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 tình dẫn dắt, định hường 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 tính 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 trong 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

DANH MỤC HÌNH VẼ