A Framework for the Analysis of File Infection Malware

09/2023 - 03/2024

Lorenzo Ippolito
Promo 2024 – Digital Security

Malware

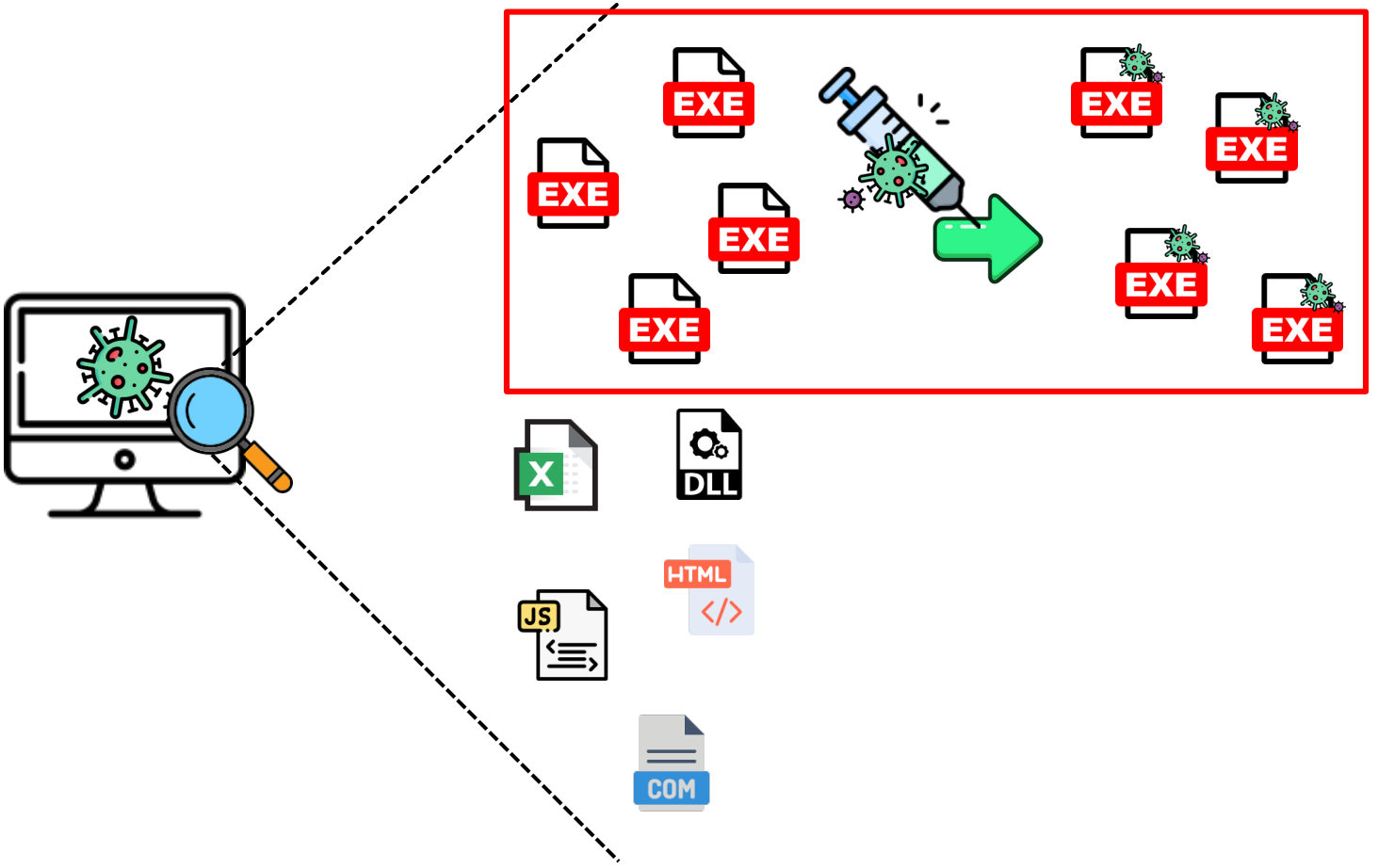
Categories:

