capacity to discriminate sequences of large length and their limitations. The generic applicability of the technique is demonstrated on two datasets of over 2.1 and 38.9 million activity events respectively. These events are collected from a commercial security product running in normal (non-attack) and under attack conditions. Results show that the technique is able to detect anomalies with a TPR of 95.47% for variable length sequences, suggesting that LSTM-based approaches outperform traditional detection mechanisms in the detection of advanced evasion attacks.

Submitted to: EAI SecureComm 2020

Deep Learning-Based Real-Time Cyberattacks Detection Using System Telemetry

Autonomy of cyber systems depends on their ability to accurately classify services as well as applications that are running on those systems. In case of mission-critical systems that are deployed in dynamic and unpredictable environment, the newly accepted applications and service interactions can either be essential to continue the mission to achieve a certain goal or cyberattacks. In particular, some of these cyberattacks are evasive Advanced Persistent Threats (APTs) where the attackers remain undetected for reconnaissance to ascertain system features for an attack e.g. Trojan Laziok. In other cases, the attacker can use the system only for computing without disrupting normal system functionalities e.g. Cryptojacking. We propose a behavioral profiling model using deep neural networks—Recurrent Neural Networks—using Performance (PERF) Counters, providing the system telemetry of 68 unique properties and 245 associated system event properties. In addition, we propose a novel model selection framework that can pick a model for classification from both light-weight machine learning models and deep learning models using a reinforcement learning utility function. Using Cryptojacking as application, we show that our models perform with high accuracy and online detection rate with various sizes of available data in classifying evasive applications and the reinforcement learning utility function reducing the use of computational resources for accurate classification.

Planned Submission: AIKE 2020

CodePro: A Deep Learning-Based Model for Behavior Profiling in Cyber Systems

The automatic profiling of applications in order to differentiate between benign and malicious applications or services is an important non-trivial research problem. Existing solutions rely on human experts to define application features. Evasive applications can easily manipulate or mask their process structure, misleading human experts and light-weight machine learning models. In particular, evasive Advanced Persistent Threats (APTs) can stay in the system undetected and make use of computational resources e.g. Cryptojacking. Some cyberattacks even mimic normal software and ascertain system features e.g. ransomware, file-less malware. In this paper, we initiate the study of modeling computing language as a natural language using deep learning for understanding and profiling applications. Since Natural Language Processing (NLP) principles are different from Programming languages, we introduce new principles for applying deep learning for application profiling and cyberattack detection. In particular, we introduce representation of assembly instructions that are extracted from the memory for each process: Assembly Blocks. Assembly blocks are sequences of instructions that are systematically related to each other. This leads us to the design and development of deep learning-based profiling model called Computing Deep Profiler (CodePro). In order to evaluate CodePro, we use cryto mining algorithms as our evasive malware application examples and compression / encoding applications as benign examples and extract their instructions sequences from memory for each application. Our experimental results show that CodePro achieves high accuracy with low false negative and false negative rates. We also show the performance of the model using online detection rate and the size of the data required to make an accurate classification as well as computing resources required for training.

Planned Submission: ICMAL 2020