**ĐẠI HỌC PHENIKAA**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN PHENIKAA**

-----\*\*\*-----

**A logo of a university

Description automatically generated**

BÀI TẬP LỚN**: AN TOÀN VÀ BẢO MẬT THÔNG TIN**

***ĐỀ TÀI: “Ứng dụng Log Analysis phát hiện brute-force attack”***

Giảng viên hướng dẫn: **Đoàn Trung Sơn**

Nhóm thực hiện: **Nhóm 11**

Thành viên:

**Nguyễn Trọng Tuấn - 23010690**

**Hoàng Xuân Phong - 23010217**

Lớp:[**An toàn và bảo mật thông tin-1-1-25(N01.TH2)**](https://canvas.phenikaa-uni.edu.vn/courses/19174)

**Mục lục**

[I . Lý do chọn đề tài 3](#_Toc211646170)

[II. Mục tiêu đề tài 4](#_Toc211646171)

[1. Mục tiêu tổng quát 4](#_Toc211646172)

[2. Các mục tiêu cụ thể 4](#_Toc211646173)

[3. Ý nghĩa của đề tài 4](#_Toc211646174)

[III. Cơ sở lý thuyết 5](#_Toc211646175)

[1. Nhật ký hệ thống (System Log) 5](#_Toc211646176)

[2. Tấn công Brute-Force (Tấn công dò mật khẩu) 5](#_Toc211646177)

[3. Phương pháp phát hiện dựa trên ngưỡng (Threshold-based Detection)...................................................................................5](#_Toc211646178)

[4. Các công nghệ sử dụng 5](#_Toc211646179)

[IV. Thiết kế và phát triển hệ thống 6](#_Toc211646180)

1. Kiến trúc hệ thống.................................................................6

2. Cấu trúc thư mục...................................................................6

3. Các module chính.................................................................6

4. Usecase…………………………………………………….6

[V. Hướng dẫn sử dụng 6](#_Toc211646181)

[VI. Kết quả và đánh giá 6](#_Toc211646182)

[VIII. Kết luận và hướng phát triển 7](#_Toc211646183)

# 

# I . Lý do chọn đề tài

Trong bối cảnh an toàn thông tin là yếu tố sống còn, các cuộc **tấn công dò mật khẩu** **(brute-force)** đang trở thành mối đe dọa phổ biến đối với mọi hệ thống. Việc **phân tích nhật ký** **(log)** thủ công để tìm ra các dấu hiệu này thường rất tốn thời gian và dễ bỏ sót, tạo ra một lỗ hổng an ninh đáng kể.

Để giải quyết khoảng trống này, việc phát triển một công cụ chuyên dụng, tự động hóa quá trình phân tích là hướng đi tất yếu. Một giải pháp tự động không chỉ giải phóng quản trị viên khỏi công việc đơn điệu mà còn nâng cao đáng kể khả năng phát hiện và phản ứng kịp thời trước các mối đe dọa tiềm tàng.

Chính vì vậy, đề tài **“Ứng dụng mini phân tích nhật ký đăng nhập (Login Analyzer)”** được thực hiện. Dự án tập trung vào việc xây dựng một công cụ gọn nhẹ, hiệu quả và dễ tiếp cận, cho phép người dùng nhanh chóng xác định các hành vi đáng ngờ và tăng cường an ninh cho hệ thống của mình.

**II. Mục tiêu đề tài**

## **1. Mục tiêu tổng quát**

Phát triển ứng dụng "Login Analyzer" nhằm tự động hóa việc phân tích nhật ký đăng nhập, phát hiện sớm các cuộc tấn công dò mật khẩu (brute-force), và trực quan hóa các cảnh báo an ninh một cách hiệu quả.

## **2. Các mục tiêu cụ thể**

* Nghiên cứu cơ sở lý thuyết về tấn công brute-force và các phương pháp phân tích log.
* Xây dựng module có khả năng đọc và phân tích (parsing) các tệp log đăng nhập để trích xuất thông tin cần thiết (IP, thời gian, trạng thái).
* Thiết kế thuật toán nhận diện địa chỉ IP có hành vi tấn công brute-force dựa trên ngưỡng tần suất đăng nhập thất bại.
* Phát triển giao diện web (sử dụng Streamlit) cho phép tải tệp log lên và hiển thị kết quả phân tích một cách trực quan qua biểu đồ, bảng biểu.
* Xây dựng chức năng xuất báo cáo các IP đáng ngờ ra tệp CSV để phục vụ việc lưu trữ và phân tích sâu hơn.

