CERIAS

The Center for Education and Research in Information Assurance and Security

PEA: PE Analyzer-Detecting PE malware using machine learning algorithms

Rushabh Vyas and Nichole McFarland, Graduate Students
Dr. Xiao Luo and Dr. Connie Justice

Abstract

Objective of this research is examine how malicious portable executable (PE) files can be detected utilizing Machine Learning (ML) algorithms. The efficiency and effectiveness of detection rely on the number of features and the learning algorithm.

In this research, features are extracted from PE files and analyzed using different ML algorithms to determine effectiveness. In the proposed system, we can achieve 98% detection rate.

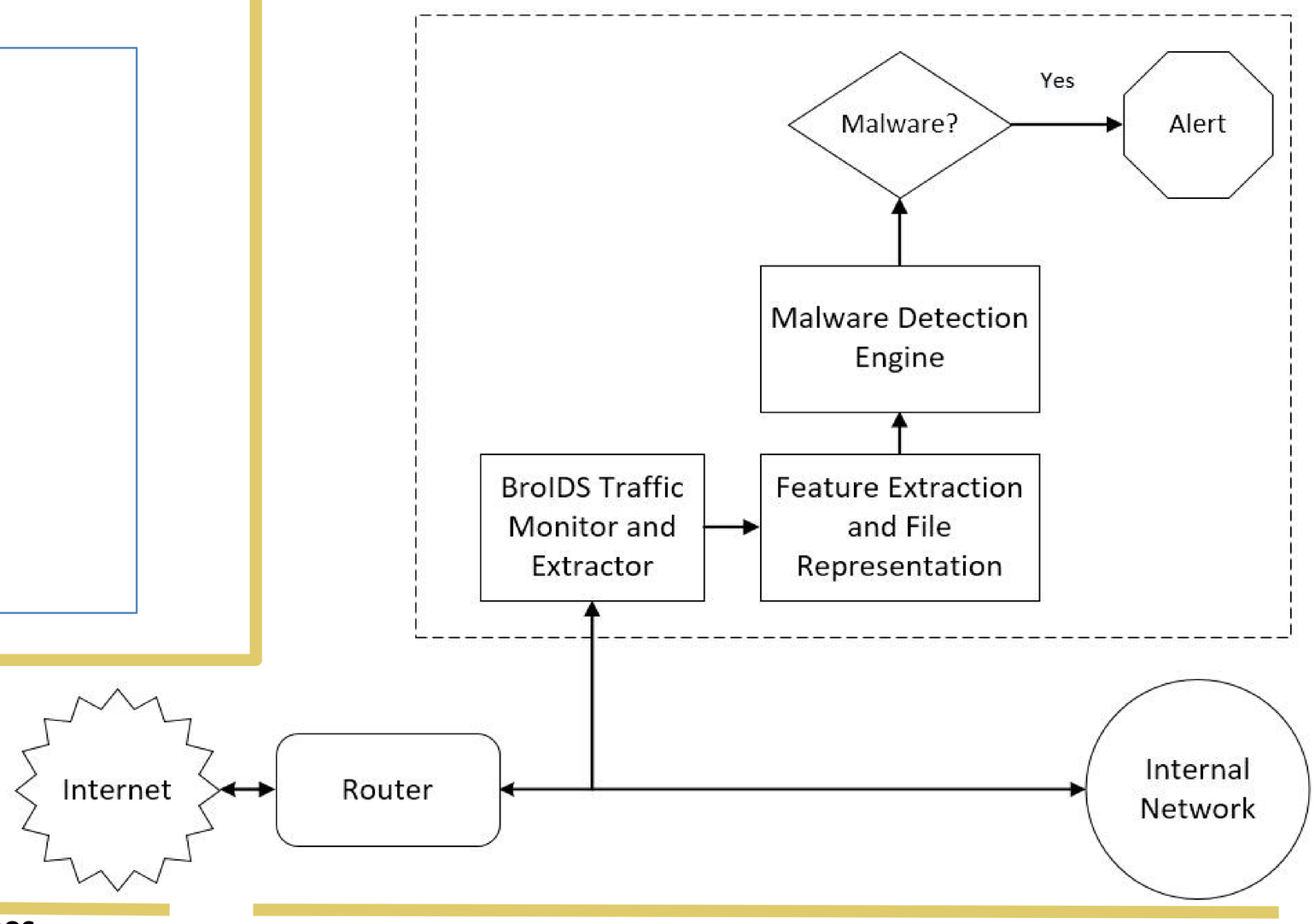
Feature Extraction
- Features from benign and malicious samples are extracted

Training
- Extracted features are given to the training algorithm

Detection
- PE file features are extracted and supplied to ML algorithm for classification

PE Features

- Entry point
- Number of Symbols
- Number of PE Sections
- File Entropy
- Compile Time
- DLL Imports
- Imported Functions



Detection Rates of Different Malware Types (Percentage)

Malware Type	k-nearest neighbors (kNN)	Decision Tree (DT)	Random Forest (RF)	Support vector machine (SVM)
Worm	97.1	97.9	98.9	72.6
Trojan	97.0	97.2	98.7	72.6
Backdoor	97.0	98.0	99.0	72.6
Virus	96.1	97.9	98.3	72.6
Weighted-Average	96.8	97.8	98.7	72.6

Future Work

- Test the proposed malware detection system on a server with more computing resources and on different network traffic settings.
- Expanding system to detect portable executable files for other OS platforms ie. Unix.
- Extensively evaluating the proposed system on big data sets using multiple evaluation metrics such as Receiver Operating Characteristic (ROC) curve
- Extensively evaluating the proposed system on big data sets using multiple evaluation metrics such as F1-measure and investigating the minimal number of features that include dynamic features for network malware detection

CERIAS

PURDUE UNIVERSITY