ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Phát hiện lưu lượng phần mềm truy cập từ xa trong mạng nội bộ ứng dụng học sâu

TRẦN ĐÌNH KIẾN GIANG
giang.tdk194265@sis.hust.edu.vn

Ngành: Công nghệ thông tin

Giảng viên hướng dẫn: PGS.TS. Trần Quang Đức _____

Chữ kí GVHD

Khoa: Kỹ thuật máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 01/2024

LỜI CẢM ƠN

Đầu tiên, em xin gửi lời cảm ơn tới PSG.TS. Trần Quang Đức, Giám đốc Trung tâm An toàn An ninh Thông tin, Đại học Bách Khoa Hà Nội, là giảng viên hướng dẫn cho em trong quá trình làm đồ án tốt nghiệp. Thầy là người đưa ra ý tưởng đề tài, nhiệt tình hướng dẫn và truyền đạt kiến thức cho em giúp em có thể hoàn thành đồ án này. Em xin cảm ơn TS. Tống Văn Vạn đã hướng dẫn em trong quá trình thu thập dữ liệu và xây dựng mô hình. Em cũng xin gửi lời cảm ơn tới các thầy cô ở Đại học Bách Khoa Hà Nội, đặc biệt là các thầy cô ở trường Công nghệ Thông tin và Truyền thông đã tận tâm giảng dạy, truyền đạt những kiến thức quý báu cho em trong suốt quá trình em học tập tại trường. Đồng thời, em xin cảm ơn các thầy và các anh ở Trung tâm An toàn An ninh thông tin Bách Khoa đã tạo điều kiện về máy móc giúp em thực hiện đồ án một cách thuận lợi hơn. Cuối cùng, em xin gửi lời cảm ơn tới gia đình, bạn bè, đồng nghiệp đã hỗ trợ và đồng hành cùng em trong suốt những năm tháng học tập và rèn luyện tại trường.

Em xin chân thành cảm ơn !

TÓM TẮT NỘI DUNG ĐỒ ÁN

Đồ án đề xuất một hệ thống phát hiện và quản lý các luồng truy cập cập từ xa sử dụng phần mềm trong mạng nội bộ. Đồ án tập trung vào hai nhiệm vụ chính, đó là xây dựng và huấn luyện một số mô hình học sâu nhằm phân loại các luồng mạng và xây dựng một hệ thống trong mạng nội bộ ứng dụng các mô hình đó một cách trực tiếp. Phương pháp phân loại lưu lượng mạng dựa trên học sâu và tải trọng đã được áp dụng, với ba loại mô hình khác nhau gồm 2DCNN, CNN kết hợp LSTM và Resnet50 đã được xây dựng một cách thủ công và được tinh chỉnh đầu vào cho phù hợp với bài toán phân loại lưu lượng mạng, với mỗi mô hình gồm bốn kích thước tải trọng đầu vào khác nhau. Bốn nhãn lớp được các mô hình phân loại bao gồm Non-RAT (không phải là phần mềm truy cập từ xa), RDP (Remote Desktop Connection), VNC (Virtual Network Computing) và TeamViewer. Hệ thống được đề xuất bao gồm các thiết bị trong mạng nội bộ, gọi là các *client*, và một thiết bị trung tâm để xử lý, gọi là *server*. Client được cài đặt Suricata và Filebeat, có nhiệm vụ ghi nhận và chuyển tiếp tất cả lưu lượng mạng xuất hiện trên thiết bị lên server. Server được cài đặt Logstash, một chương trình phân loại sử dụng ngôn ngữ Python và được import mô hình học sâu, Filebeat, và Kibana, có nhiệm vụ phân loại các luồng mạng trên client và hiển thị kết quả một cách trực quan. Tất cả các công cụ trong hệ thống đều có thể được chạy dưới dạng một service và khởi động cùng với hệ điều hành, nên không ảnh hưởng đến trải nghiệm của người dùng. Kết quả thực nghiệm cho thấy các mô hình đều có kết quả phân loại rất cao trên tập dữ liệu kiểm thử, nhưng lại kém đi khi triển khai trên thực tế với lưu lượng mạng bình thường (không phải lưu lượng phần mềm truy cập từ xa). Nhược điểm này được khắc phục khi kết hợp sử dụng mô hình học sâu với phương pháp sử dụng signature bytes (magic bytes) của các phần mềm truy cập từ xa. Ngoài ra, thời gian phản hồi trực tiếp của hệ thống trước lưu lượng mạng mới cũng là một vấn đề quan trọng được bàn luận trong đồ án. Từ đó, đồ án đưa ra tương quan so sánh mức độ hiệu quả giữa các mô hình với kích thước tải trọng đầu vào khác nhau. Kết quả cho thấy mỗi mô hình có một ưu, nhược điểm riêng và đều cho kết quả tốt khi triển khai thực tế. Bên cạnh đó, đồ án cũng chỉ ra và phân tích những hạn chế hiện có của hệ thống và đề xuất một số giải pháp trong tương lai để giúp hệ thống được hoàn thiện hơn.

Sinh viên thực hiện

(Ký và ghi rõ họ tên)

MỤC LỤC

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

Thuật ngữ	**Ý nghĩa**
CNN	Mạng tích chập (Convolutional neural network)
LSTM	Bộ nhớ dài hạn - ngắn hạn (Long short term memory)
RAT	Phần mềm truy cập từ xa (Remote Access Tool)