SubParse: Malware Artifact and Correlation Framework

Aaron Baker / Quentin Covert / Odin Bernstein / Dr. Josh Stroschein

Introduction

The goal of this research is the development of an open-source, novel, modular framework for use in the automated analysis of malware, using static and dynamic analysis, as well as data correlation from threat intelligence feeds. These features will allow a malware analyst to gather the most current and accurate information available regarding a malware sample while minimizing the effort needed to collect and aggregate that data. The modular design of the framework will allow users to customize the framework for their own purposes and needs.

Problem Definition

The speed at which malware can titrate across the internet is incredible. However, the rate of analyzing the samples found is relatively slow in comparison and, for the most part, will rely on extensive manual effort from the analyst. Over the past few years, an increase in the amount of malware being developed has been noted. Because malware analysts are often forced to analyze malware by hand, this means that many are being stretched thin. Thus, the development of tools that can automate, in whole or in part, the process of malware analysis is vital. This framework is such a tool.

Literature Review

While other projects available to the public do provide some of the functionality provided by SubParse, each is designed and limited to specific use cases and lacks the modularity required for an all-purpose tool. These other services include Suricata, CHI-RON ELK, and sqhunter, among others. Each of these tools exists for the purposes of network analysis and the determination of what threats exist on a network. Other tools such as YARA, Intrigue Core, and Xray aid in the identification of malware, reconnaissance, and information gathering. SubParse attempts to bridge the gap between these tools, offering the functionality of these systems in a singular system.

```
2022-04-19 09:36:00,981 - SubParser - DEBUG - [CUCKOOENRICHER] STARTING
2022-04-19 09:36:00,981 - SubParser - INFO - Path for file: /home/dsu/Downloads/subparse-beta-var-to-let/files/9cdcc531878d3a23d2da8
33a058efa100b91d212e4d5780cb21d5d36db0cbd01.exe
2022-04-19 09:36:00,985 - SubParser - INFO - Cuckoo has finished processing all tasks successfully!
2022-04-19 09:36:00,985 - SubParser - DEBUG - [CUCKOOENRICHER] FINISHED
2022-04-19 09:36:00,986 - SubParser - INFO - PROCESSING [34/36]: c53a2c5d825591d469aa2cc81d3bf02d
2022-04-19 09:36:00,990 - SubParser - DEBUG - Additional Parsing Needed: Not In Elastic
2022-04-19 09:36:00,992 - SubParser - DEBUG - [ALTParser] STARTING
2022-04-19 09:36:00,992 - SubParser - DEBUG - [ALTParser] FINISHED
2022-04-19 09:36:00,992 - SubParser - DEBUG - [PEParser] STARTING
2022-04-19 09:36:01,012 - SubParser - DEBUG - [PEParser] FINISHED
2022-04-19 09:36:01,012 - SubParser - DEBUG - [CUCKOOENRICHER] STARTING
2022-04-19 09:36:01,012 - SubParser - INFO - Path for file: /home/dsu/Downloads/subparse-beta-var-to-let/files/5d518f12a0fbc2de07b59
a89e2f36edfad280b3c1a2f1bb4b0cdb0f98ee10bd4.exe
2022-04-19 09:36:01,016 - SubParser - INFO - Cuckoo has finished processing all tasks successfully!
2022-04-19 09:36:01,016 - SubParser - DEBUG - [CUCKOOENRICHER] FINISHED
2022-04-19 09:36:01,018 - SubParser - INFO - PROCESSING [35/36]: 580d2373c37e5781fda8f1c9495e9c4d
2022-04-19 09:36:01,022 - SubParser - DEBUG - Additional Parsing Needed: Not In Elastic
2022-04-19 09:36:01,028 - SubParser - DEBUG - [ALTParser] STARTING
2022-04-19 09:36:01,028 - SubParser - DEBUG - [ALTParser] FINISHED
2022-04-19 09:36:01,028 - SubParser - DEBUG - [PEParser] STARTING
2022-04-19 09:36:01,116 - SubParser - DEBUG - [PEParser] FINISHED
2022-04-19 09:36:01,116 - SubParser - DEBUG - [CUCKOOENRICHER] STARTING
2022-04-19 09:36:01,117 - SubParser - INFO - Path for file: /home/dsu/Downloads/subparse-beta-var-to-let/files/12ac133adac3300ae93a8
ed238cc19d4c790c2c710549f9732510fccf8ddf40.exe
2022-04-19 09:36:01,121 - SubParser - INFO - Cuckoo has finished processing all tasks successfully!
2022-04-19 09:36:01,121 - SubParser - DEBUG - [CUCKOOENRICHER] FINISHED
2022-04-19 09:36:01,124 - SubParser - INFO - PROCESSING [36/36]: ea6de27b46dcb75eb35152d2ef0143bd
2022-04-19 09:36:01,142 - SubParser - DEBUG - Additional Parsing Needed: Not In Elastic
2022-04-19 09:36:01,150 - SubParser - DEBUG - [ALTParser] NOT COMPATIBLE
2022-04-19 09:36:01,150 - SubParser - DEBUG - [PEParser] NOT COMPATIBLE
2022-04-19 09:36:01,150 - SubParser - DEBUG - [CUCKOOENRICHER] STARTING
2022-04-19 09:36:01,150 - SubParser - INFO - Path for file: /home/dsu/Downloads/subparse-beta-var-to-let/files/f32aaffbbf2bcaee36c60
1364a794939abf6039613fea0ff833cb48a113eca2c3.unknown
2022-04-19 09:36:01,154 - SubParser - INFO - Cuckoo has finished processing all tasks successfully!
2022-04-19 09:36:01,154 - SubParser - DEBUG - [CUCKOOENRICHER] FINISHED
2022-04-19 09:36:01,154 - SubParser - DEBUG - [ELASTIC] Pushing batch sample data
2022-04-19 09:36:01,165 - SubParser - INFO - Skipping zip creation
2022-04-19 09:36:01,166 - SubParser - INFO - Finished Processing
dsu@dsu-virtual-machine:~/Downloads/subparse-beta-var-to-let/parser$
```