## **3. Ý nghĩa của đề tài**

Đề tài mang lại một công cụ gọn nhẹ, dễ sử dụng giúp các quản trị viên hệ thống nhanh chóng phát hiện các mối đe dọa an ninh, tiết kiệm thời gian so với việc phân tích thủ công. Đồng thời, đây cũng là một bài toán ứng dụng thực tế, minh họa cho khả năng của việc phân tích dữ liệu trong lĩnh vực an toàn thông tin.

**III. Cơ sở lý thuyết**

## **1. Nhật ký hệ thống (System Log)**

Nhật ký hệ thống là các tệp tin ghi lại sự kiện hoạt động của hệ điều hành hoặc ứng dụng, đóng vai trò quan trọng trong việc giám sát an ninh. Đề tài tập trung vào việc phân tích các log xác thực như SSH Log (/var/log/auth.log) và Access Log của máy chủ web, từ đó trích xuất các thông tin quan trọng như địa chỉ IP, dấu thời gian và trạng thái đăng nhập (thành công/thất bại).

## **2. Tấn công Brute-Force (Tấn công dò mật khẩu)**

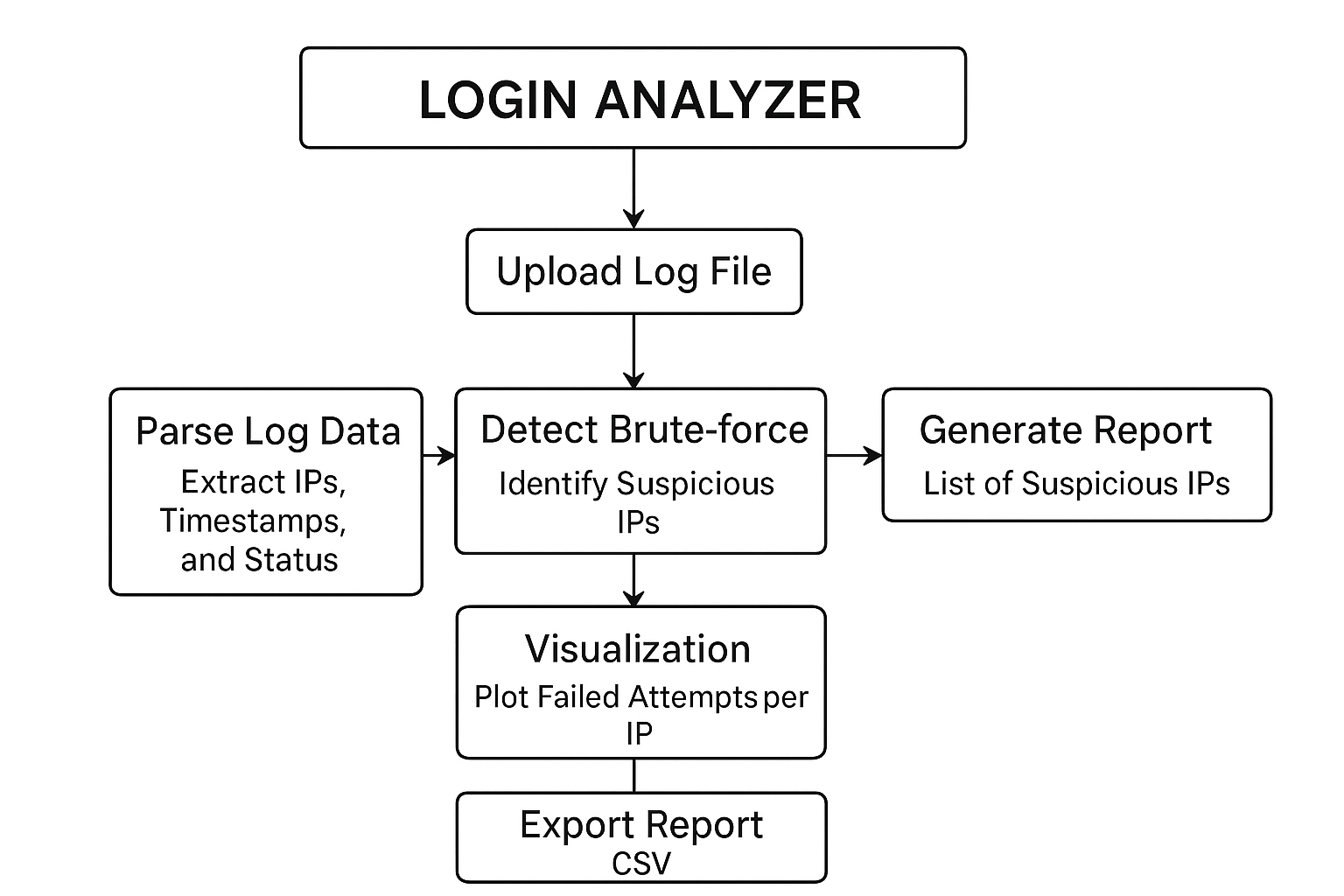
Đây là phương pháp tấn công bằng cách thử liên tục các tổ hợp mật khẩu cho đến khi thành công. Dấu hiệu đặc trưng trong log là một số lượng lớn các lần đăng nhập thất bại từ cùng một địa chỉ IP trong một khoảng thời gian ngắn, thường được thực hiện bởi các công cụ tự động. Mục tiêu của kẻ tấn công là chiếm quyền truy cập trái phép vào hệ thống.