- File Infectors
- Worm
- Grayware
- Backdoor
- Ransomware
- Clickwa

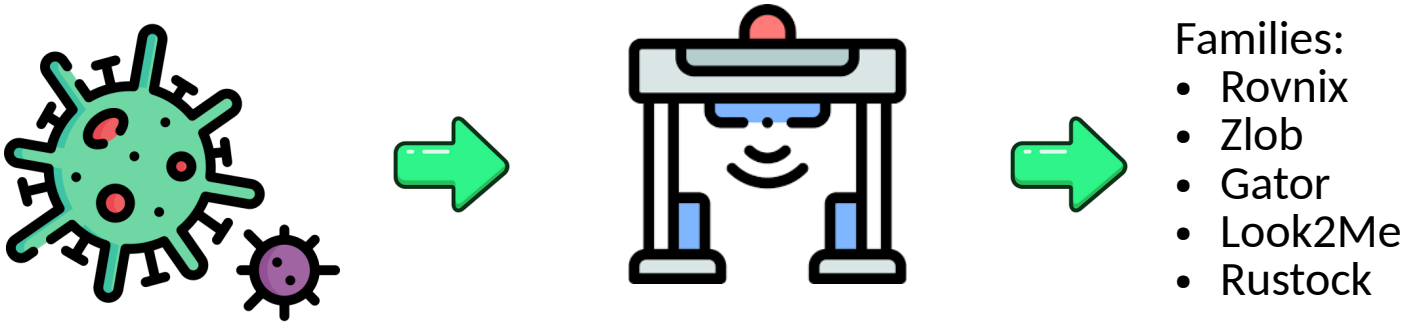
Families:

- Expiro
- Melissa
- Zeus
- Emotet
- Conficker
- Nimda
- Ramnit
- Slammer
- Wannacry
- Maze
- Ryuk

File Infection



Detection & Classification



Families:

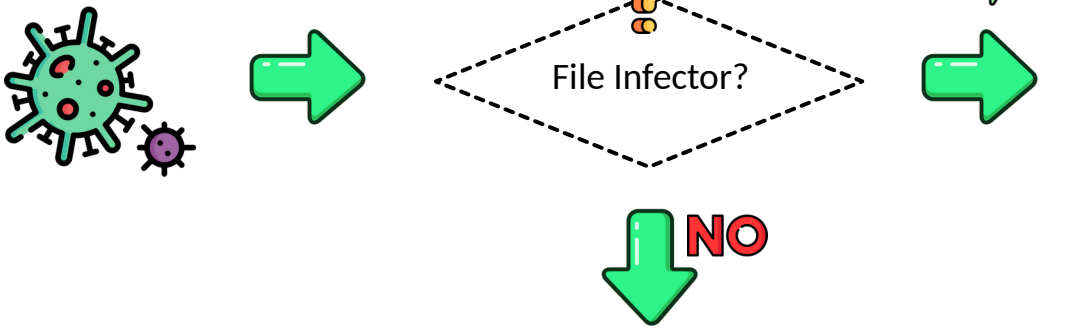
- Rovnix
- Zlob
- Gator
- Look2Me
- Rustock

File Infectors are Hard to Detect/Classify

Class	ry class. Recall Dyn.	Comb.	Fam S	core Com.
Adware		0.981	61	0.925
Backdoor		0.9	.730	0.838
Clicker	0.97		0.692	0.821
Dialer	0.994		88	0.888	0.984
Downloader	0.974		0.864	0.695	0.874
Grayware	0.932		832	0.675	0.852
Miner	0.989		7	0.807	0.962
Ransomware	0.9			0.580	0.853
Rogueware		0.99	401	0.663
Spyware	29	0.998	4	0.879
Tool	.929	1.000		0.830
Virus	0.939	0.971	0.8	.809
Wo	78	0.899	0.996	0.922	1
	0.967	**0.920**	**0.9907**	**0.848**	**0.7**

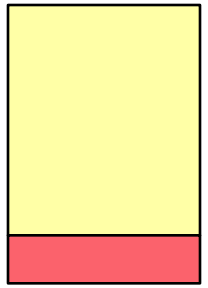
Dambra et al. «Decoding the Secrets of Machine Learning in Malware Classification: A Deep Dive into Datasets, Feature Extraction, and Model Performance». CCS '23.

