Similarity Analysis of Ransomware based on Portable Executable (PE) File Metadata

Ahsan Ayub and Ambareen Siraj

Tennessee Tech University, Cookeville, USA

Emails: mayub42@tntech.edu, asiraj@tntech.edu

IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2021)

December 4-7, 2021

What is Ransomware?

- A type of malware that takes over the system by affecting the victim machine via email, remote desktop protocol, software vulnerability, etc.
- Mainly two kinds of ransomware –
  - Locker Ransomware
  - Crypto Ransomware

Motivation

- Ransomware attacks on the computer systems of government bodies, healthcare, banking sector, airports, U.S. school districts, etc.
- The DarkSide ransomware attack in May 2021 on the colonial pipeline network, a company that supplies about half of the U.S. East Coast's gasoline
  - State of emergency declared in 18 states
  - Paid US$ 4.4 million worth of bitcoin
  - Resumed operation after 5 days of national panic

Research Questions (RQ)

RQ1. Can we identify suspicious indicators from ransomware samples' structural information?

RQ2. Is there any PE file metadata-based similarities among the studied ransomware samples as well as their families?

Portable Executable (PE) File

- A common object file on the Windows Operating System with extensions include *.exe* (executable file), *.dll* (dynamic link library), *.sys* (system file), etc.
- The PE file holds several pieces of information in different categories: File Header, Section Tables, Imports Address Table (IAT), etc.

PE File: File Header

- It contains –
  - Type of targeting machine,
  - Size of the section table,
  - Time and date that the file was created,
  - Flags indicating different attributes of file, etc.
- Additionally, optional headers include –
  - Magic number of the file
  - Size of code
  - Initialized data
  - Image
  - Subsystem required to run the image
  - DLL characteristics
  - Address of the entry point

PE File: Section Header

- This category includes –
  - Each section’s virtual address
  - Virtual size
  - Size of raw data
- Common section names are *.text* (executable code), *.data* (read/write data), *.idata* (import address table), *.edata* (export information), etc.

PE File: Import Address Table (IAT)

- Contains information about both the libraries and the imports used by the PE file
- For example, for one of the studied samples from Petya ransomware family, we find out that it uses *wininet.dll* library, Windows Internet (WinINet) application programming interface (API), that interacts with 12 imports, such as, *HttpOpenRequest*, *HttpSendRequest*, etc.

