BÁO CÁO THỰC HÀNH LAB 2

MÔN: Phương Pháp Học Máy Trong An Toàn Thông Tin

LAB 2: Machine Learning based Malware Detection

**Graphical user interface, application

Description automatically generated**

1. **THÔNG TIN CHUNG:**

Lớp: NT522.N11.ATCL

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Họ và tên** | **MSSV** | **Email** |
| 1 | Trần Hoàng Khang | 19521671 | 19521671@gm.uit.edu.vn |
| 2 | Lê Hồng Bằng | 19520396 | 19520396@gm.uit.edu.vn |
| 3 | Nguyễn Tú Ngọc | 20521665 | 20521665@gm.uit.edu.vn |

1. **NỘI DUNG THỰC HIỆN:**

|  |  |  |
| --- | --- | --- |
| **STT** | **Công việc** | **Kết quả tự đánh giá** |
| 1 | Câu hỏi 1 | 100% |
| 2 | Câu hỏi 2 | 100% |
| 3 | Câu hỏi 3 | 100% |
| 4 | Câu hỏi 4 | 100% |
| 5 | Câu hỏi 5 | 100% |
| 6 | Câu hỏi 6 | 100% |
| 7 | Câu hỏi 7 | 100% |
| 8 | Câu hỏi 8 | 100% |

BÁO CÁO CHI TIẾT

|  |
| --- |
| ***1.*** *Sinh viên so sánh kết quả băm với VirusTotal và website Python* |

**<Xem kết quả chi tiết tại file Notebook (.ipynb)>**

Kết quả hash chạy được trong file Notebook:

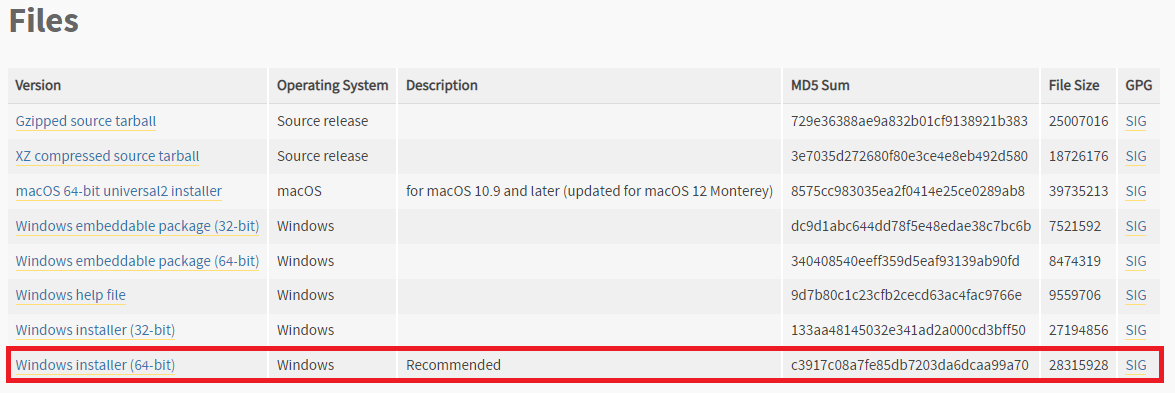
Text

Description automatically generated

MD5 hash value: **c3917c08a7fe85db7203da6dcaa99a70**

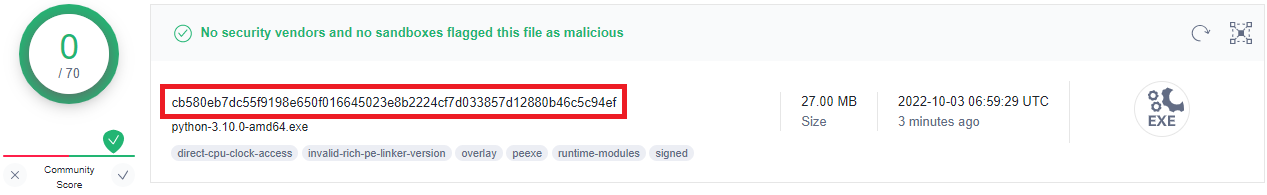
SHA256 hash value: **cb580eb7dc55f9198e650f016645023e8b2224cf7d033857d12880b46c5c94ef**

- Đối chiếu hash của MD5 với hash kiểm tra trên [trang chủ Python](https://www.python.org/downloads/release/python-3100/):



🡺 Mã hash trùng nhau, file tải về toàn vẹn và khớp data.