Our Problem

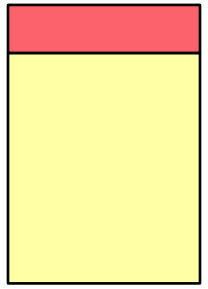


Not a File Infector

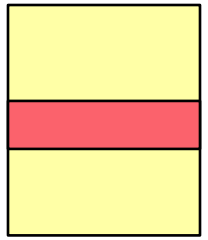
APPENDER



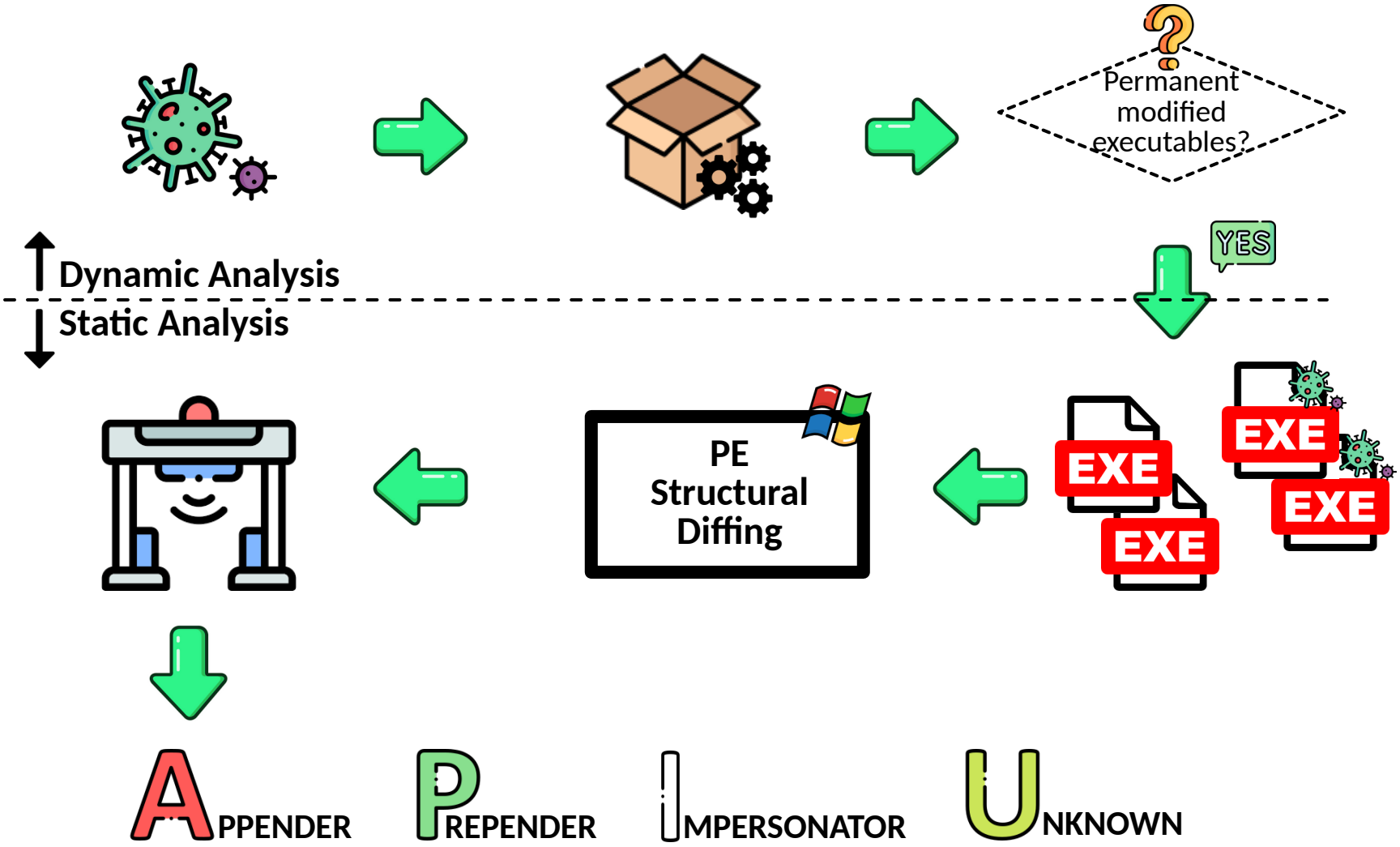
PREPENDER



CAVITY



Approach Overview



PE Structural Diffing

DOS Header
DOS Stub
Rich Header (OPT)

COFF File Header
Standard Fields
Windows Specific Fields
Data Directories

Section Table

Section 1

Section 2

Attribute Certificate Table (OPT)

Overlay (OPT)

DOS Header
DOS Stub
Rich Header (OPT)

