XÂY DỰNG GIẢI PHÁP PHÁT HIỆN LIÊN KẾT LỪA ĐẢO SỬ DỤNG MÁY HỌC

Lê Tôn Nhân - 240202025

Tóm tắt

- Lớp: CS2205.FEB2025
- Link Github của nhóm: https://github.com/Le-Ton-Nhan/CS2205.FEB2025
- Link YouTube video: https://www.youtube.com/watch?v=YnO4q4YSuDs
- Họ và Tên: Lê Tôn Nhân - 240202025

Giới thiệu

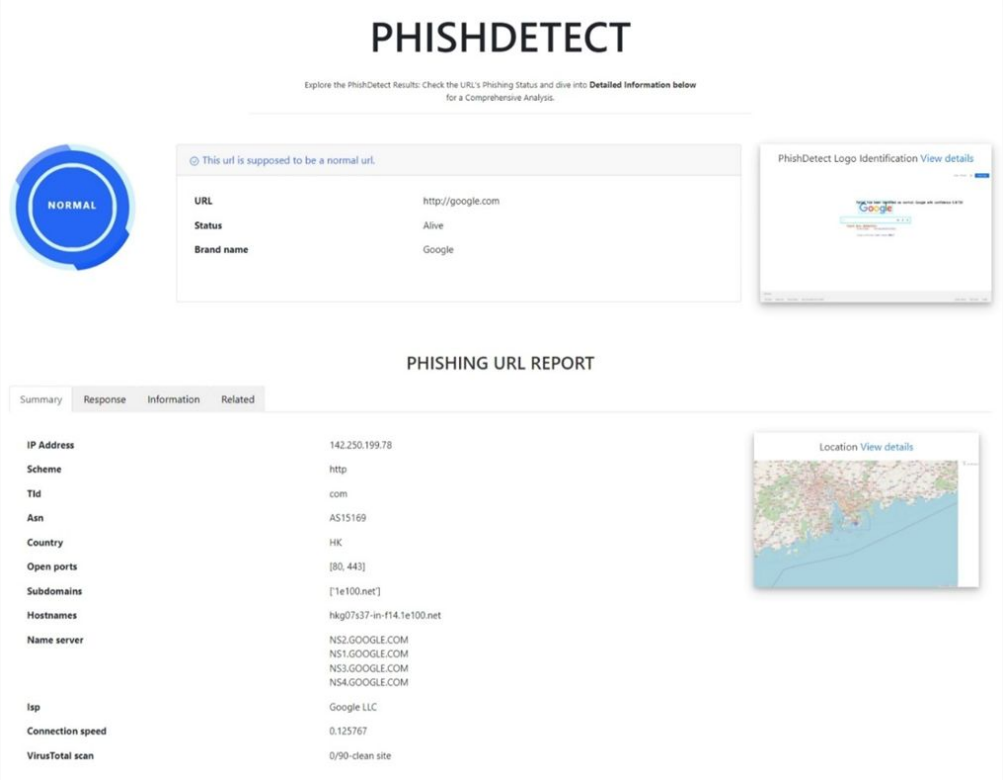
- Tấn công lừa đảo đã phát triển một cách nhanh chóng, ngày càng tinh vi và đa dạng, đến mức hiện nay, nó được cung cấp như một dịch vụ.
- Trước tình hình đó, các nhà nghiên cứu đã đề xuất nhiều phương pháp phát hiện liên kết lừa đảo khác nhau, trong đó nổi bật là hai hướng chính:
  - ***Dựa trên thông tin tĩnh***: Sử dụng các mô hình học máy để phân tích các đặc trưng như cấu trúc URL hoặc mã HTML.
  - ***Dựa trên thông tin giao diện***: So sánh các đặc điểm hình ảnh giao diện (thường là logo) với cơ sở dữ liệu các thương hiệu chính thống.

