

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

**Nghiên cứu và triển khai thử nghiệm hệ thống phát
hiện mã độc trong lưu lượng mã hóa - Phân hệ
WebUI**

Đàm Ngọc Khánh

khanh.dn205207@sis.hust.edu.vn

Ngành: Công nghệ thông tin Việt Pháp

Giảng viên hướng dẫn: ThS. Bùi Trọng Tùng

Chữ kí GVHD

Khoa: Kỹ thuật máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 06/2024

LỜI CẢM ƠN

Trong quá trình 4 năm học tập và rèn luyện tại trường, tri thức quý giá là thứ mà tôi có được. Đầu tiên, tôi xin bày tỏ lòng biết ơn sâu sắc tới các thầy cô trong Đại học Bách Khoa Hà Nội nói chung và các thầy cô trường Công nghệ thông tin và Truyền thông nói riêng đã luôn tận tình giúp đỡ tôi trong thời gian qua. Những kiến thức quý giá đó sẽ trở thành trang bị giúp tôi vững bước hơn trên con đường học tập và làm việc trong tương lai.

Đặc biệt, tôi xin gửi lời cảm ơn chân thành đến thầy Thạc sĩ Bùi Trọng Tùng đã tận tình tận tụy dẫn dắt, định hướng tôi trong suốt quá trình nghiên cứu và làm đồ án tốt nghiệp. Hơn nữa, thầy còn luôn quan tâm, hỗ trợ trong quá trình học tập, giúp tôi có được định hướng rõ ràng hơn cho tương lai của mình.

Tôi xin gửi lời cảm ơn đến gia đình đã đặt niềm tin hoàn toàn vào tôi, luôn tạo điều kiện học tập tốt nhất cho tôi. Cảm ơn anh, chị, bạn bè đã luôn đồng hành và giúp đỡ tôi trong lúc tôi cảm thấy bế tắc.

LỜI CAM KẾT

Họ và tên sinh viên: Đàm Ngọc Khánh

Điện thoại liên lạc: 0347988608

Email: khanh.dn205207@sis.hust.edu.vn

Lớp: Công nghệ thông tin Việt - Pháp 01 K65

Hệ đào tạo: Cử nhân

Tôi – Đàm Ngọc Khánh – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *học hàm học vị+điền tên giáo viên hướng dẫn*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

Hà Nội, ngày tháng năm

Tác giả ĐATN

Đàm Ngọc Khánh

TÓM TẮT NỘI DUNG ĐỒ ÁN

Hiện nay, giao thức kết nối HTTPS đang được sử dụng phổ biến trên toàn cầu, là tiêu chí bảo mật chung giúp kết nối của người dùng "an toàn" hơn. Lợi dụng điều này, các lưu lượng độc hại cũng được mã hóa, thường ẩn trong các lưu lượng bình thường. Trong thời đại mạng không gian phát triển, việc an toàn bảo mật thông tin người dùng là mục tiêu quan trọng hàng đầu. Một số doanh nghiệp sử dụng các giải pháp như phân tích luồng lưu lượng hoặc bắt và giải mã gói tin để phát hiện tấn công. Bởi chi phí cao và khi giải mã nội dung sẽ ảnh hưởng đến quyền riêng tư nên tiếp cận bằng cách sử dụng các phương pháp học máy dần trở nên phổ biến.

Trong phạm vi đồ án này, tôi cùng với Vũ Minh Long đã đề xuất xây dựng thử nghiệm một hệ thống phát hiện tấn công lưu lượng mạng mã hóa HTTPS bằng phương pháp học máy. Với việc sử dụng các phương pháp học máy truyền thống như phát hiện trên từng kết nối độc lập sẽ không thể tận dụng được quan hệ giữa các kết nối trong mạng. Hơn nữa, một ngày có rất nhiều kết nối, dù độ chính xác có cao thì không tránh khỏi có rất nhiều những cảnh báo sai. Vì vậy, chúng tôi nghiên cứu thử nghiệm một mô hình mạng nơ ron và sử dụng các phương pháp trích chọn đặc trưng, tiếp đó thực nghiệm với các bộ dữ liệu khác nhau để đưa ra độ chính xác lên tới ... % Sau đó tôi đã xây dựng một hệ thống WebUI kết hợp với BackEnd API của Long để xây dựng một hệ thống phát hiện mã độc tính hợp với hệ thống giám sát mạng, sử dụng Zeek để lấy lưu lượng theo thời gian thực.

