Project 3 Proposal

Machine Learning in Cybersecurity (WS 2020)

December 24, 2020

1 Problem Statement

The aim of the project is to perform visualization and classification of malicious code or programs commonly referred to as malware. The central idea is to combine deep learning (specifically Convolutional Neural Networks) and image processing techniques to achieve the objective [2].

2 Motivation

In the present age of technology, malware attacks are reported quite frequently. The main targets of these attacks being both financial institutions and everyday users. The damage perpetrated by such malware attacks could range from losing critical or personal data or failure in a nuclear power plant. Thus, our computing systems need protection from malware attacks round the clock. Automatic malware identification and detection tools are aimed at addressing this very problem.

3 Proposed Strategy

We leverage the power of CNNs for image classification. We first transform a malware signature into its binary representation and then convert this binary into an 8 bit vector followed by transformation to a grayscale image. The final step is to feed the image into our network for classification.

4 Related work

Traditional approaches for analysis and classification of malware relied extensively on static or dynamic analysis. The former analyses binaries without executing by inspecting features such as control flows. The latter on the hand involves executing binaries under controlled environments and stating changes in the executed environment. Static analysis provides complete coverage but suffers in the face of code obfuscations. Dynamic analysis on the other hand is quite time consuming.

Deep learning based malware classification has emerged as an interesting alternative approach to tackle the growing menace of malware. One of the works in this direction is from [3] Rieck et al. The authors used features derived from the behavioral analysis of malware to classify them. A labeled dataset of 10,072 malware samples, labeled by an anti-virus software, was used. The behaviour of all malware samples was observed in a sandbox environment and a behavioral report was generated for each. The report was used to generate a feature vector for every malware based on the frequency of some specific strings in the report. Finally, a Support Vector Machine was used for training and testing and the reported average classification accuracy was 88%. Another work is from Tian et al., [4] wherein the authors used the length of a program, to classify 7 different types of Trojans and obtained an average accuracy of 88%. The chosen dataset however, was quite small with only 721 files. In [5], [1] the same authors improved upon their above technique by using printable string information from the malware and reported a classification accuracy of 98.8%.

5 Existing code/software

Malware Classification on Microsoft Malware Classification Challenge, and they update their database with the latest implementation and algorithms. Such as HYDRA (2020), Narayanan et al. (2016): PCA features + 1-NN etc.

6 Implementation

We transform malware signatures into images. Then define the hyper-parameters for our CNN and fine-tune our model with regards to the existing code.

7 Evaluation - metrics

We will use Classification Accuracy as an evaluation metrics.

8 Evaluation - data-sets

Initially, we will use Malimg Dataset as it is available. But also hope to use a dataset from Microsoft Malware Classification Challenge, which consists of nearly half a terabyte of uncompressed data. It will be a heavy computation and time consuming. So we will use the Mailmg dataset. If time permits then we will also use Microsoft Malware Classification Challenge dataset.

9 Evaluation - baselines

We will compare our model and its result with [2] Narayanan et al. (2016): PCA features + 1-NN result reported accuracy 96.60%

10 Success criteria

We try to achieve 97% plus but as we do not have any prior experience on this track, achieving 92% can be a great success for us.

11 Team

Team Name: Alpha
Members:

1. H T M A Riyadh(s8htriya@stud.uni-saarland.de)
2. Fahad Hilal (s8fahila@stud.uni-saarland.de)

References

[1] Rafiqul Islam et al. "Classification of malware based on string and function feature selection". In: *2010 Second Cybercrime and Trustworthy Computing Workshop*. IEEE. 2010, pp. 9–17.

[2] Lakshmanan Nataraj et al. "Malware images: visualization and automatic classification". In: *Proceedings of the 8th international symposium on visualization for cyber security*. 2011, pp. 1–7.

[3] Konrad Rieck et al. "Learning and classification of malware behavior". In: *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer. 2008, pp. 108–125.

[4] Ronghua Tian, Lynn Margaret Batten, and SC Versteeg. "Function length as a tool for malware classification". In: *2008 3rd international conference on malicious and unwanted software (MALWARE)*. IEEE. 2008, pp. 69–76.

[5] Ronghua Tian et al. "An automated classification system based on the strings of trojan and virus families". In: *2009 4th International Conference on Malicious and Unwanted Software (MALWARE)*. IEEE. 2009, pp. 23–30.