Mac Malware Analysis

Frederick Berberich, Daniella Boulos, Julianna Russo, Christopher Drisdelle

MARIST

Overview

The Mac Malware project focuses on malware more specifically Backdoor (Trojan) viruses and the minimal ability to detect them on Apple computers.

Throughout the project, we ran samples of Backdoor malware which were then collected and analyzed using SpriteTree. We are developing a machine learning (ML) algorithm designed to differentiate between benign and malicious files, identifying logs that may pose a risk to devices. This is done by feeding the ML algorithm small amounts of data and then testing it to see if it can predict whether or not the file is malicious.

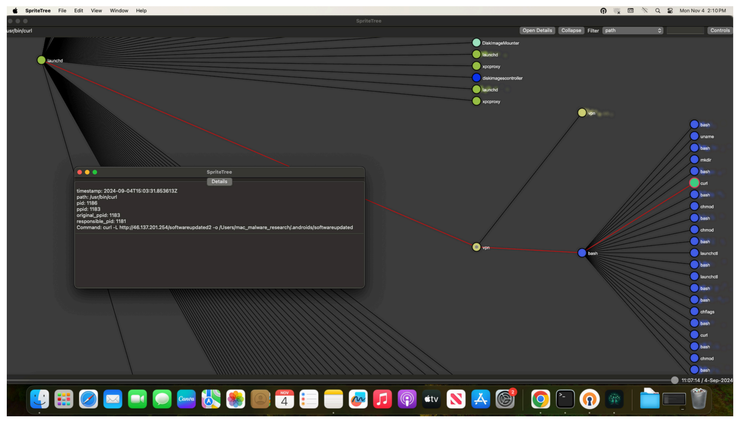
Test Bed

We used a secure testbed for data collection and analysis that consists of an Apple Mac Mini 2018 with macOS Sonoma, a Ubiquiti Dream Machine Pro Firewall, and various analytical software. This test bed allows for the deployment of various MacOS Backdoors so that malicious logs can be collected once the malware is executed.

Software used on the Mac Malware Project are:

- EsLogger
- Jupyter Notebook
- SpriteTree

EsLogger was used to collect all the files and folders edited or created from malware and converted this data into .JSON files. We then converted this data into .CSV files. Afterwards, we analyzed these files using SpriteTree and Jupyter Notebook.



Data Collection

We implemented a standardized data collection process to ensure that we collected data without impacting other devices on the network, the data was able to be cleansed for the machine learning model, and manual malware analysis could be conducted. The steps we took were:

1. **Download files**: We downloaded a variety of Backdoor malware from the Objective-See website to run and observe how they affected the system.
2. **Run malware**: After the malware was downloaded, we sandboxed the environment by deep freezing the Mac Mini before running the malware samples and then restarting the computer before continuing onto a new sample. The data that we collected was automatically documented in .JSON files using EsLogger.
3. **Analyzing samples**: Once we ran the malware samples, we analyzed them in SpriteTree, which allowed us to determine the similarities and differences between the different malware samples.
4. **Converting to .CSV**: The files were recorded in .JSON files and were unreadable so, using a Python script, we converted them to .CSV files to allow us to observe and clean the data.
5. **Cleaning data**: A lot of the data we collected had null values which could result in inaccurate conclusions when implementing the machine learning algorithm. Therefore, we narrowed down the data from over 1,000 columns to 80 columns.

Currently, we are going through the data and determining whether or not certain columns repeat with similar data sets. If this is the case, then we would need to trim the data even more.