COFF File Header
Standard Fields → **EP modified**
Windows Specific Fields
Data Directories

Section Table → **Section Table modified**

Section 1

Section 2

Section 3 → **Section added**

Attribute Certificate Table (OPT)

Overlay (OPT)

Related Work

- **File infectors**
  - Vast literature on manual analysis of file infector families (i.e., Memery, Neshta, Triusor [VirusBulletin])
  - Concept of Computer Virus [Cohen1987, Szor2005, Filiol2006]
  - Overview of malwares [Skoudis2004]
- **Executable Diffing**
  - Unlike existing tool such as *BinDiff*, we developed a component based PE Structural Diffing
- **Fuzzy Hashes**
  - We implied TLSH [Azab2014] and SSDeep [Kornblum] for PE Structural Diffing
- **Our work**
  - To our knowledge, there are no existing studies on automated file infector classification based on infection behavior

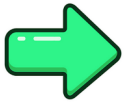
Dataset

Dambra et al. Dataset

- 67,000 samples
- 670 families
- 100 samples per families

  Collected between 2021-2022

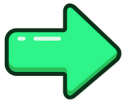
AVClass2



Subset Dataset

- 7,000 samples
- 70 families
- 100 samples per families

Analysis

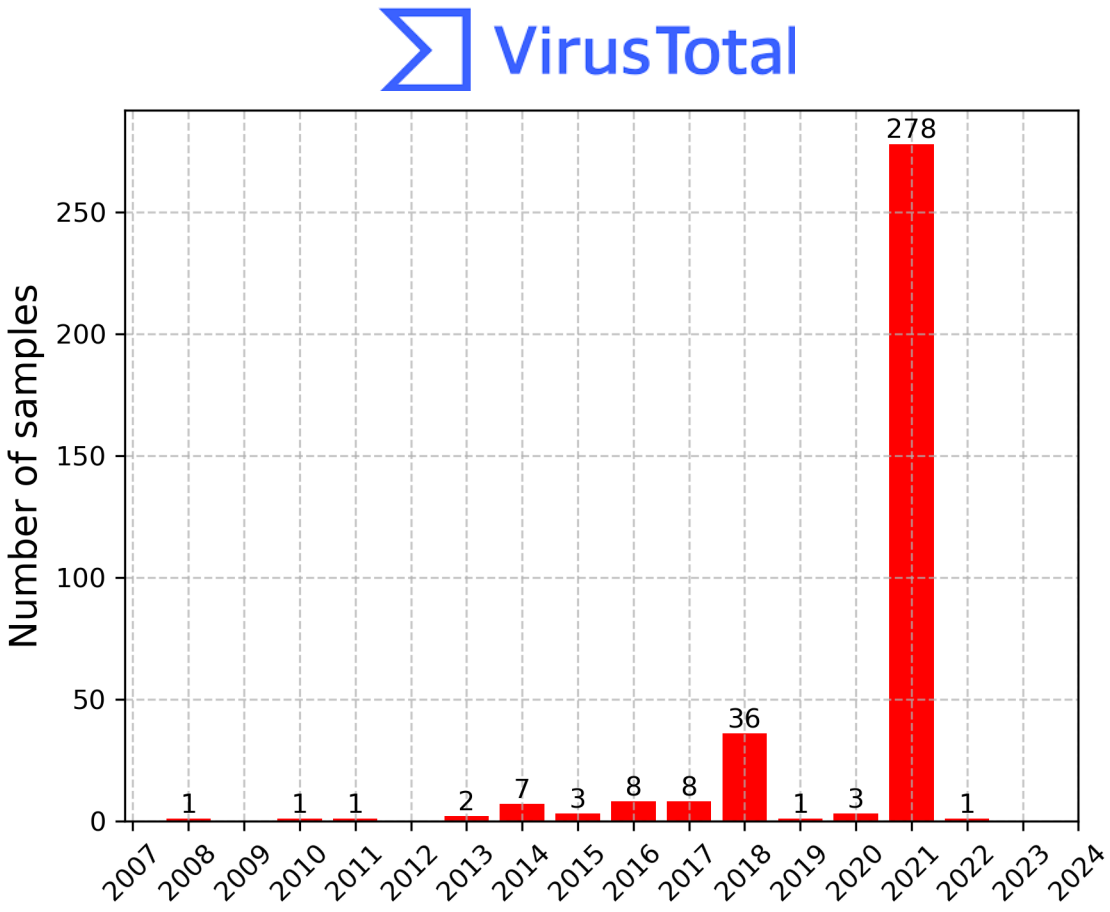


Analyzed Dataset

- 350 samples
- 70 families
- 5 samples per families

Dataset is balanced

First Seen Years



Contributions

- A novel framework for the analysis of file infection malware
- A PE Executable Differ
- Evaluation on 350 malware samples