- Đối chiếu hash của SHA256 với hash kiểm tra trên [VirusTotal](https://www.virustotal.com/gui/home/upload):



🡺 Mã hash trùng nhau, file tải về toàn vẹn và khớp data.

|  |
| --- |
| ***2.*** *Sinh viên cho biết quả của đoạn code trên* |

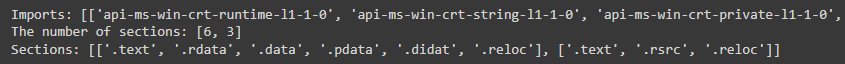
**<Xem kết quả chi tiết tại file Notebook (.ipynb)>**

Kết quả trả về như đúng mô tả, bao gồm:

- Tên các thư viện imports

- Số lượng Sections

- Tên các sections



|  |
| --- |
| ***3.*** *Sinh viên tự tìm hiểu, cài đặt (https://cuckoo.sh/docs/introduction/index.html), thực hiện và trình bày phân tích động một tập tin PE.* |

**Cài đặt Cuckoo Sandbox trên localhost (Ubuntu 20.04 LTS):**

**Lưu ý:** Khi cài đặt cuckoo, đảm bảo có một máy Host (tức là một máy chính: có thể dùng VirtualBox, VMWare, … ) và một máy Guest (Đây là môi trường Sandbox, để con Cuckoo bỏ file vào chạy thử, và môi trường này phải tắt hết tính năng bảo vệ cơ bản đi, thường là dùng WindowXP hoặc Window7 nằm trong một máy ảo của máy ảo trên). 😀

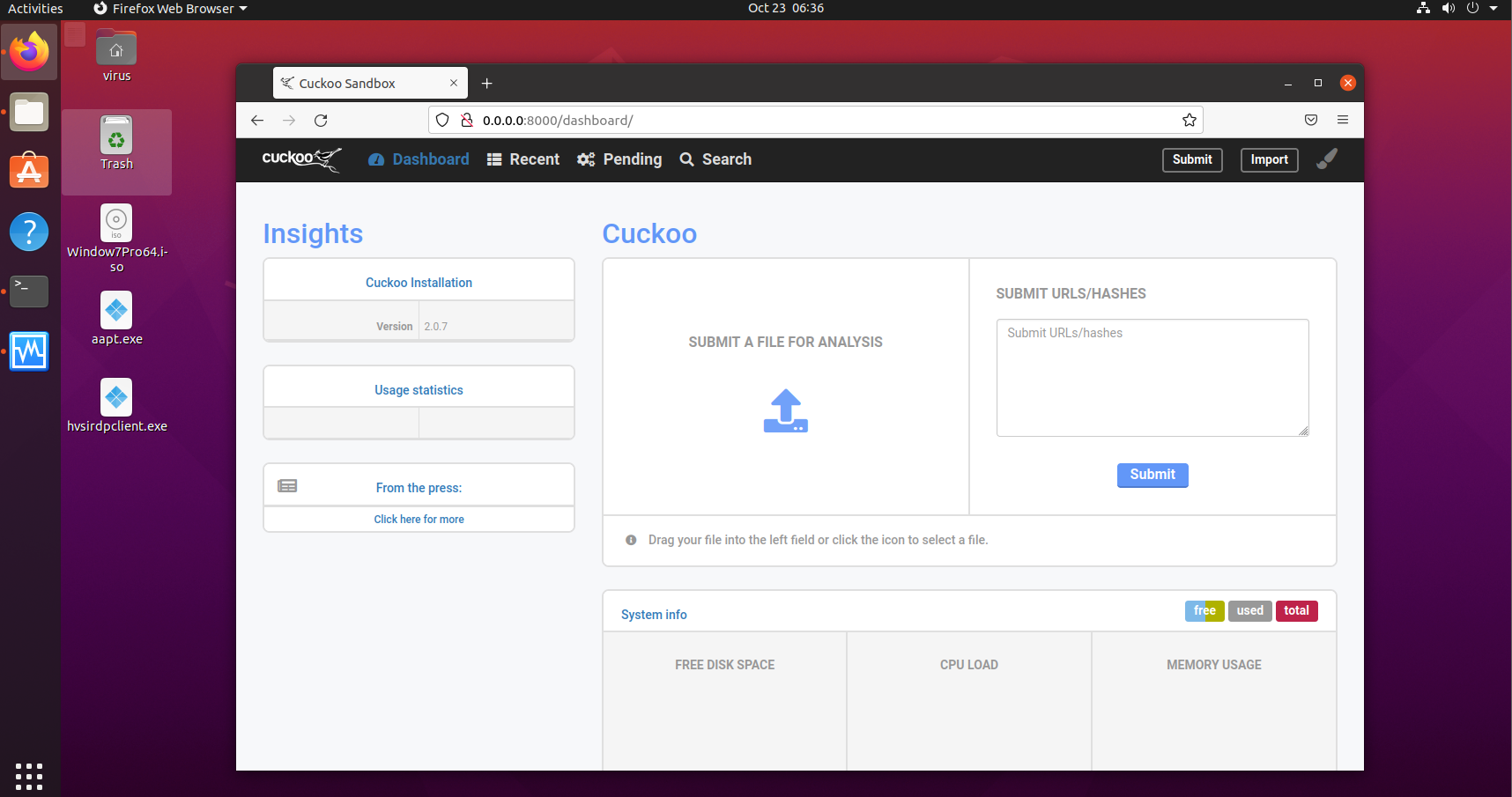
🡺 Cài đặt theo link sau (đảm bảo 100% được nhưng phải làm đúng phiên bản, kỹ và y chang theo vid): <https://www.youtube.com/watch?v=sWGmtTIzc60&ab_channel=ForeGuards>