Đồ án này đưa ra cách thức xây dựng mô đun phát hiện và các phương pháp trích chọn đặc trưng cùng với kết quả thực nghiệm đạt được trên một số bộ dữ liệu, đồng thời xây dựng một hệ thống phát hiện mã độc qua lưu lượng HTTPS cùng với các chức năng cập nhật mô hình với dữ liệu mới đã được gán nhãn, cập nhật thông số cấu hình và đưa ra các thống kê, báo cáo.

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và nhiệm vụ của đề tài.....	2
1.3 Định hướng giải pháp.....	3
1.4 Đóng góp của đồ án	5
1.5 Bố cục đồ án	6
CHƯƠNG 2. NỀN TẢNG LÝ THUYẾT	7
2.1 Giới thiệu về mạng nơ ron đồ thị GNN	7
2.1.1 Khái niệm đồ thị	7
2.1.2 Khái niệm mạng nơ ron đồ thị	7
2.1.3 Các loại mạng nơ ron đồ thị	8
2.1.4 Một số nhiệm vụ của mạng nơ ron đồ thị GNN	8
2.2 Công cụ Zeek.....	9
CHƯƠNG 3. Xây dựng mô hình phát hiện mã độc sử dụng mạng nơ ron đồ thị	12
3.1 Mô hình NF-GNN	12
3.2 Bộ xử lý đặc trưng (Feature Processor)	12
3.2.1 Trích xuất đặc trưng	12
3.2.2 Chuẩn hóa dữ liệu	13
3.3 Bộ xây dựng đồ thị (NFG-Constructor).....	13
3.4 Bộ phát hiện mã độc	14
3.5 Trích chọn đặc trưng	15
3.5.1 Mutual Information	17
3.5.2 Hồi quy Lasso.....	18

3.6 Bộ dữ liệu thực nghiệm	18
CHƯƠNG 4. Xây dựng hệ thống phát hiện tấn công	19
4.1 Biểu đồ Usecase tổng quan	19
4.2 Phân tích chức năng	20
4.2.1 Use case Xem thống kê và báo cáo	20
4.2.2 Use case Thiết lập thông số cấu hình.....	21
4.2.3 Use case Phát hiện theo thời gian thực	23
4.2.4 Use case Cập nhật mô hình	24
4.2.5 Use case quản lý người dùng	26
4.3 Thiết kế cấu trúc.....	31
4.4 Thiết kế giao diện	33
4.4.1 Bố cục chung của giao diện hệ thống	33
4.4.2 Bố cục chung của giao diện chức năng Thống kê và báo cáo	33
4.4.3 Bố cục của giao diện chức năng Thiết lập thông số cấu hình	34
4.4.4 Bố cục của giao diện chức năng Phát hiện theo thời gian thực	34
4.4.5 Bố cục giao diện chức năng Cập nhật mô hình.....	35
4.4.6 Bố cục giao diện chức năng Quản lý người dùng	35
CHƯƠNG 5. Triển khai và đánh giá thử nghiệm hệ thống phát hiện mã độc	37
5.1 Môi trường thực nghiệm.....	37
5.1.1 Môi trường thực nghiệm mô hình mạng nơ ron đồ thị phát hiện mã độc trong lưu lượng mạng HTTPS	37
5.1.2 Môi trường thực nghiệm hệ thống phát hiện mã độc trong lưu lượng mạng HTTPS.....	37
5.2 Kết quả thực nghiệm mô hình mạng nơ ron đồ thị.....	37
5.3 Kịch bản kiểm thử chức năng Thống kê và báo cáo	41
5.4 Kịch bản kiểm thử chức năng Thiết lập thông số cấu hình	42

