

BÁO CÁO LAB

Môn học: Phương pháp học máy trong an toàn thông tin

LAB 2 : Machine Learning based Malware Detection

Nhóm: 05

THÔNG TIN CHUNG:

Lớp: NT522.N21.ATCL

STT	Họ và tên	MSSV	Phân công	Hoàn thành
1	Đỗ Quang Thắng	20521893	2,6,7,8	100%
2	Nguyễn Đoàn Thiên Cung	20521146	1,4,5	100%

1. NỘI DUNG THỰC HIỆN:

BÁO CÁO CHI TIẾT

Câu 1,4,5 :

https://drive.google.com/drive/folders/13kQGk_janXoDqHyBT07cnDAZIeQJXKHw?usp=sharing

Câu 2,6,7,8:

https://colab.research.google.com/drive/1egTPqvqVpKxrKmsBUZXJh1DceB9iK_qk?usp=sharing

1. Sinh viên so sánh kết quả băm với VirusTotal và website Python.
Hàm băm của code so với VirusTotal và website Python là giống nhau

Câu 1: Sinh viên so sánh kết quả băm với VirusTotal và website Python.

```
import sys
import hashlib
filename = "/content/drive/MyDrive/Colab Notebooks/LAB2/python-3.10.0-amd64.exe"
BUF_SIZE = 65536
md5 = hashlib.md5()
sha256 = hashlib.sha256()
with open(filename, "rb") as f:
    while True:
        data = f.read(BUF_SIZE)
        if not data:
            break
        md5.update(data)
        sha256.update(data)
print("MD5: {}".format(md5.hexdigest()))
print("SHA256: {}".format(sha256.hexdigest()))
```

MD5: c3917c08a7fe85db7203da6dcaa99a70
SHA256: cb580eb7dc55f9198e650f016645023e8b2224cf7d033857d12880b46c5c94ef

cb580eb7dc55f9198e650f016645023e8b2224cf7d033857d12880b46c5c94ef

0 / 53

✓ No security vendors and no sandboxes flagged this file as malicious

cb580eb7dc55f9198e650f016645023e8b2224cf7d033857d12880b46c5c94ef

python-3.10.0-amd64.exe

peexe overlay runtime-modules signed detect-debug-environment direct-cpu-clock-access

Community Score

DETECTION DETAILS RELATIONS BEHAVIOR COMMUNITY 12



Files

Version	Operating System	Description	MD5 Sum	File Size	GPG
Gzipped source tarball	Source release		729e36388ae9a832b01cf9138921b383	25007016	SIG
XZ compressed source tarball	Source release		3e7035d272680f80e3ce4e8eb492d580	18726176	SIG
macOS 64-bit universal2 installer	macOS	for macOS 10.9 and later (updated for macOS 12 Monterey)	8575cc983035ea2f0414e25ce0289ab8	39735213	SIG
Windows embeddable package (32-bit)	Windows		dc9d1abc644dd78f5e48edae38c7bc6b	7521592	SIG
Windows embeddable package (64-bit)	Windows		340408540eeff359d5eaf93139ab90fd	8474319	SIG
Windows help file	Windows		9d7b80c1c23cfb2cecd63ac4fac9766e	9559706	SIG
Windows installer (32-bit)	Windows		133aa48145032e341ad2a000cd3bff50	27194856	SIG
Windows installer (64-bit)	Windows	Recommended	c3917c08a7fe85db7203da6dcaa99a70	28315928	SIG

