CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

BẢN NHẬN XÉT LUẬN VĂN TỐT NGHIỆP THẠC SĨ

(Dùng cho người phản biện)

Tên đề tài luận văn:	*Phát hiện xâm nhập mạng sử dụng học máy.*
Chuyên ngành:	Hệ thống thông tin.
Mã số:	8.48.01.04
Họ và tên học viên cao học:	Dương Đỗ Nhuận.
Họ và tên người nhận xét:	Phùng Văn Ổn.
Học hàm, học vị:	Tiến sĩ.
Cơ quan công tác:	Nguyên cán bộ Trung tâm Tin học VPCP.

NỘI DUNG NHẬN XÉT

I. Cơ sở khoa học và thực tiễn, sự cần thiết lựa chọn đề tài:

Khi môi trường làm việc của các tổ chức, cá nhân đều được đưa lên mạng thì cũng gia tăng các hoạt động tội phạm mạng, đã buộc các tổ chức phải tăng cường năng lực kiểm tra, giám sát nhằm bảo đảm an toàn hệ thống. Các tội phạm mạng sử dụng nhiều phương thức với mức độ ngày càng tinh vi để tấn công các hệ thống CNTT, gây thiệt hại cho các chủ thể của hệ thống. Do vậy, việc bảo đảm an toàn cho các hệ thống CNTT cần thiết phải được tăng cường.

Hiện đã có nhiều kết quả nghiên cứu và giải pháp triển khai bảo đảm an toàn cho các hệ thống CNTT. Các hướng nghiên cứu cũng đa dạng, trong đó các nghiên cứu về giải pháp kỹ thuật phát hiện, phòng chống, ngăn chặn các cuộc tấn công xâm nhập mạng vẫn đang được tiếp tục. Nhiều nghiên cứu đã ứng dụng thành công các kỹ thuật của học máy vào các hệ thống giám sát, phát hiện tấn công mạng. Vì vậy, đề tài "*Phát hiện xâm nhập mạng sử dụng học máy*" là có ý nghĩa khoa học, thực tiễn cao.

II. Về nội dung, chất lượng của luận văn, các kết quả đạt được

2.1. Về nội dung

Luận văn được trình bày trong 67 trang, gồm phần mở đầu, 3 chương nội dung, phần kết luận và tài liệu tham khảo.

Chương 1: Tổng quan về phát hiện xâm nhập mạng. Trong chương này, phần đầu chương tác giả trình bày tổng quan về xâm nhập mạng và một số dạng tấn công xâm nhập mạng điển hình vào các hệ thống CNTT như tấn công tràn bộ đệm, tấn công theo giao thức mạng, tấn công bằng mã độc,.... Phần tiếp theo, tác giả trình bày về các biện pháp phòng chống tấn công mạng như phòng thủ theo chiều sâu, phòng thủ đa dạng,... và cuối cùng là giới thiệu về phát hiện xâm nhập mạng cùng cách phân loại xâm nhập mạng theo nguồn dữ liệu (xâm nhập ở mức mạng, xâm nhập ở mức thiết bị đầu cuối), phân loại xâm nhập theo phương pháp phân tích (dựa trên dấu hiệu, dựa vào bất thường).

Chương 2: Phát hiện xâm nhập dựa trên học sâu. Trong chương này, tác giả trình bày về khái quát về học máy (machine learning), học sâu (deep learning) và ứng dụng Autoencoder trong tiền xử lý dữ liệu của các hệ thống phát hiện xâm nhập mạng. Cuối chương, tác giả trình bày về xây dựng mô hình phát hiện xâm nhập mạng dựa trên học sâu, theo đó mô hình này được triển khai trong 2 giai đoạn: giai đoạn huấn luyện và giai đoạn phát hiện và sử dụng các bộ dữ liệu được gán nhãn như NSL-KDD theo tỷ lệ 70% cho huấn luyện và 30% để kiểm thử.

Chương 3: Cài đặt và thử nghiệm. Trong chương này, tác giả trình bày tiến trình xây dựng, cài đặt hệ thống phát hiện xâm nhập mạng dựa trên mô hình đã nêu ở cuối chương 2, sử dụng ngôn ngữ Python và các thư viện Tensorflow, Sklearn,... Các dữ liệu thử nghiệm là bộ dữ liệu Phisihing Website Data với 30 đặc trưng và 2.456 mẫu dữ liệu, bộ dữ liệu NSL-KDD với 125.973 bản ghi để huấn luyện, 22.544 bản ghi để kiểm tra (huấn luyện để phát hiện các nhóm tấn công xâm nhập như DoS, R2L (chiếm quyền try cập cục bộ), U2R (để chiếm quyền truy cập root), Probe (thu thập thông tin về kiến trúc mạng); xây dựng mạng noron phát hiện xâm nhập mạng dựa trên các phương pháp học sâu SAE (Stacked Autoencoder) và SDAE (Stacked Denoise Autoencoder) với cấu trúc một tầng input, 2 tầng ẩn và một tầng output. Cuối chương, tác giả trình bày các kết quả thử nghiệm mô hình với nhiều kịch bản khác nhau. Kết quả chỉ ra là các thuật toán Naïve Bayes hiệu quả tốt hơn nhiều so với việc không sử dụng mạng Autoencoder, trong khi đó, với các thuật toán: Máy vecto hỗ trợ (SVM), Cây quyết định, Rừng ngẫu nhiên, K-láng giềng thì tính hiệu quả của việc sử dụng mạng Autoencoder không khác nhiều so với không sử dụng.

2.2. Về chất lượng và các kết quả đạt được

Về mặt lý thuyết: Luận văn đã tổng hợp được các kiến thức về tấn công mạng và phát hiện xâm nhập mạng, về kỹ thuật học sâu và ứng dụng vào việc phát hiện tấm công xâm nhập mạng; đề xuất được mô hình phát hiện tấn công mạng dựa vào kỹ thuật học sâu và quá trình xử lý dữ liệu sử dụng phương pháp trích chọn đặc trưng autoencoder và đưa dữ liệu vào huấn luyện, phát hiện tấn công mạng sử dụng một số thuật toán học máy có giám sát như máy vecto hỗ trợ (SVM), Cây quyết định, Rừng ngẫu nhiên, K-láng giềng, Naïve Bayes.

Về thực nghiệm: Tác giả đã cài đặt và thử nghiệm mô hình phát hiện tấn công xâm nhập mạng theo các kịch bản đề xuất.

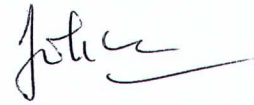
Các kết quả trên có thể làm tài liệu tham khảo cho những người quan tâm đến việc nghiên cứu về an toàn thông tin nói chung và giám sát mạng nói riêng.

III. Kết luận:

Kết quả của luận văn đáp ứng yêu cầu cơ bản của luận văn thạc sĩ. Đồng ý cho phép học viên được bảo vệ Luận văn tốt nghiệp.

Hà Nội, ngày 09 tháng 01 năm 2021

NGƯỜI NHẬN XÉT



TS. Phùng Văn Ổn