Decoding GAN-Generated Malware using Explainable AI Techniques

Matteo Tiozzo • 3rd week 2024/10/01 - 2024/10/08

Overview

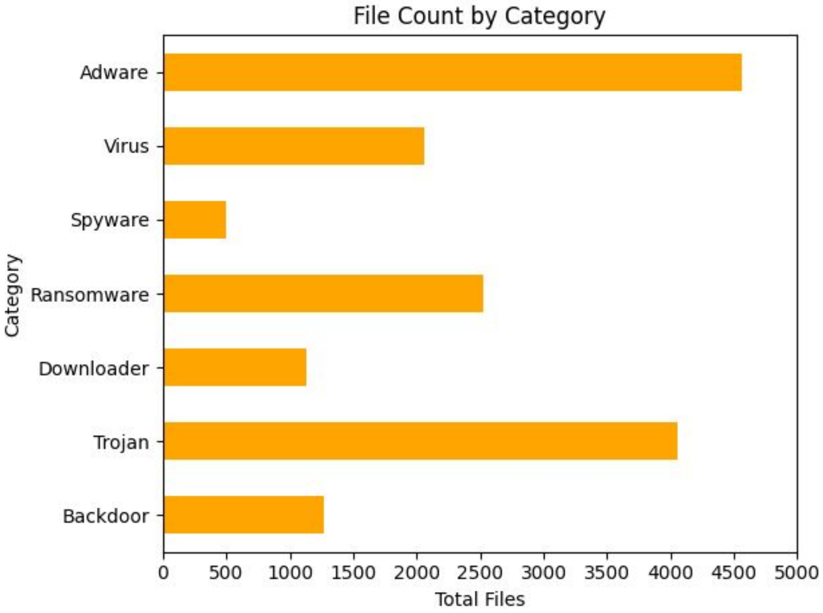
Progress

- Completed the disassembly of malware into assembly and hexadecimal code
- Extracted mnemonic instructions for each malware sample in assembly code
- Generated grayscale images
- Removed malware families with fewer than 500 samples

Major risk

- Understanding the logic behind extracting mnemonic instructions

Completed the classification of Windows malware executables



Total number of Windows malware executables downloaded

- 16114

Size of executable files

Category	Total Files	Smallest File Size (bytes)	Largest File Size (bytes)	Average File Size (bytes)
Adware	4566	2048	60801310	2276310.16
Backdoor	1270	1004	45567280	1473530.68
Downloader	1128	2201	44408472	1263060.26
Ransomware	2529	4096	16634880	611795.27
Spyware	502	3072	10961920	744435.38

Size of executable files

Category	Total Files	Smallest File Size (bytes)	Largest File Size (bytes)	Average File Size (bytes)
Trojan	4058	1024	59399864	1035019.17
Virus	2061	5616	46137193	1358951.80

Date malware uploaded to respective sites

Date	Total Files	Source
2016	7239	VirusShare
2021	2888	MalShare
2021	4	Malware Bazaar
2024	5983	VirusShare

Tools used this week

Python

- To extract mnemonic instruction from assembly and to count them
- To complete malware disassembly
- To generate graphs
- To create grayscale images

Ghidra

- To disassemble the malwares