2. Thực hiện đoạn code và in ra kết quả.

b) YARA

```

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows
PS C:\Users\DELL\Downloads\New folder (6)> ls

Directory: C:\Users\DELL\Downloads\New folder (6)

Mode                LastWriteTime         Length Name
----                -
-a----             4/6/2023  10:13 AM             179 rules.yara
-a----             4/6/2023  10:10 AM          2221001 yara-4.3.0-2120-win64.zip
-a----             3/24/2023   9:52 PM          2406912 yara64.exe
-a----             3/24/2023   9:52 PM          2353664 yarac64.exe

PS C:\Users\DELL\Downloads\New folder (6)> .\rules.yara .\rules.yara '.\Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf'
PS C:\Users\DELL\Downloads\New folder (6)> .\yara64.exe .\rules.yara '.\Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf'
is_a_pdf .\Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf
dummy_rule2 .\Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf
PS C:\Users\DELL\Downloads\New folder (6)> .\yara64.exe .\rules.yara 'Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf'
is_a_pdf Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf
dummy_rule2 Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf
PS C:\Users\DELL\Downloads\New folder (6)>

```

c) Kiểm tra PE header


```
public-apis  
youtube-dl  
core  
awesome-machine-learning  
ansible
```

5. Sinh viên cho biết quả của đoạn code trên

- Kết quả ta thu được từ đoạn code đã cho

```
0.9732142857142857  
[[122  0  0]  
 [  0 82  0]  
 [  6  0 14]]
```

6. Thực thi và kiểm tra kết quả.

g) Đo lường sự giống nhau giữa hai chuỗi

```
3:f4oo8MRwRJFGW1gC6uWv6MQ2MFS1+JuBF8BSnJi:f4kPvtHMCmubyFtQ  
3:f4oo8MRwRJFGW1gC6uWv6MQ2MFS1+JuBF8BS+EFECJi:f4kPvtHMCmubyFIsJQ  
3:f4oo8MRwRJFGW1gC6uWv6MQ2MFS1+JuBF8BS6:f4kPvtHMCmubyF0  
3:60QKZ+4CDTfDaRfKYLVL:YwKDC2mVL  
100  
39  
37  
0
```

h) Đo lường mức độ giống nhau giữa hai tập tin

```

quangthang@quangthang-VirtualBox:~$ hexdump -C python-3.10.0-amd64.exe | tail -5
01b010e0  10 9c 34 66 02 d3 51 8c  b1 64 19 f3 55 12 0e 74  |..4f..Q..d..U..t|
01b010f0  38 71 4c 2e 1c db 44 d4  f3 81 31 a5 9c 2e c6 06  |8qL...D...1....|
01b01100  4f 33 c6 8a 9a 5e 16 52  8c 4b 55 10 2b cd 45 61  |03...^..R.KU.+..Ea|
01b01110  a5 00 00 00 00 00 00 00  |.....|
01b01118
quangthang@quangthang-VirtualBox:~$ hexdump -C python-3.10.0-amd64-fake.exe | tail -5
01b010e0  10 9c 34 66 02 d3 51 8c  b1 64 19 f3 55 12 0e 74  |..4f..Q..d..U..t|
01b010f0  38 71 4c 2e 1c db 44 d4  f3 81 31 a5 9c 2e c6 06  |8qL...D...1....|
01b01100  4f 33 c6 8a 9a 5e 16 52  8c 4b 55 10 2b cd 45 61  |03...^..R.KU.+..Ea|
01b01110  a5 00 00 00 00 00 00 00  |.....|
01b01118
quangthang@quangthang-VirtualBox:~$ python3 compare.py
100
quangthang@quangthang-VirtualBox:~$

```

7. Thực thi và kiểm tra kết quả.

i) Trích xuất N-grams

```

extracted_Ngrams = binary_file_to_ngrams(binary_file)

print(extracted_Ngrams.most_common(10))

[(((0, 0, 0, 0), 24290), ((139, 240, 133, 246), 1920), ((32, 116, 111, 32), 1791), ((255, 255, 255, 255), 1671), ((108, 101, 100, 32), 1522), ((100, 32, 116, 111), 1519), ((97, 105, 97, 105), 1519), ((100, 100, 100, 100), 1519), ((100, 100, 100, 100), 1519), ((100, 100, 100, 100), 1519))]
7))