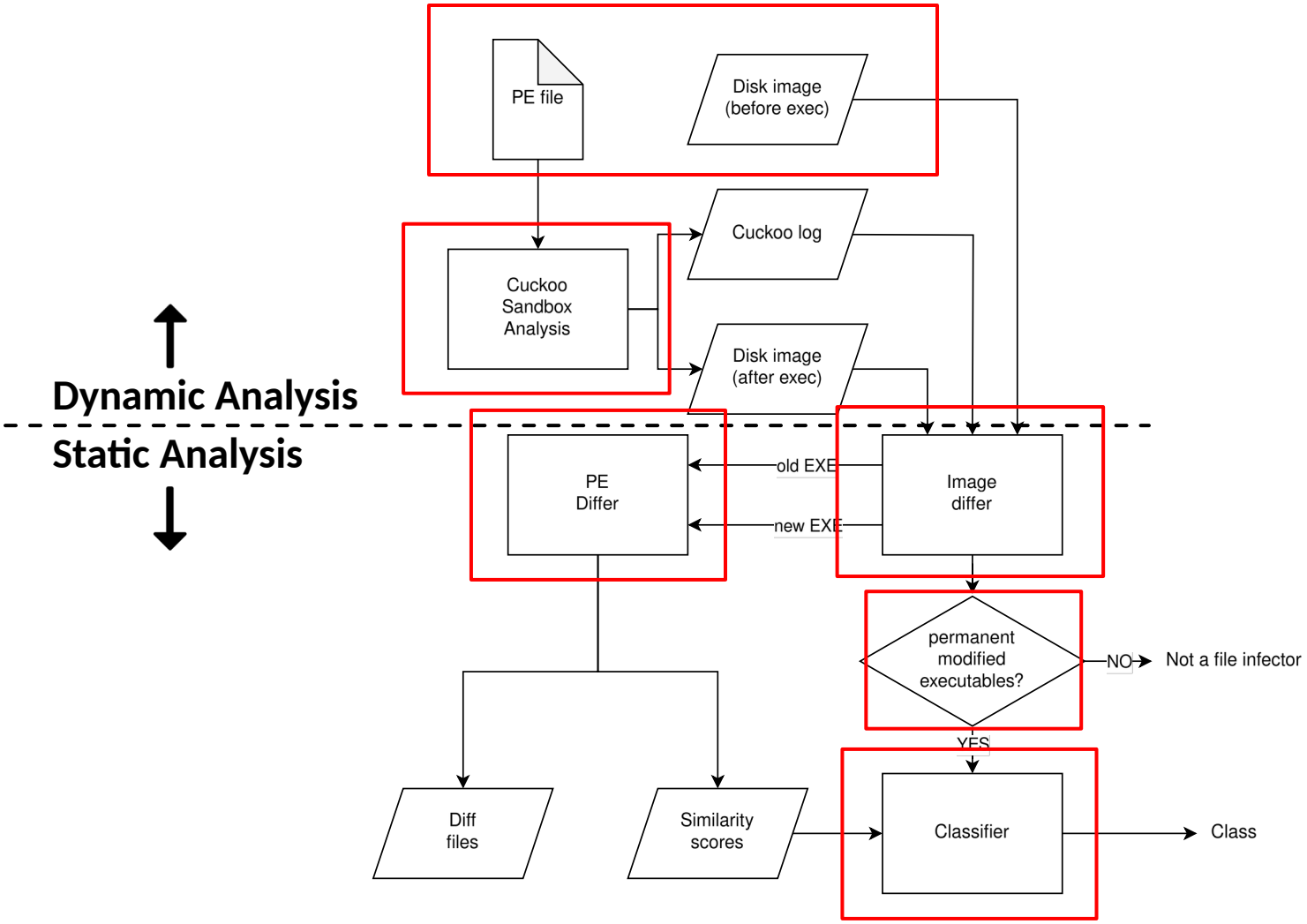
Outline

- Introduction

➔ **Approach**

- Results

Approach


Dynamic Analysis
Static Analysis
PE file
Disk image (before exec)
Cuckoo Sandbox Analysis
Cuckoo log
Disk image (after exec)
PE Differ
old EXE
new EXE
Image differ
permanent modified executables?
NO
Not a file infector
YES
Diff files
Similarity scores
Classifier
Class

