ISOT Ransomware Dataset Overview

ISOT ransomware dataset is a combination of the behavior data for a collection of ransomware samples and benign applications. The ransomware samples were obtained from Virustotal[1] under academic license, in addition to obtaining several samples from anti-malware companies. The dataset consists of a total number of 669 ransomware samples representing most of the popular ransomware families and variants that emerged in the wild. In addition to the ransomware samples, the dataset include data from 103 benign applications representing the most popular software applications used by Window users. The total size of the dataset on disk is 428 GB.

The ransomware and the benign samples were analyzed using Cuckoo sandbox [1]. In our setup, we used Windows 7 (64 bit) to run the samples and collect the behavior data.

A malware analysis sandbox (cuckoo) was installed on a host machine running Ubuntu 14:04 TLS. The sandbox used, Cuckoo, is an open source and free software that automates the analysis of suspicious files. Cuckoo sandbox consists of two main components; a host machine and an analysis machine. The whole analysis process is managed by the management software running on the host machine, while the analysis machine is an isolated environment in which the samples are executed and analyzed. The virtual networking between the host running the Cuckoo software and the analysis machine uses a host-only adapter networking layout. Full Internet access was provided for the analysis machine to enable required communications with other agents involved in the attack such as C&C servers, kill switches, and so forth, for the analyzed ransomware samples to run successfully. The outbound network traffic to the LAN network was restricted to protect against the ransomware trying to spread to other machines.

Analysis machine info

IP: 192.168.50.11

OS Platform: Windows 7 (64 bit)

Dataset Structure

After decompressing (extracting) the downloaded dataset zip file, under the main directory (see Figure 1) there is a separate subdirectory for each family named after the family name that the analyzed samples belong to, such as “Spora”, “Locky”, etc. The behavior data for the benign applications is found under the subdirectory “Benign”.

[1] http://virustotal.com

ISOT Ransomware Dataset Overview

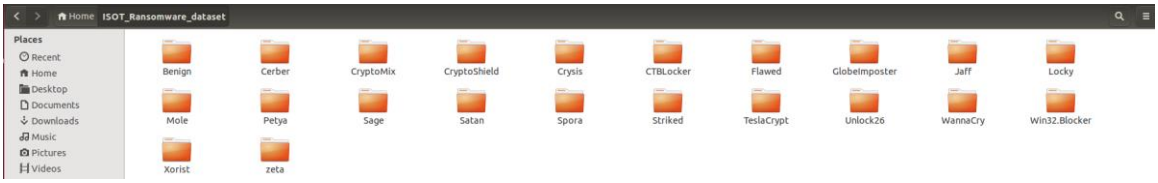


Figure 1: Dataset main directory structure

Inside the directory of each family (see Figure 2), there is a subdirectory for each sample that contains the collected data for the sample. Each subdirectory holds a numerical name which represents the ID assigned to the analysis task during analysis of the sample.

Figure 2: Family subdirectory structure

Various behavior data are collected for each analyzed ransomware and benign sample. This data is stored in different files and directories. Following is an example of a sample analysis directory structure:

```
.
|-- dump.pcap
|-- memory.dmp
|-- files.json
|-- logs
|   |-- 1232.bson
|   |-- 1540.bson
|   `-- 1118.bson
|-- reports
|   |-- report.html
|   |-- report.json
```

The different data and information that are shared for each analyzed sample are explained as follows:

dump.pcap
This is the network dump of the traffic, inside the analysis machine, generated during the analysis of the ransomware/benign sample.

memory.dmp
This file contains a full memory dump of the analysis machine. To reduce the size of the dataset, this file may not be available for all samples.

files.json A JSON-encoded entry is stored in this file for each dropped file. The JSON entries contain Meta information about all processes that touched the file, its original file path in the analysis machine, etc.

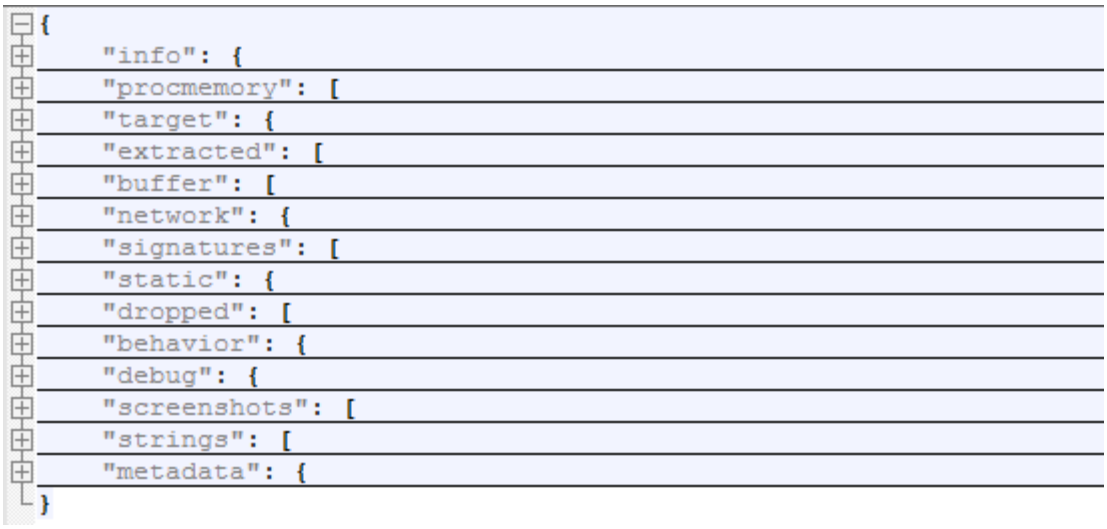
ISOT Ransomware Dataset Overview

logs/
All the raw logs generated by the process monitoring of Cuckoo sandbox are stored in this directory in .bson files.

reports/
All the analysis reports are stored in this directory. The .json version of the analysis report contains a JSON-encoded entry for the following, to name a few.

- Information about the analysis task, duration of the analysis, and several other information
- Information about the various memory regions
- Established network connections
- Information about all the processes created by the analyzed sample
- Information about the system calls initiated during the sample analysis, arguments passed, and return values
- Strings extracted from the binary file of the analyzed sample
- Information about the different operations on the file system
- Information about the different operations on Windows registry

Following is an example of a sample analysis report top-level structure:



References

[1] Cuckoo Sandbox. Cuckoo Sandbox- Automated malware analysis. Retrieved from https://cuckoosandbox.org