Experimental Setup

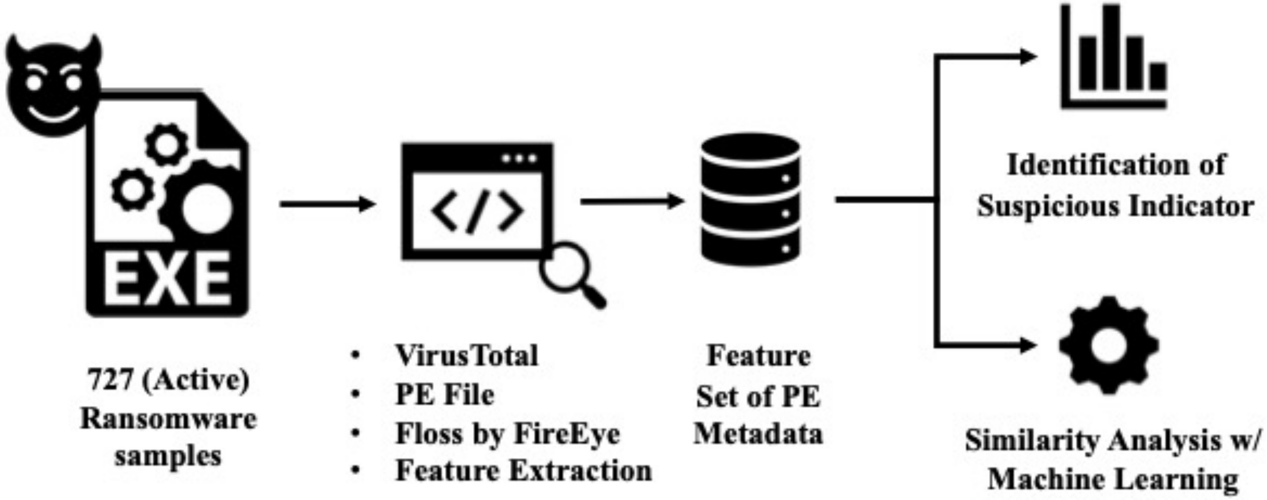

EXE
727 (Active) Ransomware samples
• VirusTotal
• PE File
• Floss by FireEye
• Feature Extraction
Feature Set of PE Metadata
Identification of Suspicious Indicator
Similarity Analysis w/ Machine Learning

Fig. 1. Framework of our approach to identify similarities among the studied ransomware samples based on PE file metadata.

Generation of Feature Set of PE Metadata

- Gather the numeric details of how many Anti-Virus (AV) engines identify the ransomware sample file as malicious or safe through VirusTotal API engine
- Utilize PEFile library, available as a Python module, to parse through the PE files' information
- Leverage FireEye Labs Obfuscated String Solver (FLOSS) to extract obfuscated strings from the samples

Observation of the Generated Feature Set

- All the samples are Non-Executable types of files that target 32-bit Microsoft Windows machines.
- 87% of the samples allocate more memory space in their PE sections than they have data written to disk.
- Total unique number of libraries and imports are 106 and 3,345 respectively with no presence of export table.
- 4% and 11% of samples show the usage of the packer and crypto libraries respectively.

Addressing RQ1: Imports

- Cursor and/or Mouse (62% samples)
- Network calls (30% samples)
  - *http*, *ftp*, *url*, and *icmp* are present in 14%, 12%, 15%, and 17% samples respectively.
- Shell execution (13% samples)
- Debugger presence checker (30% samples)
- Process-based imports (76% samples)
- File-based imports (87% samples)

Addressing RQ1: Libraries

- 19% samples use *wtsapi32.dll*
  - Remote desktop service environment
- 5% samples use *wininet.dll (*Internet Extensions for Win32)
  - Helps the sample interact with the http and ftp protocols to access online resource
- 19% samples use *psapi.dll* - Process Status Helper API
  - Enable the samples to gain information about the running processes and device drivers

