TRƯỜNG ĐH CÔNG NGHIỆP TP. HCM
Khoa Công nghệ Thông tin

THUYẾT MINH ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CẤP TRƯỜNG
(dành cho sinh viên)

I. THÔNG TIN ĐỀ TÀI

1. Tên đề tài: Sử dụng Machine Learning và Deep Learning trong an ninh mạng

2. Ngành khoa học: *(đánh dấu ✓ vào mục phù hợp)*

☐ Hóa – Sinh – Thực phẩm – Môi trường ☑ Công nghệ Thông tin – Điện – Điện tử
☐ Khoa học tự nhiên ☐ Khoa học xã hội
☐ Mỹ thuật ☐ Kinh tế
☐ Cơ khí ☐ Xây dựng

3. Thời gian thực hiện: 12 tháng

4. Kinh phí dự kiến: 10,638,600 (*số tiền bằng chữ:* Mười triệu sáu trăm ba mươi tám nghìn sáu trăm đồng)

5. Chương trình đăng ký: ☑ Tự đề xuất ☐ Đặt hàng nghiên cứu
(Trường hợp là đề tài đặt hàng cần có văn bản xác nhận hoặc đơn đặt hàng đính kèm)

6. Họ tên cá nhân đăng ký thực hiện: **Trịnh Thị Bảo Bảo**

7. Điện thoại di động: 0813309798
Thư điện tử (e-mail): ttbb08022001@gmail.com

8. Khóa học: **2019-2023**

9. Cố vấn học tập: **Nguyễn Chí Kiên**

10. Điện thoại di động: **0853848713**

11. Nguồn nhân lực thực hiện đề tài

TT	Họ và tên	Vai trò	Học vị chức danh	Đơn vị công tác
1	Trịnh Thị Bảo Bảo	Chủ nhiệm đề tài	Đại học	Trường Đại học Công Nghiệp Tp.Hồ Chí Minh
2	Nguyễn Thị Thanh Hòa	Thành viên chính	Đại học	Trường Đại học Công Nghiệp Tp.Hồ Chí Minh
3	Nguyễn Văn Anh Tuấn	Thành viên chính	Đại học	Trường Đại học Công Nghiệp Tp.Hồ Chí Minh
4	Trần Nam Bá	Thành viên chính	Đại học	Trường Đại học Công Nghiệp Tp.Hồ Chí Minh
5	Phan Lê Hoàng Việt	Thành viên chính	Đại học	Trường Đại học Công Nghiệp Tp.Hồ Chí Minh
6	Đinh Hoàng Hải Đăng	Thành viên chính	Đại học	Trường Đại học Công Nghiệp Tp.Hồ Chí Minh

II. TỔNG QUAN ĐỀ TÀI

1. Giới thiệu tóm tắt

Các hệ thống máy tính nối mạng Internet ngày nay đang phải đối diện với tấn công mạng (cyber attacks) ngày càng tinh vi và có qui mô lớn. Vì vậy các cơ chế phòng vệ cũng cần phải phản ứng nhanh, linh hoạt, và dễ dàng triển khai trên qui mô lớn. Từ trước đến nay, Machine Learning (Học Máy) đã được áp dụng rộng rãi để xây dựng các chiến lược tấn công và phòng thủ trong an ninh mạng. Phía tấn công sử dụng Machine Learning để vô

hiệu hoá các chiến lược phòng thủ. Phía phòng thủ sử dụng Machine Learning để ngăn ngừa tấn công và giảm thiểu các rủi ro và tác động đến hệ thống.

Trong dự án nghiên cứu này, chúng tôi kết hợp cả học máy không giám sát (unsupervised learning), học máy có giám sát (supervised learning), và học máy tăng cường (reinforcement learning) để thiết kế và xây dựng các mô hình an ninh mạng. Chúng tôi sẽ sử dụng unsupervised learning để cho việc phát hiện bất thường (anomaly detection) và phát hiện đột nhập (intrusion detection). Các cơ chế deep learning sẽ được sử dụng để tạo features cho mô hình Machine Learning. Hệ thống còn sử dụng supervised learning để học từ các tấn công và hành vi của phía tấn công trong quá khứ. Cuối cùng, reinforcement learning sẽ được sử dụng để hệ thống đưa ra chiến lược phòng thủ tối ưu sau các giai đoạn phòng thủ.