Giới thiệu

- **Nhược điểm:**
  - ***Dựa trên thông tin tĩnh***: Phụ thuộc vào dữ liệu gắn nhãn, đòi hỏi công sức lớn trong thu thập và cập nhật. Kẻ tấn công có thể áp dụng kỹ thuật né tránh để vượt qua hệ thống.
  - ***Dựa trên thông tin giao diện***: Gặp khó khăn trong việc theo kịp sự thay đổi liên tục của giao diện website và cần cập nhật thường xuyên cơ sở dữ liệu thương hiệu.

Giới thiệu

- **Input**: đoạn văn bản hoặc email, tin nhắn SMS chứa liên kết muốn kiểm tra.
- **Output**: kết quả dự đoán liên kết có phải là lừa đảo hay không, cùng các đặc trưng trích xuất từ liên kết được cung cấp.



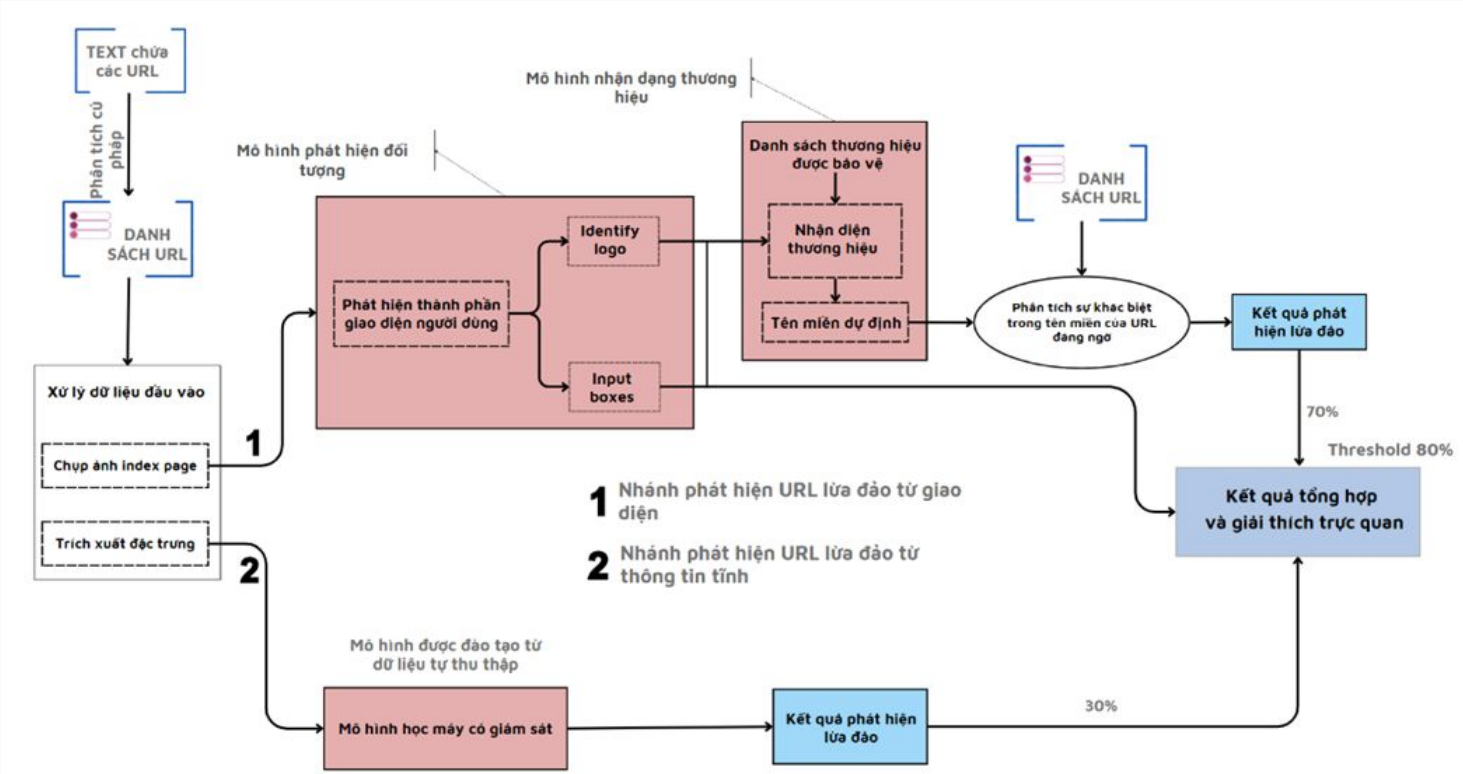
Hình 1: Minh hoạ trang hiển thị kết quả chính trên giao diện web

