Prudent Practices for Designing Malware Experiments: Status Quo and Outlook

Jeff Malavasi
Dept. of Computing Security
Rochester Institute of Technology
Email: jm3378@rit.edu

I. Review

The authors of the paper aim to create a framework that can be used to accurately and safely evaluate malware. They then analyze the work of previous researchers in order to show that the communities lack of standardization can drastically limit its ability to effectively study malware.

The framework is broken down into four categories: correct datasets, transparency, realism and safety. The researchers found that some of the most important factors include removing benign samples from the dataset, investigating errors, and testing in a real world, contained environment. Additionally, they found that datasets should contain a variety of samples from various families. Authors should rationalize why specific samples were omitted or selected and should test them on a variety of systems [1].

After defining the framework, the authors used their guidelines to score 36 previous experiments based on a qualitative assessment. Multiple scorers were used to remove rater bias. Experiments were given the benefit of the doubt and the authors noted that not all characteristics are relevant to every experiment. It was found that 25% of the reviewed papers lacked the necessary criteria to ensure an accurate result [1]. Additionally, not a single paper separated datasets or removed legacy samples. The authors found that separating training and testing data by malware family was necessary to ensure accuracy. Removing outdated binaries was shown to improve realism by reducing the false negative rate of the model. Another notable finding was that many of the experiments lacked repeatability, since authors did not include a description of the test environment or operating system. Finally, experiments also omitted real world testing scenarios and those that did lacked the necessary safety precautions to properly contain the sample. This can lead to research that may cause more harm than good.

As has been demonstrated by the authors, studying malware with a scientific framework, greatly improves the accuracy of the research. The authors noted that building a comprehensive and diverse dataset was vital to removing bias. Finally, they showed that properly building a test environment requires real world analysis, containment and monitoring to establish safety.

References

[1] C. Rossow, C. Dietrich, C. Grier, C. Kreibich, V. Paxson, N. Pohlmann, H. Bos, and M. van Steen, "Prudent practices for designing malware experiments: Status quo and outlook," in *IEEE*, 2012.