Spotless Sandboxes: Evading Malware Analysis Systems Using Wear-and-Tear Artifacts

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

I. Review

The authors of the paper propose a novel sandbox evasion technique that aims to detect wear-and-tear artifacts on the test system. As the use of dynamic malware analysis expands, malware authors are adding code blocks to detect sandboxes in order to appear benign during classification and testing. Typically, bad actors identify sandboxes by investigating features of the underlying hardware and operating system. However, in this paper the authors show that by analyzing factors that indicate how much the machine has been used, such as browser and file activity, can be 92% effective at detecting sandboxes. Finally, the researchers show how they can even use these factors to predict the age of a machine, which hopefully can be used in the future to artificially age a sandbox machine to prevent evasion [1].

The researchers begin by determining which factors could be used to identify that a system has been used by a real operator, versus more indirect artifacts of sandbox activity. The artifacts were broken down into five categories: system, disk, network, registry and browser [1]. In order to determine which factors would be most effective, the researchers built a probe tool written in C++. The tool was designed to not collect any PII and was ran against 270 real user machines and 23 sandbox environments. However, only 15 sandboxes were able to return data back to the researchers, so the remaining were removed from analysis. Generally, they found that real user systems had broader and larger bands for the majority of the tested artifacts.

The data was fed into a decision-tree model which selected the best features. The authors found that the model could still achieve greater than 90% accuracy with as few as 29 features [1]. Additionally, they found that the network and disk based classifiers were the most effective, however this can vary a bit depending on the underlying containment policies of a sandbox.

Finally, the researchers wanted to determine if they could accurately ”age” a machine based on the presence of wear-and-tear artifacts. They found that they could not only do this with precision, but they could reference data from other machines to determine if the age of the machine was artificially tampered with [1]. The authors hope that this data could be used to proactively age sandbox machines in order to make it harder for malware authors to use wear-and-tear artifacts for evasion.

Overall the authors were extremely successful in showing that wear-and-tear artifacts can be used to evade sandboxes and even to determine the age of a specific machine. Their solution was applicable to many different types of sandboxes including virtualized, bare metal and emulated. However, there were also a few limitations. First of all, although they were able to achieve high accuracy, the initial dataset was extremely small. To further support the effectiveness of the model, the researchers should continue to sample real user machines as well as incorporate other operating systems into their model. Finally, while they propose that sandboxes should artificially age their test machines, they do not measure how much of a difference this would make in the real world.

Malware authors are constantly looking for new ways to evade sandboxes in order to be falsely classified as benign. The researchers create a novel system to achieve sandbox detection as well as a way to potentially subvert this attack. Future research will be needed to expand their model to other types of machines.

References

[1] N. Miramirkhani, M. P. Appini, N. Nikiforakis, and M. Polychronakis, “Spotless sandboxes: Evading malware analysis systems using wear-and-tear artifacts,” in *2017 IEEE Symposium on Security and Privacy (SP)*, 2017, pp. 1009–1024.