Cuckoo Sandbox

- Input : Sample
- Output :

Cuckoo report

JS report.js


```
{
    "info": {
        "added": 1707408441.368206,
        "started": 1707408454.89419,
        "duration": 425,
        "ended": 1707408880.598961,
        "owner": null,
        "score": 3.8,
        "id": 1,
        "category": "file",
        "git": {
            "head": "13cbe0d9e457be3673304533043e992ead1ea9b2",
            "fetch_head": "13cbe0d9e457be3673304533043e992ead1ea9b2"
        },
        "monitor": "2deb9ccd75d5a7a3fe05b2625b03a8639d6ee36b",
        "package": "",
        "route": "internet",
        "custom": null,
        "machine": {
            "status": "stopped",
            "name": "Win7",
            "label": "Win7",
            "manager": "VirtualBox",
            "started_on": "2024-02-08 16:07:35",
            "shutdown_on": "2024-02-08 16:14:40"
        },
```

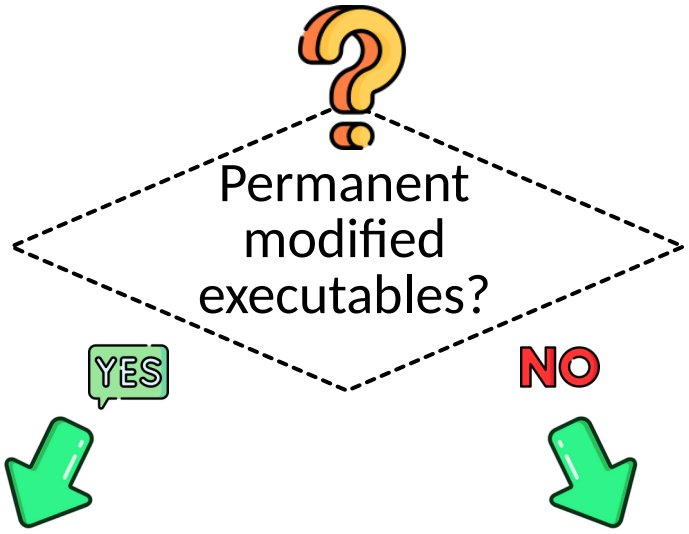

Disk Image (after exe)

Image Differ

- Input : Disk Image (orig.) & Disk Image (infect.) & Cuckoo report
- Output : Permanent modified executable dictionary

```
//Windows/assembly/NativeImages_v2.0.50727_64/MSBuild/1a154709cdfe214029ea88c51ab2b579
    321b2b1d180b4eb0bb5402fe6417f7f892a9dd68089419274f35dd555820cd35
    065b8ef72d98dd901b0199cf1a007c7696ee29279245f67393db4ea7f01f414f

//$Recycle.Bin/S-1-5-21-483214431-1722755210-2890981749-1001/$R7W2UHH/lib64/python3.10/site-packages/setuptools
    5c1af46c7300e87a73dacf6cf41ce397e3f05df6bd9c7e227b4ac59f85769160
    256ceba9c8e9e2303748398d2e08be9257756cbdcc7160924e4a787328334d58

//$Recycle.Bin/S-1-5-21-483214431-1722755210-2890981749-1001/$R7W2UHH/lib64/python3.10/site-packages/setuptools
    28b001bb9a72ae7a24242bfab248d767a1ac5dec981c672a3944f7a072375e9a
    639bb16efaf06be54434215a9eaadcea35aa3fb600a8359211bc3a5b05e3aa35

//Python27/Lib/site-packages/pip/_vendor/distlib
    34f60fa6decf22356a00112ed42cda6db0f21c7909a6ec3efea66aff8f07d23d
    b8e087f1b2166047bdb40c7335ddde85fa13e7118ad9831221edb9f309446c5a

//Program Files/LibreOffice/program
    a9e7d53b51e332a9a182e1cbb801ba243e98535aaf99991a53a4925865fdee1d
    12d70059c2d49067ca281a60542ea9cca660c0025825795a9d1dfe2822136e58

//Program Files/Oracle/VirtualBox Guest Additions
    631a44ff469772323d877168757d67bfde69125a492bd3875b9177c5226c5ea3
    3c9d086f6b4c4acee8dd328a465fccfc8a62020c25180ad8d93e83ca348c043e
```

Is it a File Infector?



