An Evaluation of Prominent Commercial Malware Detector

Class 2I

Team

Al Azhar Rizqi R. F
01

Lenka Melinda F.
12

Malik Abdul A.
13

Background

- Malware (malicious software) poses a significant threat to computer systems and data security.
- Commercial malware detectors, including antivirus software, are crucial for defending against malware attacks.
- There is limited independent research evaluating the effectiveness of commercial malware detectors.
- The U.S. Navy initiated the Artificial Intelligence Applications to Autonomous Cybersecurity (AI ATAC) prize challenges to address this need for evaluation

Problem Formulation

- The effectiveness of commercial malware detectors is not well-understood due to limited research.
- Existing evaluations often lack depth or have limitations in terms of sample size and diversity.
- The challenge is to comprehensively evaluate the performance of prominent commercial malware detectors.

Research Objectives

1. Evaluate six prominent commercial endpoint malware detectors, a network malware detector, and a file-conviction algorithm from a cyber technology vendor.
2. Conduct the evaluation as part of the AI ATAC prize challenges funded by the U.S. Navy.
3. Use a dataset of 100,000 files, balanced with 50% benign and 50% malicious samples, including approximately 1,000 zero-day program executables.

Research Objectives

4. Develop a rigorous evaluation process that involves delivering files to virtual machines, allowing static detection, executing the files, and observing dynamic detection.
5. Consider resource usage and time-to-detection statistics in the evaluation.
6. Design a software framework to automate and parallelize the experiment.
7. Create a cost-benefit model to integrate recall, precision, time to detection, and resource requirements, enabling a ranking methodology for cyber competitions.
8. Provide insights into the state of commercial malware detection and highlight areas for improvement.

Related Research

Although the study doesn't mention related research, it can be assumed that it builds upon prior research in the field of cybersecurity, particularly in the area of malware detection and security tools evaluation.

Proposed Method

- Framework Design: The researchers would have started by creating a systematic framework for their evaluation. The actions, practices, and metrics that will be utilized to evaluate the effectiveness of the commercial malware detectors are described in this framework.
- Malware Samples: A wide range of malware samples would have been gathered by the researchers. For a thorough assessment, these samples most certainly contain a variety of kinds (such as viruses, worms, and Trojan horses) and variations.
- Include Normal Data: Normal or benign data examples would have been provided in addition to malware samples. This makes it possible to evaluate the detectors' ability to discriminate between harmful and benign files in a balanced manner.

Proposed Method

- Test settings and Use Cases: To replicate various real-world settings, the researchers would have established several test scenarios or use cases. To test the detectors' performance in various scenarios, they may, for example, create a scenario involving email attachments, online downloads, or USB devices.
- Execution of Tests: In the actual trials, the described test scenarios would be executed by commercial malware detectors. In order to determine if the samples are malicious or benign, the detectors would examine the supplied samples.

Proposed Method

- Performance Analysis: Following the completion of the studies, the researchers would have examined the data to identify the advantages and disadvantages of every commercial malware detector. Comparing detection rates, false positives, and false negatives in various settings may be part of this study.
- Validation and Reproducibility: To make sure the given approach is reliable, it probably underwent validation. To ensure that the outcomes are consistent, this could need doing the trials more than once.

Proposed Method

- Measurements and metrics:

1. Detection Rates: This statistic shows the proportion of real malware samples that the detectors properly classify as harmful.
2. False Positive Rates: This measures the proportion of legitimate data that is mistakenly labeled as malware.
3. False Negative Rates: This statistic shows the proportion of real malware that the detectors miss.

Data used

A collection of malware samples gathered from multiple sources is used in the case. This dataset probably includes a wide range of malware samples to test how well commercial malware detectors can distinguish between various threats.

Experiments Conducted

The tests entail running various commercial malware detectors on the dataset of malware samples. The performance of these tools is assessed based on their capacity to recognize and respond to the dangers included in the dataset.

Potential Developments

From the journal, potential developments include:

- Improved malware detectors, focusing on recall rates, cost-efficiency, and adaptability.
- Advancements in machine learning models for better zero-day threat detection.
- Tailored security strategies based on variations in detection capabilities.
- Further evaluations with diverse datasets for comprehensive assessments.
- Optimization of resource usage to reduce costs.
- Enhanced transparency and accountability in cybersecurity.
- Policymakers considering findings for regulation and standards.
- More comparative studies on malware detection tools.
- Enhanced cybersecurity training to educate employees.

Result

In this endpoint malware experiment:

- Detection costs significantly outweigh initial and resource costs.
- Precision is near perfect for all tools.
- Differentiation comes from recall and detection time.
- Baseline 2 performs well, with high recall and fast detection.
- Median detection times vary significantly among tools.
- Network-based ML detector achieves high recall but has latency and drawbacks.
- Algorithm-based detector has good recall and low latency.
- ML-based tools excel at detecting zero-day malware.
- Tool efficacy varies across different file types.
- Baseline 2 is cost-effective, with low annual malware cost ($204K).

Advantages

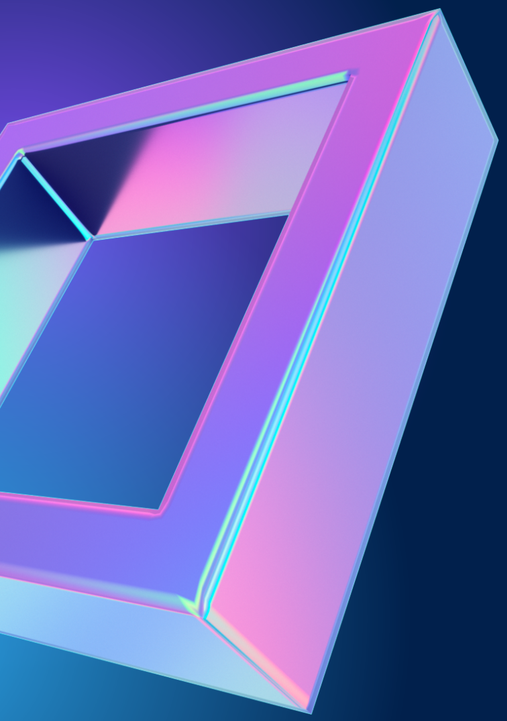
1. Relevance to cybersecurity.
2. Practical experiments with a substantial dataset.
3. Comparison of detection technologies.
4. Precise metrics and cost-benefit analysis.
5. Highlighting variation in detection capabilities.

Disadvantages

1. Limited scope (focus on commercial tools).
2. Dependency on the dataset used.
3. Potential bias due to Navy funding.
4. Lack of specific tool details.
5. Incomplete recommendations for improvement.

Summarize

This article evaluates commercial malware detectors, funded by the US Navy, using 100,000 files, including zero-day executables. It assesses precision, recall rates, and costs. Results show high precision, better recall for machine learning tools, and varied detection across file types, emphasizing the need for better evaluation and tool enhancement.



Thank You