Mục tiêu

- Cung cấp một hệ thống phát hiện hiệu quả và chính xác để xác định liên kết lừa đảo, giúp người dùng tránh tiếp cận và rơi vào các cuộc tấn công lừa đảo trực tuyến.
- Các module trong hệ thống được thiết kế tường minh và mở rộng, nhằm hỗ trợ cho các nghiên cứu tiếp theo và phát triển công nghệ phát hiện lừa đảo.
- Nâng cao chất lượng bộ dữ liệu, đảm bảo tính đáng tin cậy và đa dạng, tạo cơ sở cho các nghiên cứu và phát triển tiếp theo trong lĩnh vực phát hiện liên kết lừa đảo.

Nội dung và Phương pháp

- Hệ thống đề xuất kết hợp phát hiện liên kết lừa đảo từ thông tin tĩnh và thông tin giao diện.



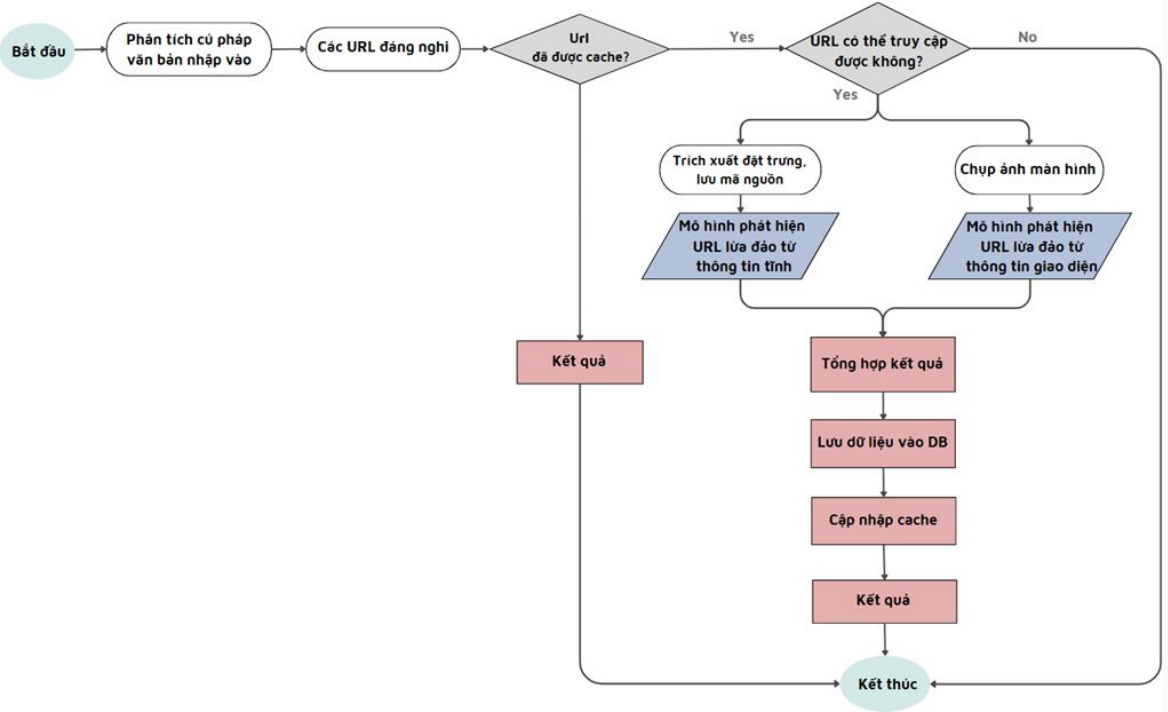
Hình 2: Tổng quan hệ thống đề xuất

Nội dung và Phương pháp

- ***Phát hiện liên kết lừa đảo từ thông tin tĩnh***: Trích xuất các đặc trưng từ liên kết, có thể chia các đặc trưng thành 3 loại:
  - Lexical: Đặc trưng từ chuỗi liên kết như độ dài URL, số lượng chữ số, tham số,... phản ánh cấu trúc đường dẫn.
  - External service: Thông tin từ máy chủ như quốc gia đăng ký, tên miền, cổng mở, thời gian tồn tại,... giúp nhận diện nguồn gốc liên kết.
  - Content-based: Phân tích mã HTML để phát hiện các yếu tố đáng ngờ như thẻ script, tệp nhúng, đối tượng ẩn,...
- ***Phát hiện liên kết lừa đảo từ thông tin giao diện***:
  - Phát hiện logo và ô nhập liệu bằng mô hình học sâu (sử dụng ResNet50, RPN và Fast-RCNN). So sánh logo phát hiện được với danh sách logo thương hiệu mục tiêu bằng mô hình Siamese để xác định tên miền dự kiến. Nếu tên miền thực tế không khớp với tên miền dự kiến, liên kết bị đánh giá là lừa đảo.

Kết quả dự kiến

- Hệ thống được phát triển dưới dạng ứng dụng web với 4 module chính:
  - Xử lý dữ liệu đầu vào
  - Phát hiện liên kết lừa đảo từ thông tin tĩnh
  - Phát hiện liên kết lừa đảo từ thông tin giao diện