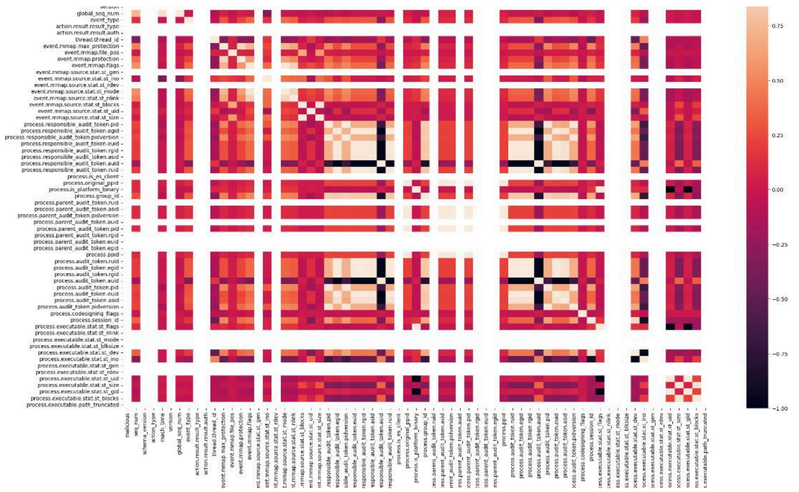
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	malicious	seq_num	schema_version	action_type	mach_time	version	global_seq_num	time	event_type	action.result.resu	action.result.resu	thread.thread_id	event.mmap.max_	event.mmap.file_
2		0	0	1	1	277763000000	7	0	2024-03-25T18:	20	0	0	30715	1	0
3		0	1	1	1	277763000000	7	1	2024-03-25T18:	20	0	0	30715	1	0
4		0	2	1	1	277763000000	7	2	2024-03-25T18:	20	0	0	30715	1	0
5		0	3	1	1	277763000000	7	3	2024-03-25T18:	20	0	0	30715	1	0
6		0	0	1	1	277763000000	7	4	2024-03-25T18:	13	0	0	30715	0	0
7		0	1	1	1	277763000000	7	5	2024-03-25T18:	13	0	0	30715	0	0
8		0	0	1	1	277763000000	7	6	2024-03-25T18:	25	0	0	30715	0	0
9		0	4	1	1	277770000000	7	7	2024-03-25T18:	20	0	0	30715	1	0
10		0	5	1	1	277770000000	7	8	2024-03-25T18:	20	0	0	30715	1	0
11		0	6	1	1	277770000000	7	9	2024-03-25T18:	20	0	0	30715	1	0
12		0	7	1	1	277770000000	7	10	2024-03-25T18:	20	0	0	30715	1	0
13		0	2	1	1	277770000000	7	11	2024-03-25T18:	13	0	0	30715	0	0
14		0	3	1	1	277770000000	7	12	2024-03-25T18:	13	0	0	30715	0	0
15		0	4	1	1	277772000000	7	13	2024-03-25T18:	13	0	0	31438	0	0
16		0	1	1	1	277772000000	7	14	2024-03-25T18:	25	0	0	30715	0	0
17		0	5	1	1	277775000000	7	15	2024-03-25T18:	13	0	0	31438	0	0
18		0	2	1	1	277775000000	7	16	2024-03-25T18:	25	0	0	31438	0	0
19		0	8	1	1	277775000000	7	17	2024-03-25T18:	20	0	0	30715	1	0
20		0	9	1	1	277775000000	7	18	2024-03-25T18:	20	0	0	30715	1	0
21		0	10	1	1	277775000000	7	19	2024-03-25T18:	20	0	0	30715	1	0
22		0	11	1	1	277775000000	7	20	2024-03-25T18:	20	0	0	30715	1	0
23		0	6	1	1	277775000000	7	21	2024-03-25T18:	13	0	0	30715	0	0
24		0	7	1	1	277775000000	7	22	2024-03-25T18:	13	0	0	30715	0	0
25		0	3	1	1	277775000000	7	23	2024-03-25T18:	25	0	0	30715	0	0

Data Analysis

When analyzing the data, we are looking at every folder and directory, using SpriteTree, which is edited when the malware is running, and learn everything we can about the directory.

We look for variables collected, using EsLogger, and see if there are any similarities between these different variables.

Lastly, we are looking for telltale signs that a file is malicious and that we can have the machine learning algorithm concentrate on and focus throughout the time it is running.



Future Goals / Research

Our aim moving forward is to meticulously clean the data we've collected so that we can effectively train our machine learning model to determine whether a file is malicious or benign.

We also expect to continue our research and expand the scope of it to incorporate other types of malware such as ransomware. To support this, we plan on refining our machine learning model and training it to accurately detect different types of malware.