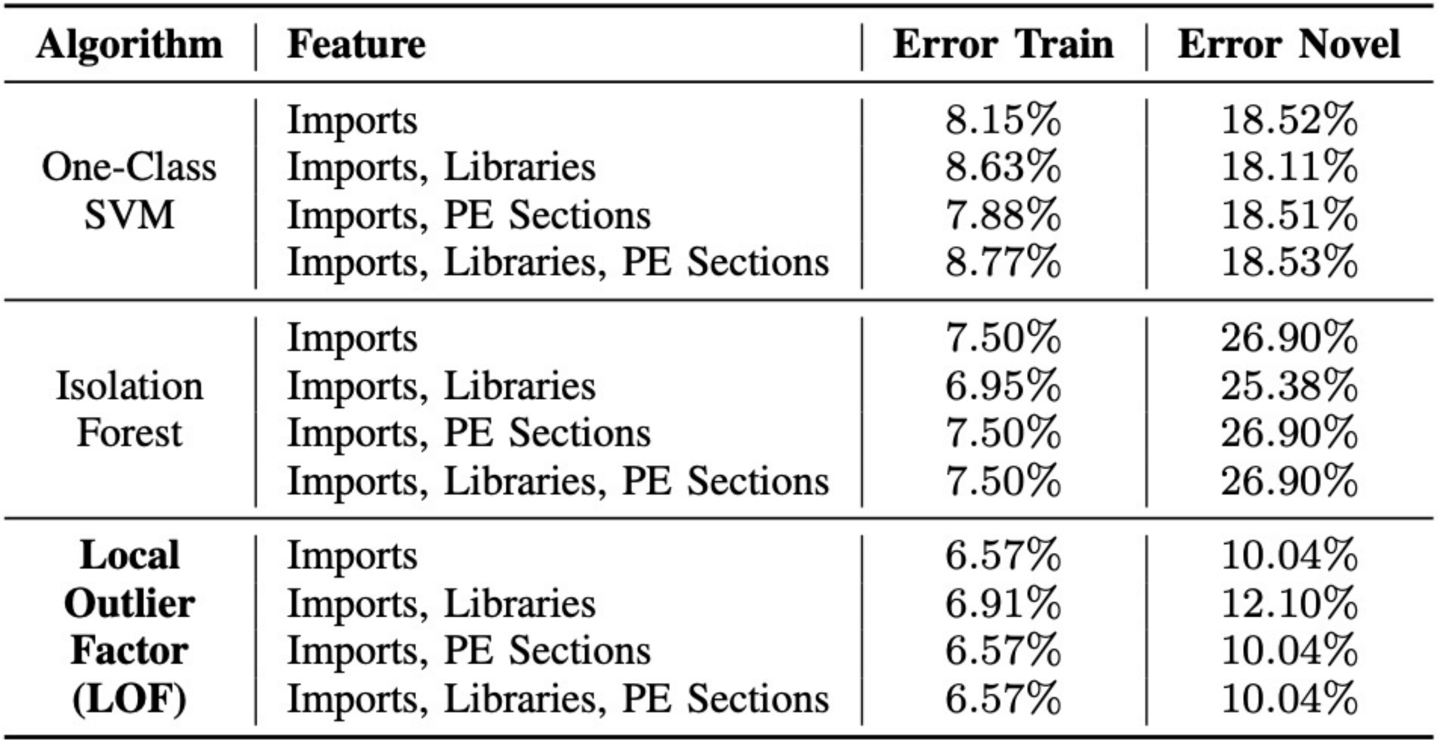
Addressing RQ1: Strings

- Encryption-based keywords: “encrypt”, “decrypt”, “RSA”, and “AES” keywords are present in 16.3%, 25.27%, 48.1%, and 22.1% samples respectively.
- Ransom-based notice: “payment” or “pay”, “bitcoin” or “btc”, and “usd” keywords are present in 14.09%, 7.74%, and 10.5% samples respectively.
- File Path: “C://” and “/windows” keywords are present in 6.35% and 7.18% samples respectively.

Addressing RQ2: One-Class Classification (1/3)

- Feature spaces include PE Metadata, Imports, Libraries, and PE Sections
- Applied one-class classification algorithms to identify similarities among samples
  - One-Class SVM
  - Isolation Forest
  - Local Outlier Factor (LOF)
- Performed 5-fold cross-validation to report the evaluation of each model through Error Train and Error Novel