Classify!

Not a File Infector

Fuzzy Hashes

- Fuzzy hashes allow to compute similarities between executables
- Similarity score range from 0 to 100
- Fuzzy hashes used are *TLSH* and *SSDeep*

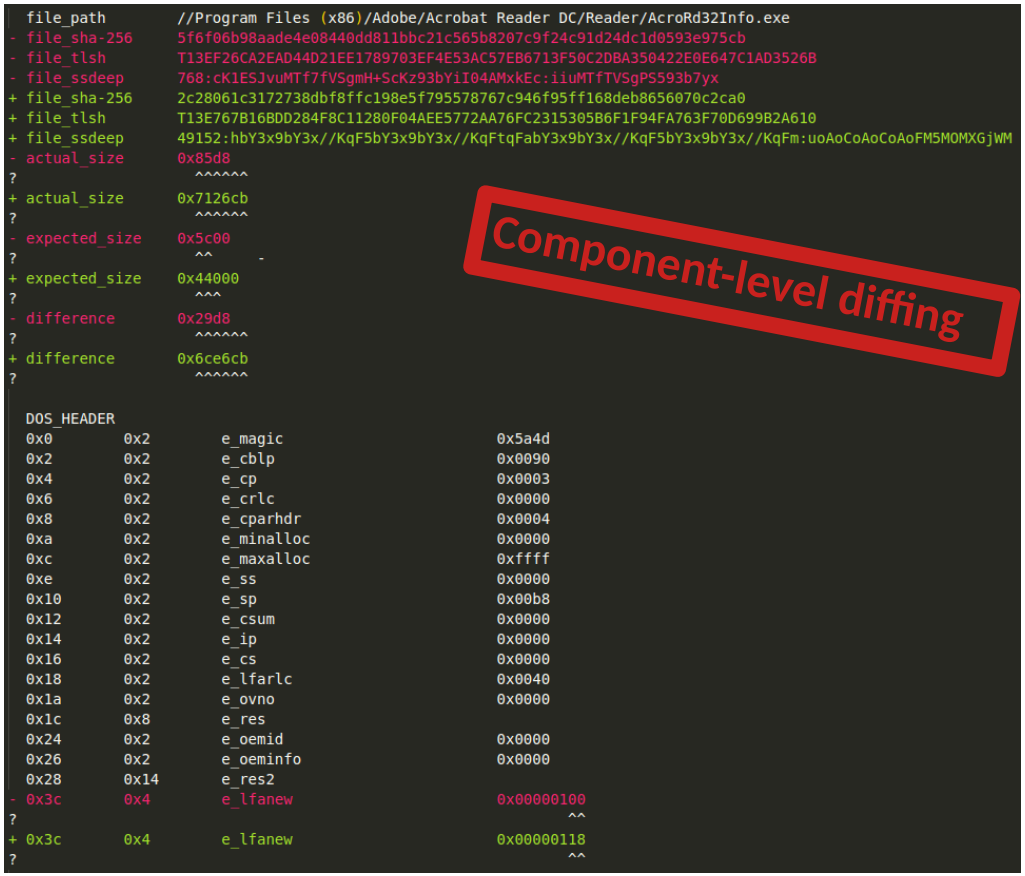
T134537C21B981C073C446107A592DC6B19F
7BBC312675C983BB961BBB9F313D1E72E24A