```

j) Chọn N-grams tốt nhất



```
[ ]
print("Frequency :")
print(X_top_K2_freq)
print("Mutual information :")
print(X_top_K2_mi)
print("Chi-squared :")
print(X_top_K2_ch2)

Frequency :
[[94271, 54969, 19, 1, 2992, 117, 13605, 3610, 264, 13443, 220, 6, 211, 429, 1249, 146, 115, 2940, 259, 62, 234, 344, 96, 520, 2, 186, 159, 836, 24, 521, 74, 190, 107]

Mutual information :
[[ 0 48 4 18 0 28 62 13 21 7]
 [ 0 12 41 25 2 17 0 23 14 4]
 [ 0 2 0 4 1 0 20 1 0 0]
 [ 0 8 2 20 3 10 5 12 3 13]
 [ 2 282 555 104 27 409 139 278 560 303]
 [ 26 191 12 122 18 133 56 448 69 121]
 [ 0 1 1 1 0 4 1 4 1 2]
 [ 6 37 5 107 16 32 28 3 7 56]
 [ 5 24 15 43 5 37 36 27 18 9]
 [ 29 25 21 49 24 17 40 24 29 42]
 [ 1 16 5 47 21 12 81 4 2 66]
 [ 1 22 9 57 2 10 86 8 12 39]
 [ 0 2 0 4 2 3 3 2 2 0]
 [ 2 15 0 4 1 5 12 2 2 9]
 [ 2 55 63 96 5 33 30 6 48 101]
 [ 2 221 290 223 40 246 361 149 204 217]
 [1001 72 47 86 894 49 49 36 44 48]
 [ 43 58 44 68 55 60 84 57 71 59]
 [ 1 26 3 27 1 6 15 8 2 11]
 [ 2 4 4 16 3 5 8 11 5 7]]

Chi-squared :
[[ 0 48 4 18 0 28 62 13 21 7]
 [ 0 12 41 25 2 17 0 23 14 4]
 [ 0 2 0 4 1 0 20 1 0 0]
 [ 0 8 2 20 3 10 5 12 3 13]
 [ 2 282 555 104 27 409 139 278 560 303]
 [ 26 191 12 122 18 133 56 448 69 121]
 [ 0 1 1 1 0 4 1 4 1 2]
 [ 6 37 5 107 16 32 28 3 7 56]
 [ 5 24 15 43 5 37 36 27 18 9]
 [ 29 25 21 49 24 17 40 24 29 42]
 [ 1 16 5 47 21 12 81 4 2 66]
 [ 1 22 9 57 2 10 86 8 12 39]
 [ 0 2 0 4 2 3 3 2 2 0]
 [ 2 15 0 4 1 5 12 2 2 9]
 [ 2 55 63 96 5 33 30 6 48 101]
 [ 2 221 290 223 40 246 361 149 204 217]
 [1001 72 47 86 894 49 49 36 44 48]
 [ 43 58 44 68 55 60 84 57 71 59]
 [ 1 26 3 27 1 6 15 8 2 11]
 [ 2 4 4 16 3 5 8 11 5 7]]
```

8. Thực thi và kiểm tra kết quả.

```
print("Classifier score: \n", classifier.score(X_test, y_test))

/content/drive/MyDrive/Năm 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/aspnetca.exe:
'DOS Header magic not found.'
/content/drive/MyDrive/Năm 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/AppVStreamingUX.exe:
'PE' object has no attribute 'DIRECTORY_ENTRY_IMPORT'
/content/drive/MyDrive/Năm 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Malicious PE Samples 2/Build.exe:
'utf-8' codec can't decode byte 0xd2 in position 6: invalid continuation byte
/content/drive/MyDrive/Năm 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/ADSchemaAnalyzer.exe:
'DOS Header magic not found.'
/content/drive/MyDrive/Năm 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/adamuninstall.exe:
'DOS Header magic not found.'
/content/drive/MyDrive/Năm 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/appcmd.exe:
'DOS Header magic not found.'
Classifier score:
0.8571428571428571
```