Chúng tôi sẽ sử dụng các tập dữ liệu an ninh mạng (ví dụ TCP dump) để kiểm thử và đánh giá các mô hình học máy.

Giới thiệu tóm tắt về đề tài (nêu vấn đề, mục tiêu và phương pháp nghiên cứu, không quá 300 từ)

2. Tổng quan tình hình nghiên cứu và tính cấp thiết tiến hành nghiên cứu

a) Tình hình nghiên cứu quốc tế

1. An ninh mạng là tập hợp các công nghệ và quy trình được thiết kế để bảo vệ hệ thống điện tử, mạng lưới, máy tính, chương trình và dữ liệu khỏi những cuộc tấn công truy cập hoặc thay đổi hay phá hủy có chủ đích. Hiện nay có các phương thức tấn công an ninh mạng như: U2R, R2L, DoS, Wormhole, … Những nghiên cứu trước đây đa phần tập trung về việc phát triển hệ thống phát hiện xâm nhập mạng (Intrusion Dectection System – IDS), ví dụ như Revathi và Malathi [1] chỉ tập trung vào nghiên cứu phát hiện lạm dụng, còn Sahoo và cộng sự [2] trình bày công thức về việc phát hiện URL độc hại sử dụng học máy, nhưng chưa trình bày chi tiết về kĩ thuật, … .

b) Tình hình nghiên cứu trong nước

Tình hình nghiên cứu trong nước

c) Đánh giá kết quả các công trình nghiên cứu đã công bố *(ưu, khuyết, những tồn tại…)*

Những nghiên cứu trước đây đa phần tập trung về việc phát triển hệ thống phát hiện xâm nhập mạng (Intrusion Dectection System – IDS), ví dụ như Revathi và Malathi [1] chỉ tập trung vào nghiên cứu phát hiện lạm dụng, còn Sahoo và cộng sự [2] trình bày công thức về việc phát hiện URL độc hại sử dụng học máy, nhưng chưa trình bày chi tiết về kĩ thuật, … . Các thuật toán như K-NN, Decision Tree, SVM, Bayes, … để phát hiện xâm nhập bất thường.

d) Tính cấp thiết tiến hành nghiên cứu *(tính mới, tính khoa học)*

Trong nghiên cứu này, chúng tôi sử dụng *unsupervised learning* để phát hiện xâm nhập, làm lõi cho IDS. Sử dụng *deep learning* để chọn các features cần thiết đưa vào mô hình

machine learning để học dựa trên các tấn công và hành vi tấn công trong quá khứ. Để cải thiện mô hình, *reinforce learning* sẽ tối ưu sau giai đoạn phòng thủ bằng cách đưa ra các chiến lược.

3. Mục tiêu của đề tài

a) Mục tiêu tổng quát.

Trong dự án nghiên cứu này, chúng tôi kết hợp cả học máy không giám sát (unsupervised learning), học máy có giám sát (supervised learning), và học máy tăng cường (reinforcement learning) để thiết kế và xây dựng các mô hình an ninh mạng.

b) Mục tiêu cụ thể.

Các mô hình Machine Learning sẽ được áp dụng cho các tập dữ liệu an ninh mạng phổ biến. Mục tiêu của chúng tôi là xây dựng các mô hình an ninh mạng mới có hiệu năng cao, có thể áp dụng trong thực tế. Chúng tôi sẽ hoàn thành bài báo để xuất bản trên tạp chí của IUH.

4. Nội dung thực hiện, phương pháp nghiên cứu

(Luận cứ rõ cách tiếp cận vấn đề nghiên cứu, thiết kế nghiên cứu, phương pháp nghiên cứu, kỹ thuật sẽ sử dụng gắn với từng nội dung chính của đề tài; so sánh với các phương pháp giải quyết tương tự khác và phân tích để làm rõ được tính mới, tính sáng tạo của đề tài)

4. 1 Nội dung 1: Nghiên cứu về an ninh mạng

- *Cách tiếp cận:* Từ những tài liệu liên quan mà chúng tôi thu thập được.

