BÁO CÁO QUÁ TRÌNH PHÁT TRIỂN CÔNG CỤ GIÁM SÁT VÀ KIỂM TRA BẢO MẬT

# GIỚI THIỆU

Đây là một công cụ cho phép người dung giám sát trạng thái của website, đồng thời pháp hiện sơ bộ các vấn đề bảo mật tiềm ẩn dựa trên cơ sở dữ liệu CVE và công nghệ được sử dụng trên hệ thống mục tiêu.

# TÍNH NĂNG CHÍNH

## Giám sát hệ thống

### Chức năng đã phát triển

* Một thread chạy liên tục để gửi request đến mục tiêu nhằm kiểm trả status code ở response.
* Cho phép dừng nếu muốn.
* Cho phép redirect tối đa 10 lần để tránh redirect vô tận.
* Thông báo lỗi trên trình duyệt khi lỗi kết nối quá 5 lần (ConnectionAborted, 4xx, 5xx, …)

### Khó khăn, vướng mắc

* Hiện tại, chưa phát triển thêm cơ chế nào khác để phát hiện trạng thái website ngoài status code.

### Hướng phát triển tiếp theo

* Bổ sung khả năng đo response time, theo dõi hiệu năng.

## Dò quét lỗ hổng

### Chức năng đã phát triển

* Kiểm tra công nghệ web được sử dụng bằng Wappalyzer.
* Kiểm tra CPE, từ đó tìm CVE “có thể” tồn tại trên mục tiêu dựa vào CPE và cơ sở dữ liệu của NVD.
* Kiểm tra tính hiện diện của 1 CVE bằng các kịch bản của Nuclei
* Cập nhật dữ liệu **recent, modified** mỗi 2 tiếng; Cập nhật lại **toàn bộ dữ liệu** và **template Nuclei** mỗi 1 ngày.
* Kiểm tra cập nhật lần cuối và thực hiện cập nhật ngay lập tức nếu trễ lịch.

### Khó khăn, vướng mắc

* Thư viện Wappalyzer được sử dụng không phải từ Wappalyzer gốc, không thể đảm bảo được tính **chính xác** và tính **đủ** của engine quét công nghệ. Đây là một điều quan trọng, vì các module sau đó phụ thuộc rất nhiều vào module đầu tiên này. Cho đến hiện tại, tạm thời hiệu suất và độ chính xác vẫn ổn so với bản chính thức, nhưng không thể đảm bảo cho tương lai.
* Danh sách CPE không có format rõ ràng, do chỉ những công nghệ hoặc 1 dải version của công nghệ đó có CVE ảnh hưởng mới được ghi vào. Điều này dẫn đến việc quá trình index có thể bị trùng lặp CPE, làm tăng thời gian index lên nhiều.
* Số lượng template (kịch bản) của Nuclei là quá ít so với số lượng CVE, đồng thời tính chính xác còn phụ thuộc vào chất lượng của template. Hơn nữa, có những template yêu cầu thực hiện chỉnh sửa thủ công mới có thể thực hiện kịch bản trong template (Với template yêu cầu xác thực, người dùng phải cung cấp *username* và *password* thủ công). Vì vậy, giải pháp này chỉ có thể đảm bảo *true positive*, còn không đảm bảo **false positive**, **true negative** và **false negative**.
* Hiện tại, cơ sở dữ liệu đang được kéo về từ NVD Data Feeds. Dữ liệu được tải về và xử lý tại local. Cập nhật dữ liệu theo cách này đem lại nhiều lợi ích, bao gồm: cải thiện tốc độ truy vấn, không phụ thuộc vào API và kết nối Internet. Tuy nhiên, việc **cập nhật toàn bộ** dữ liệu diễn ra khá lâu (gần 10’). Hơn nữa, đã có thông báo chính thức của NVD là Data Feeds sẽ đóng cửa sớm để nhường chỗ cho API. Vì vậy, giải pháp này chỉ là tạm thời và không đảm bảo tính lâu dài.
* Về quét **lỗ hổng bảo mật/lỗ hổng nghiệp vụ**:
  + Tự động/bán tự động hóa quá trình quét **lỗ hổng bảo mật/lỗ hổng nghiệp vụ** là một bài toán khó. Chiến lược ban đầu đặt ra gồm 2 bước chính: *Crawling* và *Scanning*, và cả 2 bước này đều phát sinh rất nhiều trở ngại:
  + Về Crawling:
    - Khó tự động hóa quá trình xác thực do quá nhiều cơ chế đăng nhập.
    - Quản lý session phức tạp.
  + Về Scanning:
    - Lỗ hổng logic/luồng nghiệp vụ vẫn cần test thủ công.
    - Có nhiều payload nhưng chưa có cơ chế chọn lọc hiệu quả, dẫn đến thời gian quét quá lâu.
  + Về công cụ:
    - Hiện tại, Burp Suite Pro đang là công cụ rất mạnh, các công cụ khác tích hợp lại rất khó khăn mà hiệu quả không cao bằng Burp Suite Pro.

### Hướng phát triển tiếp theo

* Cải thiện thuật toán lọc CPE để tăng tốc độ index và độ chính xác, giảm trung lặp.
* Xử lý thêm trường hợp template Nuclei cần payload (Tìm kiếm template, sau đó chọn ra những template cần điền payload. Sau khi người dùng điền payload, thực hiện quét bằng Nuclei).
* Chuyển đổi từ **NVD Data Feeds** sang **NVD API** khi Data Feeds bị gỡ bỏ.
* Tích hợp ML để phân tích response, tạo ra các payload thông minh hơn, tự động đăng nhập và duy trì session.

## Quản lý chế độ quét

### Chức năng đã phát triển

* Cho phép người dùng thao tác đến đối tượng (theo dõi/bỏ theo dõi, quét thủ công, xóa).
* Danh sách các trang web được theo dõi, liên tục kiểm tra trạng thái các trang web đó và quét tự động theo lịch trình.
* Liên tục theo dõi WAF mỗi phút.

### Khó khăn, vướng mắc

* Do đầu vào chỉ có URL, chỉ có thể sử dụng các công cụ có sẵn như wafw00f (Kiểm tra WAF dựa vào dấu hiệu ở header response) để kiểm tra WAF dựa trên 1 tập WAF được định nghĩa sẵn. Vì vậy, với những WAF không có trong cơ sở dữ liệu của công cụ, chưa thể kiểm tra được.
* Chưa thể test nhiều chức năng theo dõi WAF do không thể tự set up môi trường test.
* Có thể xem xét để tích hợp thêm check WAF bằng nslookup với các WAF dạng cloud-based (Nếu được cung cấp IP của WAF).

### Hướng phát triển tiếp theo

* Kiểm tra loại WAF, kiểm tra xem có đang online không.

## Báo cáo và cảnh báo

### Chức năng đã phát triển

* Cảnh báo qua Gmail khi phát hiện CVE mới/CVE được chỉnh sửa/Mất kết nối với URL mục tiêu.

### Khó khăn, vướng mắc

None

### Hướng phát triển tiếp theo

* Tạo log để debug.

## Tối ưu hiệu suất, ẩn danh

### Chức năng đã phát triển

None

### Khó khăn, vướng mắc

None

### Hướng phát triển tiếp theo

* Cải thiện hiệu quả quét, giảm nguy cơ bị phát hiện là bot bằng cách random hóa và giới hạn tần suất gửi request.

# HƯỚNG DẪN SỬ DỤNG