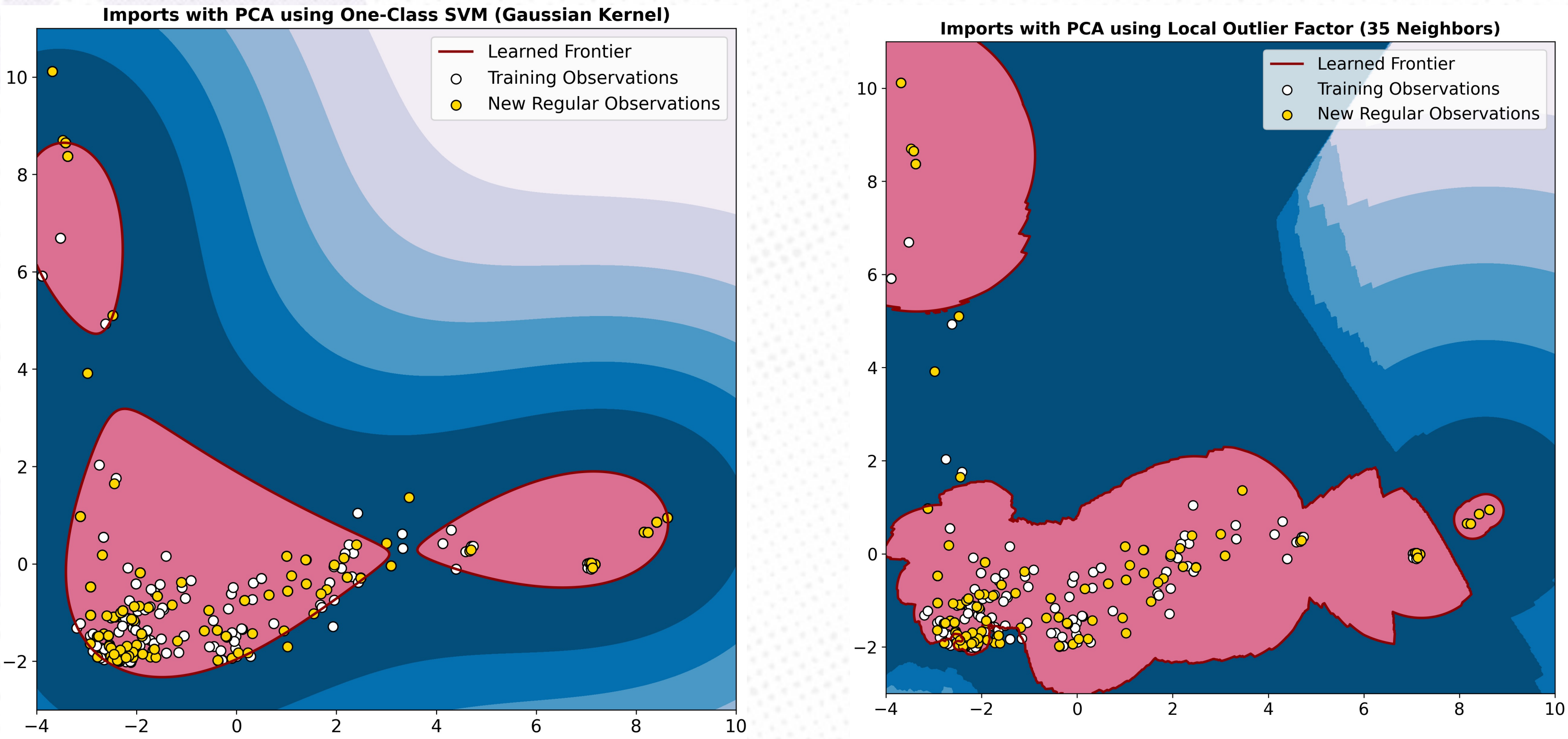
Addressing RQ2: One-Class Classification (2/3)



Algorithm	Feature	Error Train	Error Novel
One-Class SVM	Imports	8.15%	18.52%
	Imports, Libraries	8.63%	18.11%
	Imports, PE Sections	7.88%	18.51%
	Imports, Libraries, PE Sections	8.77%	18.53%
Isolation Forest	Imports	7.50%	26.90%
	Imports, Libraries	6.95%	25.38%
	Imports, PE Sections	7.50%	26.90%
	Imports, Libraries, PE Sections	7.50%	26.90%
Local Outlier Factor (LOF)	Imports	6.57%	10.04%
	Imports, Libraries	6.91%	12.10%
	Imports, PE Sections	6.57%	10.04%
	Imports, Libraries, PE Sections	6.57%	10.04%

Fig. 2. Performance of One-Class Classification algorithms in different experimental settings.

Addressing RQ2: One-Class Classification (3/3)



Imports with PCA using One-Class SVM (Gaussian Kernel)

Imports with PCA using Local Outlier Factor (35 Neighbors)

Fig. 3. Visualization of the learned cluster region of One-Class SVM (left) and Local Outlier Factor (right) classifiers for the Imports with PCA feature space

Summary

- We identify suspicious indicators on the generated PE metadata of ransomware based on the exploratory data analysis tasks and domain knowledge.
- We leverage the powerful one-class classification algorithms to capture the similarities among all the studied ransomware samples.
- We encourage the organizations to use the 3-2-1 rule, that is to keep 3 back-ups of their data: 2 on different storage types while 1 on offsite.

Acknowledgement

The work reported in this paper has been entirely supported by Cybersecurity Education, Research & Outreach Center (CEROC) at Tennessee Tech University…

THANK YOU!

Implementation

https://github.com/TnTech-CEROC/static_ransomware_analysis