  - Tổng hợp kết quả từ 2 nhánh và đưa ra kết luận



Hình 3: Luồng hoạt động của ứng dụng web

Tài liệu tham khảo

[1] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, Gail-Joon Ahn: Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale. USENIX Security Symposium 2020: 361-377

[2] Changbo Hu, Qun Li, Zhen Zhang, Keng-hao Chang, Ruofei Zhang: A Multimodal Fusion Framework for Brand Recognition from Product Image and Context. ICME Workshops 2020: 1-4

[3] Penghui Zhang, Adam Oest, Haehyun Cho, Zhibo Sun, RC Johnson, Brad Wardman, Shaown Sarker, Alexandros Kapravelos, Tiffany Bao, Ruoyu Wang, Yan Shoshitaishvili, Adam Doupé, Gail-Joon Ahn: CrawlPhish: Large-scale Analysis of Client-side Cloaking Techniques in Phishing. SP 2021: 1109-1124

[4] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao: YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. CVPR 2023: 7464-7475

[5] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, Jian Sun: YOLOX: Exceeding YOLO Series in 2021. CoRR abs/2107.08430 (2021)

[6] Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao: You Only Learn One Representation: Unified Network for Multiple Tasks. CoRR abs/2105.04206 (2021)

[7] Qiang Chen, Yingming Wang, Tong Yang, Xiangyu Zhang, Jian Cheng, Jian Sun: You Only Look One-Level Feature. CVPR 2021: 13039-13048

[8] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao: YOLOv4: Optimal Speed and Accuracy of Object Detection. CoRR abs/2004.10934 (2020)

[9] Yun Lin, Jun Sun, Gordon Fraser, Ziheng Xiu, Ting Liu, and Jin Song Dong. Recovering fitness gradients for interprocedural boolean flags in search-based testing. In ISSTA, pages 440–451, 2020.

[10] Diego A. Velázquez, Josep M. Gonfaus, Pau Rodríguez, F. Xavier Roca, Seiichi Ozawa, Jordi Gonzàlez: Logo Detection With No Priors. IEEE Access 9: 106998-107011 (2021)