## **3. Phương pháp phát hiện dựa trên ngưỡng (Threshold-based Detection)**

Đây là phương pháp cốt lõi của ứng dụng. Hệ thống sẽ đọc và phân tích (parse) các tệp log để đếm số lần đăng nhập thất bại từ mỗi địa chỉ IP. Sau đó, kết quả được so sánh với một ngưỡng (threshold) được định nghĩa trước (ví dụ: 10 lần thất bại trong 5 phút). Nếu một IP vượt quá ngưỡng này, nó sẽ bị đánh dấu là nghi ngờ và được đưa vào báo cáo cảnh báo.

## **4. Các công nghệ sử dụng**

* **Python:** Ngôn ngữ lập trình chính để xử lý logic và phân tích dữ liệu.
* **Streamlit:** Framework để xây dựng giao diện ứng dụng web một cách nhanh chóng.
* **Pandas:** Thư viện dùng để cấu trúc hóa, xử lý dữ liệu log và xuất báo cáo CSV.
* **Matplotlib / Plotly:** Các thư viện dùng để trực quan hóa kết quả phân tích thành các biểu đồ.

IV. Thiết kế và phát triển hệ thống  
**1. Kiến trúc hệ thống:**   
Ứng dụng gồm 3 lớp chính:   
- Giao diện người dùng (UI) — Streamlit   
- Bộ xử lý logic (Core) — Python   
- Hiển thị & báo cáo (Visualization) — matplotlib, pandas   
  
****

**Sơ đồ tổng quan kiến trúc và Flow hoạt động**

**2. Cấu trúc thư mục:**

│ login-analyzer

├── app.py # File chính chạy ứng dụng Streamlit

├── utils.py # Xử lý logic

├── requirements.txt # Danh sách các thư viện cần cài đặt

├── README.md # Tài liệu mô tả dự án

├── sample\_logs/ # Thư mục chứa dữ liệu log mẫu

│ ├── auth\_log\_01.csv # Log đăng nhập mẫu 1

│ ├── auth\_log\_02.csv # Log đăng nhập mẫu 2

│ └── README.txt # Mô tả nội dung và định dạng của logfile  
  
**3. Các module chính:**   
**3.1. Module Upload File**

* **Chức năng:**  
  Cho phép người dùng tải lên các tệp log chứa thông tin hoạt động đăng nhập của hệ thống (định dạng .csv, .txt, hoặc .json).
* **Hoạt động:**  
  Khi người dùng chọn tệp, ứng dụng sẽ đọc nội dung và chuyển dữ liệu đến module **Log Parser** để xử lý.
* **Ý nghĩa:**  
  Cung cấp giao diện thân thiện, hỗ trợ người dùng không có kiến thức kỹ thuật vẫn có thể sử dụng hệ thống.

