Decoding GAN-Generated Malware using Explainable AI Techniques

Matteo Tiozzo • 2nd week 2024/10/01 - 2024/10/08

Overview

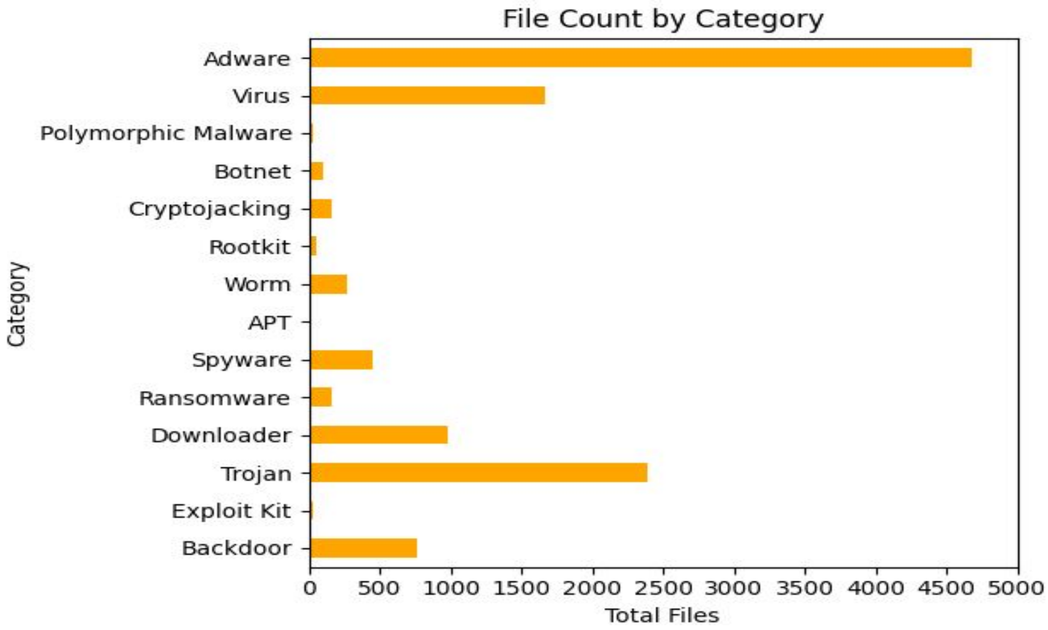
Progress

- Finished classification of windows malware executables
- Disassembled windows malware executables and obtained assembly and hexadecimal code

Major risk

- Use of Ghidra as a new and never used tool

Finished classification of windows malware executables



Total windows malware executable downloaded

- 12538

Size of executable files

Category	Total Files	Smallest File Size (bytes)	Largest File Size (bytes)	Average File Size (bytes)
Backdoor	770	1004	45567280	869437.39
Exploit Kit	35	10240	5267459	802590.11
Trojan	2391	1024	74521669	1189001.06
Downloader	987	1164	44408472	1183877.82
Ransomware	162	15072	12552896	568090.28

Size of executable files

Category	Total Files	Smallest File Size (bytes)	Largest File Size (bytes)	Average File Size (bytes)
APT	7	21520	1110624	292162.00
Worm	275	2560	51541088	1430851.33
Rootkit	55	2784	7535776	383208.78
Cryptojacking	159	22016	27193824	978693.14
Botnet	106	1789	20228444	697703.65

Size of executable files

Category	Total Files	Smallest File Size (bytes)	Largest File Size (bytes)	Average File Size (bytes)
Polymorphic Malware	33	19968	9308981	704173.73
Virus	1673	2363	46137193	1457855.05
Adware	4676	2048	60801310	2260712.84
Spyware	456	3577	21699976	774230.16

Date malware uploaded to respective sites

Date	Total Files	Source
2021	5144	MalShare
2013	997	VirusShare
2016	7604	VirusShare
2021	123	Malware Bazaar
2024	+10000	VirusShare

Tools used for classification of Windows executable malware

Python

- To download malware
- To classify malware superfamilies and subfamilies
- To create directory and subdirectory for each malware
- To create graphs and tables

AVClass

- A command line tool to tag / label malware samples starting from VirusTotal JSON report as input

Tools used for classification of Windows executable malware

VirusShare, MalShare and Malware Bazaar

- To download the malware dataset

VirusTotal

- To generate JSON report for each malware

Tools used to disassemble Windows executable malware and obtain assembly and hexadecimal code

Python

- To grab malware executables and pass them to Ghidra
- To save assembly and hexadecimal code

Ghidra

- In headless mode, to disassemble executable files in assembly code and hexadecimal

Currently

This is what I am currently doing:

- Downloading and labeling additional malware samples
- Learning about AI architecture (Xception, Inception, ecc..)