5.5 Kịch bản kiểm thử cho chức năng Phát hiện theo thời gian thực	43
5.6 Kịch bản kiểm thử chức năng Cập nhật mô hình	44
5.7 Kịch bản kiểm thử chức năng Quản lý người dùng	46
5.8 Đánh giá chất lượng mã nguồn với SonarQube	51
CHƯƠNG 6. KẾT LUẬN	53
6.1 Kết luận	53
6.2 Hướng phát triển trong tương lai	53
TÀI LIỆU THAM KHẢO.....	54

DANH MỤC HÌNH VẼ

Hình 1.1	Quy trình xây dựng mô hình	3
Hình 1.2	Kiến trúc chung của hệ thống phát hiện mã độc trong lưu lượng mạng HTTPS	5
Hình 2.1	Mạng nơ ron đồ thị	7
Hình 2.2	Minh họa các nhiệm vụ của mạng nơ ron đồ thị	9
Hình 2.3	Zeek Cluster	10
Hình 4.1	Biểu đồ Usecase tổng quan	19
Hình 4.2	Biểu đồ tuần tự Thống kê và báo cáo	21
Hình 4.3	Biểu đồ tuần tự Thiết lập thông số cấu hình	23
Hình 4.4	Biểu đồ tuần tự use case Phát hiện theo thời gian thực	24
Hình 4.5	Biểu đồ tuần tự use case Cập nhật mô hình	26
Hình 4.6	Biểu đồ tuần tự use case Quản lý người dùng	28
Hình 4.7	Thiết kế cấu trúc WebUI của hệ thống	32
Hình 4.8	Phân vùng bố cục giao diện hệ thống	33
Hình 4.9	Phân vùng bố cục giao diện chức năng Thống kê và báo cáo	34
Hình 4.10	Phân vùng bố cục giao diện chức năng Thiết lập thông số cấu hình	34
Hình 4.11	Phân vùng bố cục giao diện chức năng Phát hiện theo thời gian thực	35
Hình 4.12	Phân vùng bố cục giao diện chức năng Cập nhật mô hình	35
Hình 4.13	Phân vùng bố cục giao diện chức năng Quản lý người dùng	36
Hình 5.1	Kết quả thực nghiệm với phương pháp Mutual Information	38
Hình 5.2	Kết quả thực nghiệm với phương pháp hồi quy Lasso	39
Hình 5.3	Giao diện Thống kê và báo cáo trước thay đổi	41
Hình 5.4	Giao diện Thống kê và báo cáo sau khi thay đổi	42
Hình 5.5	Giao diện Thiết lập thông số cấu hình	42
Hình 5.6	Giao diện sau khi Thiết lập thông số cấu hình thành công	43
Hình 5.7	Kết quả dự đoán khi gửi giả lập các lưu lượng từ file PCAP được dán nhãn bình thường	43
Hình 5.8	Kết quả dự đoán khi gửi giả lập các lưu lượng từ file PCAP được dán nhãn độc hại	44
Hình 5.9	Giao diện khi người dùng chọn chức năng Cập nhật mô hình mà mô hình không trong trạng thái đang cập nhật	44

Hình 5.10	Giao diện khi người dùng tải file không hợp lệ	45
Hình 5.11	Giao diện kết quả khi người dùng tải file không hợp lệ . . .	45
Hình 5.12	Giao diện khi người dùng tải file hợp lệ	46
Hình 5.13	Giao diện khi người dùng chọn chức năng Cập nhật mô hình khi mô hình đang được cập nhật	46
Hình 5.14	Giao diện Quản lý người dùng	47
Hình 5.15	Giao diện Quản lý người dùng khi thực hiện tìm kiếm	47
Hình 5.16	Giao diện Quản lý người dùng khi chọn thêm người dùng . . .	48
Hình 5.17	Giao diện Quản lý người dùng khi chọn thêm người dùng với đầu vào theo bảng 5.7	49
Hình 5.18	Giao diện Quản lý người dùng khi chọn thêm người dùng với đầu vào theo bảng 5.8	50
Hình 5.19	Giao diện Quản lý người dùng khi chọn xóa người dùng	50
Hình 5.20	Giao diện Quản lý người dùng khi chọn xóa người dùng và thực hiện thành công	51
Hình 5.21	Kết quả đánh giá chất lượng phần mềm với công cụ SonarQube	52