**3.2. Module Log Parser**

* **Chức năng:**  
  Phân tích và trích xuất thông tin quan trọng từ tệp log, bao gồm: địa chỉ IP, tài khoản người dùng, thời gian đăng nhập, trạng thái đăng nhập (thành công hoặc thất bại).
* **Hoạt động:**  
  Module thực hiện quá trình làm sạch dữ liệu (chuẩn hóa định dạng thời gian, loại bỏ dòng lỗi hoặc thiếu dữ liệu) trước khi chuyển tiếp cho khối thống kê.
* **Kết quả:**  
  Tạo ra một tập dữ liệu sạch, thống nhất, sẵn sàng cho việc phân tích thống kê và phát hiện bất thường.

**3.3. Module Statistic Engine**

* **Chức năng:**  
  Tính toán và tổng hợp các chỉ số thống kê liên quan đến hoạt động đăng nhập.
* **Các thông số chính bao gồm:**
  + Tổng số lượt đăng nhập
  + Số lần đăng nhập thành công và thất bại
  + Số lượng người dùng duy nhất
  + Phân bố đăng nhập theo thời gian
* **Mục tiêu:**  
  Cung cấp cái nhìn tổng quan về tình hình đăng nhập của hệ thống trong một giai đoạn cụ thể.

**3.4. Module Visualization**

* **Chức năng:**  
  Trực quan hóa dữ liệu bằng biểu đồ, đồ thị giúp người dùng dễ dàng quan sát và so sánh các chỉ số.
* **Biểu đồ hiển thị:**
  + Biểu đồ cột: thể hiện số lượng đăng nhập theo thời gian
  + Biểu đồ tròn: thể hiện tỷ lệ đăng nhập thành công và thất bại
  + Biểu đồ đường: mô tả xu hướng hoạt động theo khung giờ hoặc ngày
* **Công nghệ:**  
  Sử dụng thư viện **Streamlit**, **Matplotlib** hoặc **Plotly** để hiển thị biểu đồ động trên giao diện web.

**3.5. Module Detection & Alert**

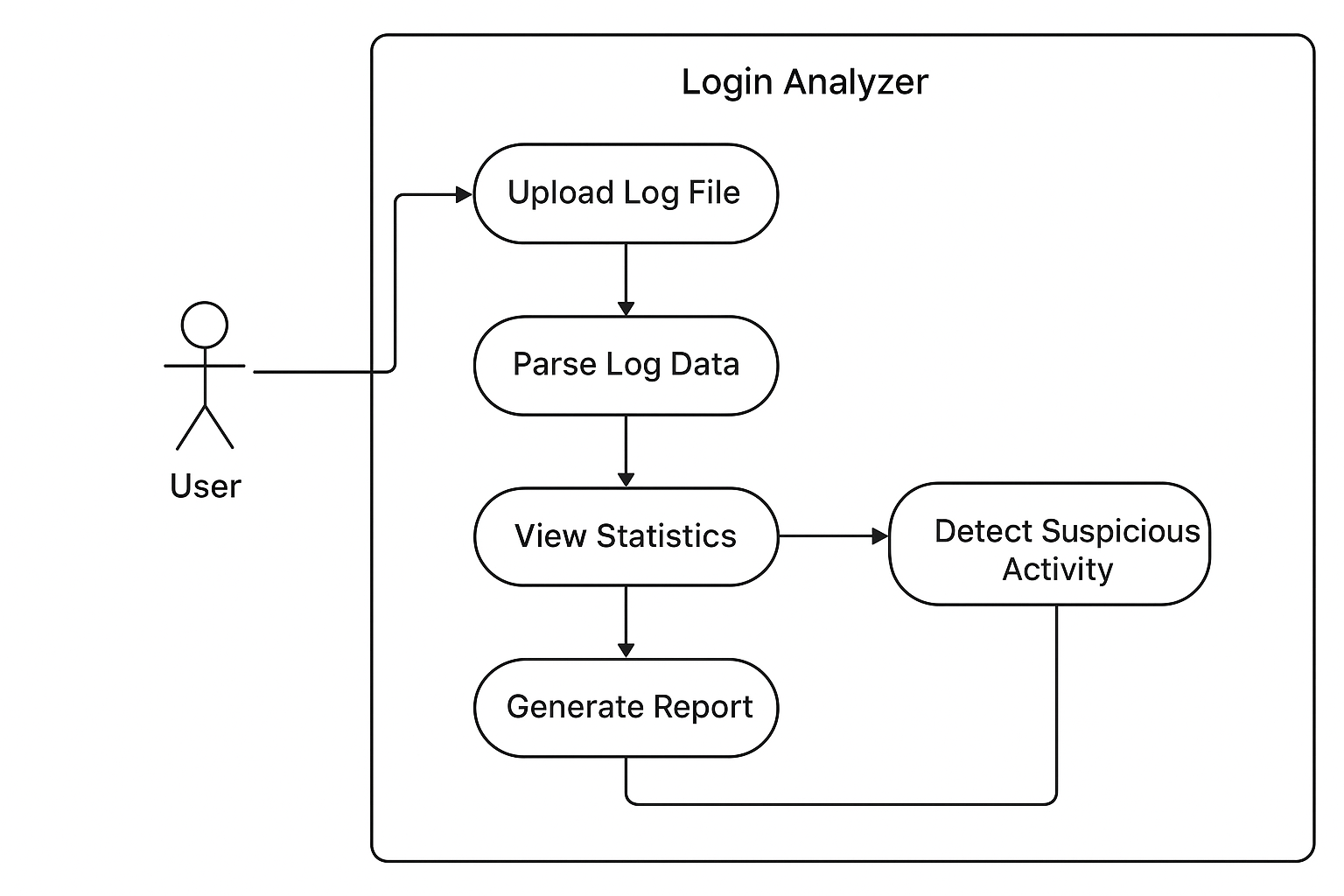
* **Chức năng:**  
  Phát hiện các dấu hiệu bất thường hoặc tiềm ẩn nguy cơ tấn công trong dữ liệu log.
* **Các quy tắc phát hiện bao gồm:**
  + Địa chỉ IP có số lần đăng nhập thất bại liên tiếp vượt ngưỡng cho phép
  + Một tài khoản đăng nhập từ nhiều IP khác nhau trong thời gian ngắn
  + Hoạt động đăng nhập vào khung giờ bất thường
* **Kết quả:**  
  Danh sách cảnh báo được hiển thị trực tiếp trên giao diện để người quản trị theo dõi.

