GR-83 Analysis of Malware Detection Techniques

Abstract

Today's digital world is pervaded with malware. In response to this reality, there are copious studies being conducted around the world on how best to improve the detection of malware, as malware becomes more sophisticated with every passing year. In the following report, we will discuss some current studies of interest on malware detection techniques and propose some of our own suppositions on how these suggested techniques can be improved upon.

Research Statement & Conjecture

In *Malware Intrusion Detection For System Security* [1], Katkar et. al suggest a two-stage system of malware detection. The first stage is a blacklist of websites which are known to host malware. The second is a classification model trained to detect malware. The objective of this research assignment is to validate the performance of this model against a larget dataset, employ standard data-cleaning procedures to imrpove performance, and assess the system against common tactics used by malware to evade detection.

The primary shortfall of the proposed solution [1] is that it does not identify a system for tracking the behavior of a program beyond first inspection. "The detection of malware is done on the latest files which have been downloaded on the system" (Katkar).This is important because malicious code can, and often does, wait for some period before doing anything malicious. A partial solution to this problem is to perpetually monitor the behavior of all files.

Katkar et. al do not mention specific approaches used for data cleaning, which may have a significant impact on model performance according to Letteri [7]. Specifically, Letteri found that after removing outliers and excluding all but the 6 best performing features, a Multi-Layer Perceptron model outperforms Random Forest.

Methods

The project addresses the efficacy of the two-stage system proposed as proposed in "Malware Intrusion Detection For System Security," by Katkar, Shukla, Shaikh and Dange [1] from two directions.

The first is from a theoretical perspective. To this end, we are researching evasion tactics that are currently being employed by malicious programs to see if there are any well-known tactics that can evade this detection strategy.

The second is from an experimental approach in which we replicate the experiment performed by Katkar, Shukla, Shaikh and Dange in order to critically assess the performance of the classification algorithm. To this end, we will determine the accuracy (specifically with respect to data that the model has not yet seen) in addition to False Positive/Negative ratios.

Potential improvements to the baseline model include:

- ❑ Experimenting with alternative classification models.
- ❑ Outlier detection and deletion.
- ❑ Hyperparameter tuning.

Antivirus Sofware Evasion Techniques

Malware is evolving at a rapid pace. The techniques used to analyze and detect malware must evolve as well, as the evasion techniques in newer malware are making them increasingly difficult to detect. In the survey by Robert Grimes [5], the author discusses dynamic analysis techniques, both manual and automated, to detect malware and to study how each invasion operates. Grimes further classifies the known types of evasion techniques that malware venders use and their efficacy against different types of analysis and detection approaches.

***Static detection**:*
A static detection metric is one which does not require the suspicious program to be executed. One of the most popular forms of static detection is signature matching. This method relies on a database of known malware fingerprints. When a file is downloaded, its fingerprint is checked against the database. Most commercial antivirus programs employ this technique due to its low false negative (falsely identified as illegitimate) rate [8]. However, as the method relies on a database of known malware, it is not useful in detecting novel malware.

Dynamic detection:
The bulk of malware detection in this current environment relies on dynamic detection, which is analysis of the behavior of the program or process at run-time. There are two main categories of Dynamic detection: manual and automated.
Malware authors, being aware of recent developments in malware detection, enlist strategies to better evade detection. The malware industry recognizes that if a debugger is deployed – which is often the tool used to identify malware – it will be used to thwart their malicious intents. The malware will also probe its environment to discern whether the environment is "production" vs. a sandbox, which is an environment used to inspect suspicious code

Conclusions

We have replicated the experiment shown in *Malware Intrusion Detection For System Security* [1] on a larger dataset of 200k samples, achieving a similar score of 98.9% accuracy. FalsePositive: 1.4%, FalseNegative: 0.9%

MTA-KDD'19: A Dataset for Malware Traffic Detection suggests a Multi-Layer Perceptron classifier performs better than Random Forest after cleaning the dataset. We were not able to replicate this finding on our dataset.

To conclude, simple classification algorithms are effective for detecting malware. But we cannot ignore the evasive tactics employed by malicious programmers. One clear shortfall of the system proposed in [1] is the absence of continuous monitoring. This is necessary in order counter a hard-coded delayed activation of malicious behaviour.

If a malicious program can detect the presence of a resource monitor, then it may still evade detection by halting malicious activity until the detection system finishes collecting the dynamic behavioural data required for classification. Unless, of course, the static attributes of the program are sufficient to label it as malicous.

https://github.com/JoelStansbury/os6025_project

Contact Information

Kiefer Bazan, KSU Student: kbazan@students.Kennesaw.edu
Joel Stansbury, KSU Student:jstansb2@students.Kennesaw.edu
Robert White, KSU Student: rwhit168@students.Kennesaw.edu

References

[1] Katkar, S. Shukla, D. Shaikh and P. Dange, "Malware Intrusion Detection For System Security," 2021 International Conference on Communication information and Computing Technology (ICCICT), 2021, pp. 1-5, doi: 10.1109/ICCICT50803.2021.9510161.
[2] Amir Afianian, Salman Niksefat, Babak Sadeghiyan, and David Baptiste. 2019. Malware Dynamic Analysis Evasion Techniques: A Survey. ACM Comput. Surv. 52, 6, Article 126 (January 2020), 28 pages. DOI: https://doi.org/10.1145/3365001.
[3] Bedell, Crystal, "Information Security Threats – Malware". https://searchsecurity.techtarget.com/definition/worm
[4] Firch, Jason, "9 Common Types of Malware". https://purplesec.us/common-malware-types/
[5] Grimes, Robert, "9 Types of Malware and how to recognize them". https://www.csoonline.com/article/2615925/security-your-quick-guide-to-malware-types.html
[6] Constntin, Lucien, "What is adware? How it works and how to protect against it. https://www.csoonline.com/article/3406422/what-is-adware-how-it-works-and-how-to-protect-against-it.html
[7] Letteri Ivan, "MTA-KDD'19: A Dataset for Malware Traffic Detection". https://github.com/IvanLetteri/MTA-KDD-19/blob/master/ITASEC2020.pdf
[8] Heena, B. M. Mehtre: "Advances In Malware Detection-An Overview". https://arxiv.org/abs/2104.01835

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING

Author(s)
Advisors(s)