- *Phương pháp nghiên cứu, kỹ thuật sử dụng:* Cơ sở lý thuyết và thực nghiệm

- *Kết quả:* Chương trình và báo cáo

4. 2 Nội dung 2: Nghiên cứu về thuật toán

- *Cách tiếp cận:* Từ các nguồn tài liệu

- *Phương pháp nghiên cứu, kỹ thuật sử dụng:* Cơ sở lý thuyết và thực nghiệm

- *Kết quả:* Chương trình và báo cáo

4. 3 Nội dung 3: Đưa ra chiến lược phòng thủ tối ưu sau các giai đoạn phòng thủ

- *Cách tiếp cận:* Từ các nguồn tài liệu

- *Phương pháp nghiên cứu, kỹ thuật sử dụng:* Cơ sở lý thuyết và thực nghiệm

- *Kết quả:* Chương trình và báo cáo

4. 4 Nội dung 4: Viết báo cáo tổng kết

- *Cách tiếp cận:* Tổng hợp các bản báo cáo từ các nội dung trên

- *Phương pháp nghiên cứu, kỹ thuật sử dụng:* Viết báo cáo

- *Kết quả:* Báo cáo tổng kết

5. Kế hoạch triển khai

Nội dung	Công việc thực hiện	Kết quả phải đạt	Thời gian (bắt đầu, kết thúc)	Cá nhân chủ trì
1	Nghiên cứu về an ninh mạng	Báo cáo	1-2/2021	Tất cả thành viên
2	Nghiên cứu về thuật toán (supervised learning, unsupervised learning, reinforcement learning, deep learning)	Báo cáo	1-4 tháng	Tất cả thành viên
3	Đưa ra chiến lược phòng thủ tối ưu sau các giai đoạn phòng thủ	Báo cáo	5-12 tháng	Tất cả thành viên
4	Viết báo cáo	Báo cáo tổng kết	12/2021	Tất cả thành viên

6. Sản phẩm

Dạng III: Bài báo;

TT	**Tên bài báo dự kiến**	**Nơi công bố**(IUH, ISI, SCOPUS)	**Ghi chú**
1	Sử dụng Machine Learning và Deep Learning trong an ninh mạng.	IUH	

***Lưu ý:** bài báo là sản phẩm đề tài NCKH không được đăng ký xét thưởng quy định tại Khoản 3 Điều 11 Quy chế chi tiêu nội bộ*

7. Khả năng ứng dụng và phương thức chuyển giao kết quả nghiên cứu *(dự kiến)*

Mục tiêu của chúng tôi là xây dựng các mô hình an ninh mạng mới có hiệu năng cao, có thể áp dụng trong thực tế.

III. PHÂN BỔ KINH PHÍ THỰC HIỆN (đơn vị: đồng)

1. Cơ cấu phân bổ kinh phí

Tổng kinh phí	Khoản chi		
	Công lao động	Nguyên vật liệu, thiết bị, máy móc	Chi khác
10,638,600	**10,638,600**	**0**	**0**

2. Phụ lục giải trình các khoản chi

Công lao động (khoa học, phổ thông)

TT	Người thực hiện	Chức danh	Nội dung công việc	Tổng số ngày công	Hệ số tiền công/b>	Đơn giá	Thành tiền
1	Trịnh Thị Bảo Bảo	Chủ nhiệm nhiệm vụ	Lao động khoa học	4	0.71	1,490,000	4,231,600
2	Nguyễn Văn Anh Tuấn	Thành viên chính	Lao động khoa học	2	0.43	1,490,000	1,281,400
3	Nguyễn Thị Thanh Hòa	Thành viên chính	Lao động khoa học	2	0.43	1,490,000	1,281,400
4	Trần Nam Bá	Thành viên chính	Lao động khoa học	2	0.43	1,490,000	1,281,400
5	Đinh Hoàng Hải Đăng	Thành viên chính	Lao động khoa học	2	0.43	1,490,000	1,281,400
6	Phan Lê Hoàng Việt	Thành viên chính	Lao động khoa học	2	0.43	1,490,000	1,281,400
Cộng							**10,638,600**

CHỦ NHIỆM ĐỀ TÀI **ĐƠN VỊ CHỦ QUẢN** **TIỂU BAN KHOA HỌC**

Trịnh Thị Bảo Bảo **Huỳnh Trung Hiếu**