Scope

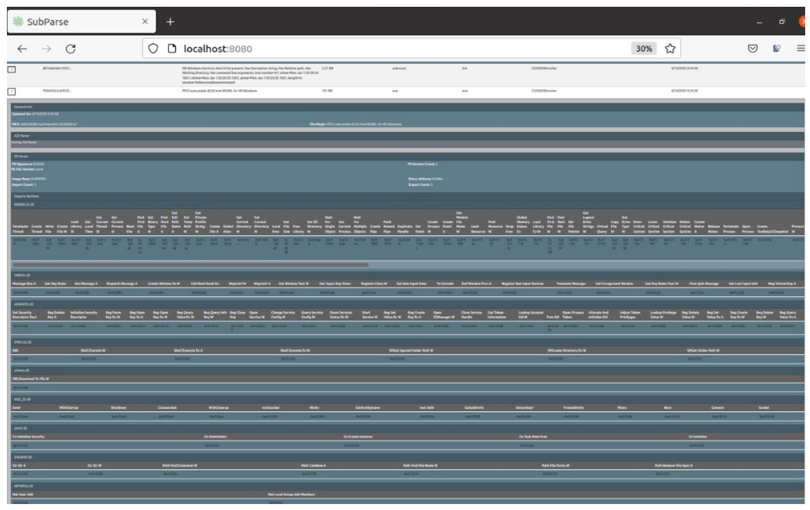
The scope of this project is the development of the basic framework of SubParse, with at least two “parsing engines” that extract information from specific file types and at least two “enrichment engines” which collect information from external sources.

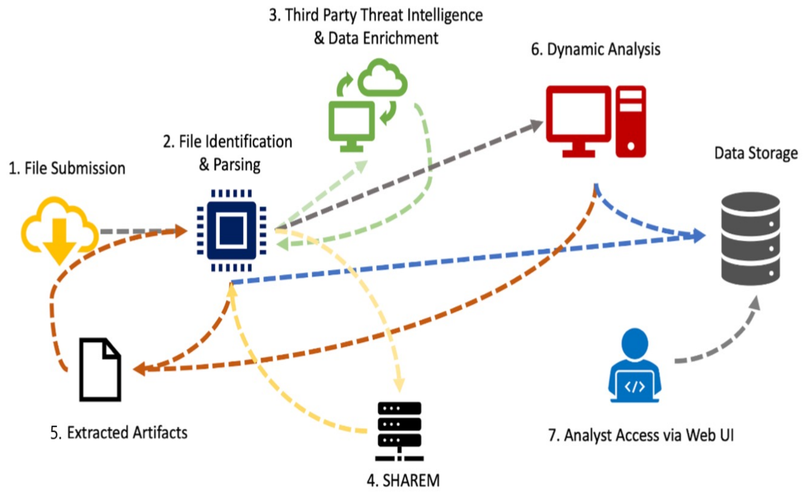
Objectives

- Develop the central framework of SubParse which acts as an engine for calling the modules and presenting the information in a searchable, corelatable manner.
  - This includes the logic for identifying file types, calling parsing engines, calling enrichment engines, and presenting the information in a VUE web instance.
- Develop a small number of modules for parsing very common kinds of malware. These file types currently include:
  - PEParser, which parses Portable Executable (PE) files.
  - ELFParser, which parses Executable and Linkable Format (ELF) files.
- Develop a small number of enrichment modules for information gathering. These currently include:
  - AbuseEnricher, which pulls information from the Abuse.ch API.
  - CuckooEnricher, which pulls information from a Cuckoo Sandbox for dynamic analysis.

Anticipated Results

The anticipated result of this project is the completion of the basic SubParse framework and the dissemination of the tool via open-source distribution. The current results from SubParse are encouraging. The tool allows for the parsing and enrichment of thousands of malware samples in a matter of minutes, doing what would likely take hours or days of work in minutes and presenting the findings in a visual, searchable manner.





SubParse Framework

File Submission: The user submits files for parsing and analysis. This can be done in either a “one-shot” style or through the use of a file folder being watched by SubParse. The user will also tell SubParse which enrichment engines to use during this step.

File Identification & Parsing: SubParse identifies each file’s type and calls the appropriate parsing engine, which collects information relevant to the file. This currently supports PE and ELF files.

Third Party Threat Intelligence & Data Enrichment: Any enrichment engines called at the time of file submission are then run. These engines will collect information about each sample through the third party APIs they are designed for.

Extracted Artifacts: Information gathered during parsing and enrichment are combined into a single entry and added to an Elastic Search database.

Dynamic Analysis: SubParse is capable of launching and using an instance of Cuckoo/CAPEv2 for the automatic, dynamic analysis of malware samples, should the analyst choose to use this feature.

Analyst Access via Web UI: Analysts may then access the information gathered through a VUE web UI, which allows for dynamic searching on any field found during the parsing and enrichment.

Conclusion

SubParse is able to generate in-depth results for different types of files such as portable executable (PE) files. Currently, with the help of external threat intelligence fields, SubParse is able to aid in the data correlation efforts of analysts to identify the impact of samples it parses. With the use of our dynamic framework, we have been able to successfully speed up the time that was previously required for analysts to manually collect information in both static and dynamic ways and extended it to fit common analysis needs.

DAKOTA STATE UNIVERSITY
MADRID LAB