**3.6. Module Export Report**

* **Chức năng:**  
  Tổng hợp kết quả phân tích và xuất thành báo cáo dưới dạng **CSV**, **PDF** hoặc **Excel**.
* **Nội dung báo cáo bao gồm:**
  + Thống kê tổng hợp
  + Danh sách IP nghi ngờ
  + Các biểu đồ trực quan
* **Mục đích:**  
  Giúp người dùng lưu trữ, chia sẻ hoặc nộp kết quả phân tích cho bộ phận an ninh hoặc giảng viên hướng dẫn.

**4. Usecase**

**4.1. Sơ đồ Usecase**

****

**Sơ đồ Usecase của ứng dụng Login Analyzer**

**Tác nhân (Actor):**

* **Người dùng (User/Admin):** Là người tải tệp log, xem thống kê, theo dõi cảnh báo và xuất báo cáo.
* **Hệ thống Login Analyzer:** Thực hiện việc xử lý, phân tích, trực quan hóa và lưu trữ dữ liệu log.

**4.2. Mô tả chi tiết Usecase**

**Use Case 1: Upload File Log**

| **Thuộc tính** | **Nội dung** |
| --- | --- |
| **Tên Use Case** | Upload File Log |
| **Tác nhân** | Người dùng |
| **Mô tả** | Người dùng tải lên tệp log đăng nhập (định dạng .csv hoặc .txt) để hệ thống xử lý. |
| **Tiền điều kiện** | Người dùng có tệp log hợp lệ trong máy tính. |
| **Hậu điều kiện** | Tệp log được tải lên thành công và sẵn sàng để phân tích. |
| **Luồng sự kiện chính** | 1. Người dùng nhấn “Upload File”.2. Chọn tệp log từ máy tính.3. Hệ thống đọc dữ liệu và hiển thị bản xem trước.4. Nếu file không hợp lệ, hệ thống thông báo lỗi. |

**Use Case 2: Phân tích Log (Log Parser & Statistic Engine)**

| **Thuộc tính** | **Nội dung** |
| --- | --- |
| **Tên Use Case** | Parse Log Data |
| **Tác nhân** | Hệ thống |
| **Mô tả** | Hệ thống tự động xử lý dữ liệu log: tách thời gian, địa chỉ IP, kết quả đăng nhập. |
| **Tiền điều kiện** | Đã có file log được tải lên. |
| **Hậu điều kiện** | Dữ liệu được làm sạch và chuyển sang module thống kê. |
| **Luồng sự kiện chính** | 1. Hệ thống phân tích từng dòng log.2. Trích xuất các trường dữ liệu quan trọng.3. Lưu trữ dữ liệu đã chuẩn hóa để phục vụ phân tích. |