T1BA666B02B69DBCF8C4765030477793F25B
29FC211560EA5F73D4BB252E34683BA29B26

Similarity: **67**

PE Differ

- Input : Original EXE & Infected EXE
- Output : Diff file



Classifier Features

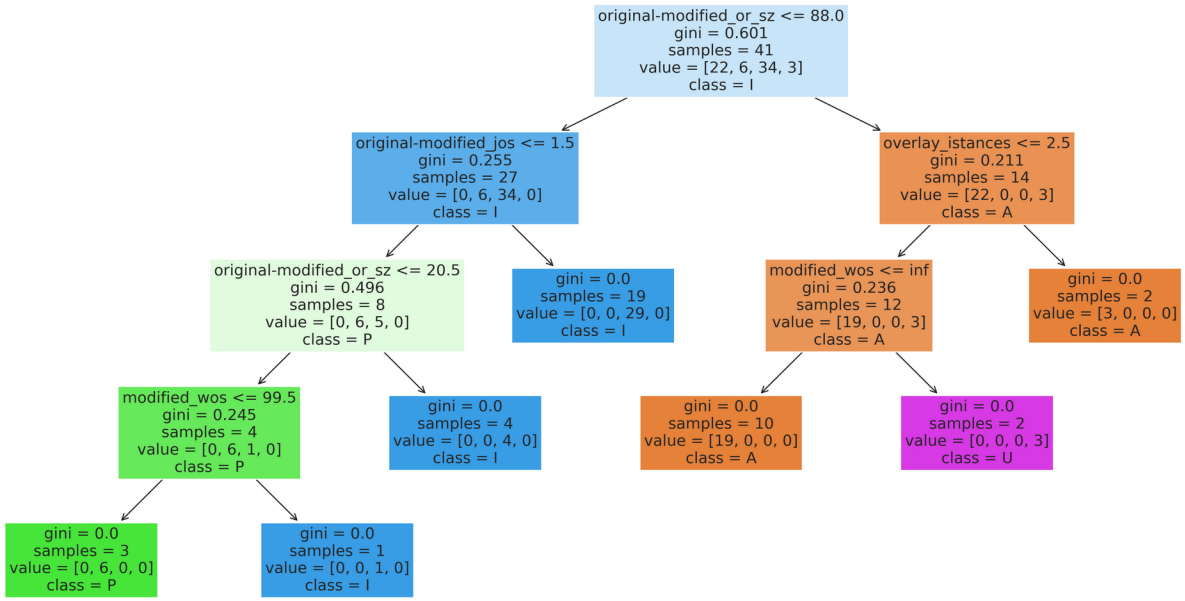
Ground Truth Labels

family	type	st added	st extend	st remov	orig es mod	ep mod	overlay ratio
expiro	A	0	1	0	✓	✗	✗
triusor	A	4	0	1	✗	✓	✗
wapomi	A	1	0	0	✗	✓	✗
wlksm	A	0	1	0	✓	✗	✗
lamer	P	3	0	all	-	-	637.2
induc	P	3	0	all	-	-	1.0
neshta	P	8	0	all	-	-	1.0
shodi	P	4	0	all	-	-	6.4
sinau	P	8	0	all	-	-	1.0
sivis	P	3	0	all	-	-	4.8
soulclose	P	3	0	all	-	-	1.0
xiaobaminer	P	7	0	all	-	-	53.3
memery	P	4	0	all	-	-	55.9
pidgeon	P	26	0	all	-	-	0.8
detroie	P	8	0	all	-	-	20.3
gogo	P	3	0	all	-	-	33.1
lmir	P	8	0	all	-	-	0.1
stihat	P	18	0	all	-	-	13.4
xolxo	P	70	0	all	-	-	82.6
xorer	P	3	0	all	-	-	1.8
virlock	I	2	0	all	-	-	✗
grenam	I	10	0	all	-	-	✗

22 file infector families
4 Appenders, 16 Prependers, 2 Impersonators

Classifier

- Random Forest Classifier on 4 different labels (A, P, I, U)
- Training Set and Testing Set divided as 70 % and 30 %



Outline

- Introduction
- Approach

➔ **Results**

Results: Execution

- 350 samples analyzed, 5 for each family
- 2 families with 0 Windows API calls
- About 97% of the samples detonated (>50 Win API)
- 94 samples of 22 families permanently modified executables, thus labelled as file infectors

Results: Classifier

	Accuracy	Macro Avg	Weighted Avg	A	P	I	U
Precision	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Recall	1.0	1.0	1.0	1.0	1.0	1.0	1.0
F1-score	1.0	1.0	1.0	1.0	1.0	1.0	1.0

POTENTIAL OVERFITTING

Limitations

- A fraction of samples may not have detonated
- The 70 families may not capture all file infector types, e.g., we did not encounter *Cavity infectors*
- Only 350 samples analyzed
- Overfitting in the random forest classifier

Conclusions

- A novel framework for the analysis of file infection malware
- A PE Executable Differ
- Evaluation on 350 malware samples

Future work

- Analysis of a more extensive sample pool
- Additional features beyond similarity scores
- Improving the capabilities of PE Differ

Questions?

Thanks!