Intro

* Why are we doing this
* Why we used cuckoo
* How we used cuckoo

Python Scripts

We needed an executable to give to Cuckoo to process and give us the raw data. The first problem we encountered was Cuckoo not accepting our executable. We used pyinstaller [2] to turn our script into a .exe, but Cuckoo gave errors because of the python version, and at the time, we used python 3.11. The highest python version that cuckoo can use is python version 2.7 [3]. After going to python 2.7, we lost packages that were not available for python 2.7.

Our solution was to use python 2.7 and simplify our python script and disregard turning it into a .exe but give it to cuckoo as a .py file.

Data (Cuckoo Logs, PCAP Logs)

After giving cuckoo our script, cuckoo gave us the three files, analysis log, behavioral log, and PCAP log.   
Analysis logs are the logs that cuckoo generates when a file, processes, and errors are created; behavioral logs are analysis logs in more detail; and PCAP logs have the network traffic.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 1.1 Raw Analysis Log

Calendar

Description automatically generated with low confidence

Figure 1.2 Raw PCAP Log

Figure 1.3 Raw Behavioral Log

Given this data, we need to process and clean up for the machine learning algorithm to make sense of the data.

Processed Data

To process our analysis data, we had to filter anything that was not a valid path. Using regex to filter and save a txt file it was then possible to turn the paths as a term for TF-TDF.

Text

Description automatically generated

Figure 2.1 New Analysis Log

We decided that past encryption was unnecessary since we are trying to warn when it detects a virus before encryption. Thus, we removed data when it showed signs of when WannaCry encrypted.

Text

Description automatically generated

Figure 2.2 Pre-Encryption Log

Since WannaCry’s encrypted files have the file extension .WNCRYT, we used regex to find when we encountered .WNCRYT and stop appending it to the new log, Pre-Encryption log.

We noticed that each path is different for every scenario, we decided on two ways to combat this, one where we generalize the paths and to remove the paths completely. Figure 2.3 and Figure 2.4 are generalized and removed logs respectfully.

Text

Description automatically generatedText

Description automatically generated

Figure 2.3 Generalized Log

For our PCAP data, we needed to

Classifiers

Predictions & Thresholds (Tenfold)

Performance

References

[1] https://cuckoosandbox.org

[2] https://pyinstaller.org/en/stable/

[3] https://cuckoo.readthedocs.io/en/latest/installation/host/requirements/

[4]