Link Github gốc: (để cài nhanh không cần động tới vid)

<https://github.com/ForeGuards/Cuckoo-Installation-Guide/blob/main/installation.txt#L186>

**Lưu ý:** Thêm một bước quan trọng trong khi cài (Video không nói). Tắt hẳn “Window Firewall” trên Window 7 trước khi lưu Snapshot: Snapshot1

Sau khi cài xong, ta có giao diện web:



Phần demo trích xuất feature file còn lại, hãy xem video của mình thao tác trên

<https://youtu.be/jw74rjz_q80>

|  |
| --- |
| ***4.*** *Tương tự sinh viên hãy làm các câu truy vấn về Python và Powershell* |

**<Xem kết quả chi tiết tại file Notebook (.ipynb)>**

Kết quả trích xuất vào 3 folder như trong code:

Graphical user interface, application

Description automatically generated

**Note**: Folder “Sample” không liên quan.

|  |
| --- |
| **5.** *Sinh viên cho biết quả của đoạn code trên* |

**<Xem kết quả chi tiết tại file Notebook (.ipynb)>**

Kết quả đạt 95.8%

Text

Description automatically generated

|  |
| --- |
| **6**. *Sinh viên cho biết quả của đoạn code trên* |

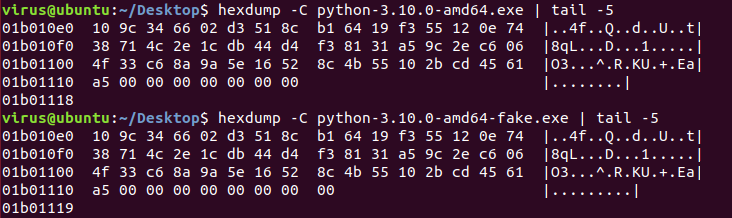
File đầu tiên test hash và độ giống nhau của 4 string

Text

Description automatically generated

- Ở 4 dòng đầu, ta thấy 4 file cho ra mã hash hoàn toàn khác nhau vì nội dung khác biệt nhau. Ở 4 dòng sau, so sánh hash1-hash1 (giống nhau hoàn toàn), hash1-hash2 và hash1-hash3(gần giống) và hash1-hash4(hoàn toàn khác).

Tiếp theo, ta có một file “Cài đặt python” **gốc** và một file **fake** được thêm *1 byte vào cuối* *file* bằng lệnh “truncate”.



Chạy code check độ giống nhau tương tự như khi so sánh với chuỗi ta có kết quả được xem là giống nhau 100%



Vậy thêm 1 byte hoàn toàn trick được SSDeep.

|  |
| --- |
| **7.** *Sinh viên cho biết quả của đoạn code trên* |

**<Xem kết quả chi tiết tại file Notebook (.ipynb)>**

Kết quả khi lấy N-grams với từng phương pháp:

A screenshot of a computer

Description automatically generated with medium confidence

|  |
| --- |
| **8.** *Sinh viên hoàn thành các bước trên* |

**<Xem các bước chi tiết tại file Notebook (.ipynb)>**

Đã comment giải thích