D4.2 - NGSOTI Data collection blog post

NGSOTI Project: 101127921
DIGITAL-ECCC-2022-CYBER-03 D4.2

Team CIRCL/NGSOTI

2024-03-14

Contents

List of Figures

1 D4.2 - NGSOTI Data collection blog post

1.1 Disclaimer

1.2 Distribution and License

1.3 Deliverable Definition

The identifier of the deliverable is D4.2 and it adheres to the definition outlined in the grant agreement **Public report with key findings of data collected in NGSOTI such as new discoveries, high level statistics to attacked schools to use NGSOTI**. The deliverable name is **(D4.2) - NGSOTI data collection blog post** and the overall objective/alignment is described in the executive summary.

This deliverable corresponds to a blogpost published on: https://www.d4-project.org/2024/10/02/Enhancing-Detection-Engineering-with-Automated-Malware-Sandboxing.html

2 Enhancing Detection Engineering with Automated Malware Sandboxing

2.1 Introduction

In the complex field of incident response, effective training for Security Operations Center (SOC) operators is critical. One of the key challenges in SOC training is providing realistic, data-driven environments that accurately simulate the threats and incidents operators will face. Additionally, detection engineers need reliable and actionable data to create robust detection rules that align with real-world security monitoring systems. However, gathering and analyzing real-world malware samples, which is essential to this process, can be time-consuming and prone to errors when done manually. In this blog post, we introduce an approach to solving these challenges through automation. We explore how a Kunai-based sandbox can streamline the collection and analysis of malware samples, offering

a practical solution. By leveraging this sandbox infrastructure, the project opens up new opportunities for more efficient malware analysis while supporting a wide range of CPU architectures, including those specific to IoT and mobile devices.

2.2 The Need for Realistic Data

One prerequisite for offering cyber ranges or training solutions in the context of detection engineering and security monitoring is the collection of real-world malware samples. To provide high-quality training and realistic experiences, these samples can be used as injects in various training scenarios or for testing detection rules.

A common approach is to collect such data manually by running and monitoring malware samples, preferably in a confined environment such as a virtual machine (VM). However, this approach has several drawbacks: it lacks reproducibility under identical experimental conditions and involves repetitive, error-prone tasks (uploading files, running monitoring tools/malware samples, monitoring network traffic, conducting post-analysis, etc.). Thus, this process is an ideal candidate for automation. Our first motivation for creating this new project is to address these challenges. Our second goal is to provide detection engineers with a reliable way to **generate actionable** data from malware samples.

2.3 The Concept of a Kunai-Based Sandbox

Malware sample sandboxing is a frequent task performed at various stages of a security alert's lifecycle, from incident/malware triage to more detailed malware analysis. This task is typically supported by numerous tools, ranging from open-source options like Cape Sandbox to paid alternatives like Joe Sandbox, VMRay, or Any Run. While these solutions are excellent in many respects—such as defeating anti-sandboxing techniques and providing deep insight into a sample's capabilities—we believe they are not always the best tools for gathering actionable information for detection engineers. For many organizations, there is no direct mapping between the data collected from malware analysis platforms (sandboxes) and their monitoring systems. As a result, a task that should be simple—building detection rules tailored to an organization's security monitoring tools—can become challenging. To solve this issue, we propose a simple yet powerful sandboxing infrastructure based on QEMU for virtualization and Kunai for sample monitoring. This infrastructure can serve multiple purposes: analyzing malware samples using the same tools employed for

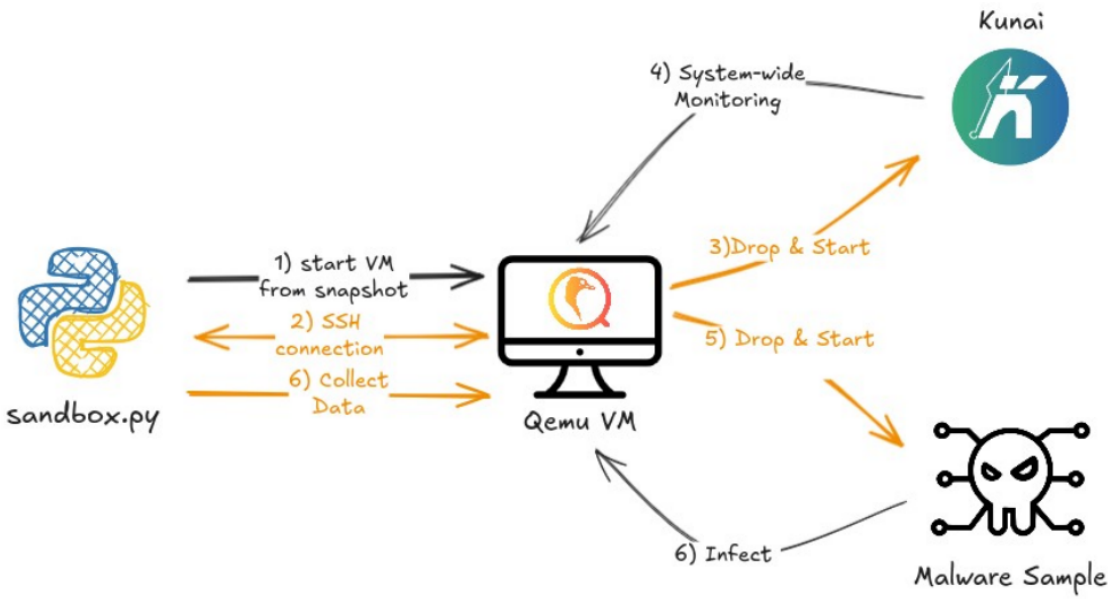

Figure 1: Kunai sandbox process

The image depicts kunai-sandbox process

2.4 Project Status

The source code for the project is available in the Kunai sandbox repository. Additionally, our under-construction open dataset, extracted using this sandbox, can be found at NGSOTI malware dataset. Currently, the sandboxing system can run Linux malware samples within a virtual environment, monitor them using Kunai, and capture the network traffic generated by the system. Another key feature of our sandbox is its support for multiple CPU architectures (currently **Intel 32/64bits** and **ARM 64bits**), enabling the analysis of a broader range of malware samples. We believe **ARM** achitecture support is crucial, as it can be used to analyze malware samples specific to **IoT** or **mobile** (phones, tablets, etc.) devices.

2.5 Limitations

While our approach provides a great opportunity for detection engineers to obtain data that is directly usable for creating Kunai-based detection rules, we must remember that it does not achieve the same level of stealthiness as other sandboxing platforms, which often rely on custom hypervisors. Therefore, our approach should not be considered a replacement for dedicated sandboxing platforms but rather a complement that facilitates detection engineering-related tasks.

3 Conclusion

The NGSOTI project aims to bridge the gap between theoretical knowledge and practical skills for SOC operators by offering realistic, data-driven training experiences. By automating the collection and analysis of malware samples through the Kunai-based sandbox, we provide a straightforward, efficient, and repeatable method for detection engineers to generate actionable insights. This approach is not intended to replace traditional sandboxing but rather to complement it. With support for multiple CPU architectures, including those specific to IoT and mobile devices, the sandbox expands the possibilities for analyzing and generating data from a wider range of malware, enhancing the diversity of scenarios that NGSOTI can offer. As the project progresses, we look forward to further enriching the open dataset and continuing to develop solutions that address the evolving challenges in detection engineering.