**Use Case 3: Xem thống kê và trực quan hóa (Visualization)**

| **Thuộc tính** | **Nội dung** |
| --- | --- |
| **Tên Use Case** | View & Visualize Statistics |
| **Tác nhân** | Người dùng |
| **Mô tả** | Hiển thị số lượng đăng nhập, tỷ lệ thành công/thất bại, thống kê theo IP, theo thời gian. |
| **Tiền điều kiện** | Hệ thống đã xử lý dữ liệu log. |
| **Hậu điều kiện** | Người dùng xem được thống kê và biểu đồ trực quan. |
| **Luồng sự kiện chính** | 1. Người dùng chọn mục “Statistics”.2. Hệ thống hiển thị biểu đồ và số liệu tổng hợp.3. Người dùng có thể lọc theo thời gian hoặc IP. |

**Use Case 4: Phát hiện hoạt động bất thường (Detection & Alert)**

| **Thuộc tính** | **Nội dung** |
| --- | --- |
| **Tên Use Case** | Detect Suspicious Activity |
| **Tác nhân** | Hệ thống |
| **Mô tả** | Phân tích dữ liệu để phát hiện các IP đăng nhập thất bại nhiều lần hoặc hành vi bất thường. |
| **Tiền điều kiện** | Dữ liệu log đã được phân tích. |
| **Hậu điều kiện** | Danh sách các IP hoặc tài khoản nghi ngờ được hiển thị. |
| **Luồng sự kiện chính** | 1. Hệ thống quét toàn bộ dữ liệu.2. Phát hiện hành vi đáng ngờ theo quy tắc (login fail > 3 lần, IP lạ, v.v).3. Hiển thị cảnh báo trên giao diện. |

**Use Case 5: Xuất báo cáo (Export Report)**

| **Thuộc tính** | **Nội dung** |
| --- | --- |
| **Tên Use Case** | Export Report |
| **Tác nhân** | Người dùng |
| **Mô tả** | Cho phép xuất dữ liệu thống kê và kết quả phân tích ra tệp CSV hoặc PDF. |
| **Tiền điều kiện** | Hệ thống đã hoàn tất phân tích và có dữ liệu hiển thị. |
| **Hậu điều kiện** | File báo cáo được lưu vào máy người dùng. |
| **Luồng sự kiện chính** | 1. Người dùng nhấn “Export Report”.2. Chọn định dạng tệp (CSV/PDF).3. Hệ thống tạo file và tự động tải về. |

**V. Hướng dẫn sử dụng**

Cài đặt các thư viện cần thiết của Python

`pip install streamlit`

`pip install pandas`

`pip install matplotlib`

`pip install seaborn`

`pip install plotly`

`pip install scikit-learn`

`pip install numpy`   
**B1:** Cài thư viện: `pip install -r requirements.txt`   
**B2:** Chạy: `python -m streamlit run app.py`   
**B3:** Truy cập: http://localhost:8501   
**B4:** Upload file log, hệ thống tự động phân tích và hiển thị biểu đồ.   
**B5:** Xuất kết quả ra file CSV nếu cần lưu lại.

**VI. Kết quả và đánh giá**  
Ứng dụng hoạt động ổn định, có thể phân tích hàng ngàn dòng log trong vài giây.   
Giao diện đơn giản, hiển thị biểu đồ dễ hiểu. Kết quả thử nghiệm cho thấy hệ thống nhận diện chính xác IP brute-force.

**VIII. Kết luận và hướng phát triển**  
**1. Kết luận:**- Login Analyzer đáp ứng tốt mục tiêu ban đầu:

- Khả năng xử lí dữ liệu log thực tế

- Phát hiện tấn công brute-force hiệu quả

- Trực quan hóa dữ liệu bảo mật đơn giản mà mạnh mẽ  
**2. Hướng phát triển:**  
- Giám sát real-time.   
- Tích hợp email cảnh báo.   
- Hỗ trợ nhiều định dạng log hơn.