HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Trần Hoàng Anh

NGHIÊN CỨU PHƯƠNG PHÁP LAI TRONG PHÁT HIỆN MÃ ĐỘC BOTNET TRÊN THIẾT BỊ IOT

Chuyên ngành: Hệ thống thông tin
Mã số: 8.48.01.04

TÓM TẮT LUẬN VĂN THẠC SỸ

(Theo định hướng ứng dụng)

Hà Nội - 2021

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: **TS. Ngô Quốc Dũng**

Phản biện 1: **PGS. TS. Hoàng Hữu Hạnh**

Phản biện 2: **PGS. TS. Nguyễn Linh Giang**

Luận văn này được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: Ngày 28/8/2021

Có thể tìm hiểu luận văn này tại:

Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

Ngày nay, cuộc cách mạng công nghiệp 4.0 đang làm biến đổi nhanh chóng nền công nghiệp ở mọi quốc gia và diễn ra trên phạm vi toàn cầu. Với nhiều thành phần khác nhau, đặc điểm nổi bật nhất của cuộc cách mạng công nghiệp lần thứ 4 đó là việc dịch chuyển các hệ thống máy móc sản xuất truyền thống sang các hệ thống tự động hoá có khả năng tự hành một cách thông minh (Cyber Physical Systems - CPSs), xóa nhòa biên giới giữa thực và ảo.

Hiện nay, cách mạng công nghiệp 4.0 vẫn đang ở giai đoạn sơ khai. Xu hướng số hóa hay công cuộc chuyển đổi số đang xuất hiện ở mọi lĩnh vực và ngành kinh tế. Trong những bước chuyển đổi đó, công nghệ Vạn vật kết nối (IoT – Internet of Things) phát triển một cách mạnh mẽ và là một trong những thành phần không thể thiếu của mọi hệ thống hiện nay. Các thiết bị IoT được ứng dụng không chỉ trong trong công nghiệp mà còn góp phần cải thiện đời sống của con người.

Phần lớn các thiết bị IoT, vì mục đích triển khai một cách dễ dàng trên diện rộng, đã được sản xuất hàng loạt với khả năng bảo mật khá lỏng lẻo. Trong một thời đại mà các thiết bị IoT vốn phân tán khắp nơi trên thế giới, sự kém bảo mật của những thiết bị này đã trở thành đích ngắm của vô vàn các cuộc tấn công bằng mã độc. Trong số những cuộc tấn công đó, loại tấn công lây lan với tốc độ chóc mặt nhất và nguy hiểm nhất là các cuộc tấn công sử dụng mã độc IoT Botnet. Hiện nay, các loại mã độc IoT Botnet phức tạp như Mirai đang dần xuất hiện với một tỷ lệ đáng báo động. Tuy nhiên, do đặc điểm hạn chế về các tài nguyên trong hệ thống, việc phát hiện mã độc IoT Botnet đang ngày càng đa dạng và đổi mới là cực kỳ khó khăn. Số lượng các nghiên cứu về phát hiện IoT Botnet hiện nay vẫn còn khá ít. Vì vậy, đây có thể coi là một hướng nghiên cứu ứng dụng khá cấp thiết trong tình hình hiện nay.

Với những lý do đã đề cập ở trên, tôi xin lựa chọn đề tài “Nghiên cứu phương pháp lai trong phát hiện mã độc Botnet trên thiết bị IoT”.

Chương 1: MÃ ĐỘC IOT BOTNET VÀ CÁC HƯỚNG PHÁT HIỆN

1.1. Tổng quan về mã độc IoT và IoT Botnet

1.1.1. Khái niệm mã độc IoT

Mã độc IoT là loại mã độc được tạo ra để nhắm tới các thiết bị IoT, lợi dụng các điểm yếu sẵn có trong lỗi kiến trúc và mã mềm kém chất lượng của thiết bị để thực hiện các hành động như đánh cắp thông tin cá nhân của người sử dụng, xây dựng mạng botnet hoặc thậm chí phá hủy cả hạ tầng mạng. Các hướng tấn công của mã độc IoT rất đa dạng, tuy nhiên chúng chủ yếu nhắm vào các tầng vật lý và tầng mạng của thiết bị IoT.

1.1.2. Phân loại mã độc IoT

Dựa trên những đặc điểm của thiết bị IoT và các tính chất đặc trưng của mã độc, có thể phân loại độc IoT thành 5 loại: Worm, Trojan, Rootkit, Spyware, và Botnet.

1.1.3. Mã độc IoT Botnet và nguy cơ tấn công từ chối dịch vụ

IoT Botnet là một mạng lưới được tạo nên bởi các thiết bị đã lây nhiễm IoT bot. Về cơ bản, chức năng của IoT bot cũng tương tự như một bot truyền thống. Mạng lưới IoT Botnet sẽ thường được điều khiển bởi một botmaster, cho phép thực hiện các hoạt động tấn công với sự hỗ trợ của các bot. Cách gọi mã độc IoT Botnet được đề cập trong luận văn là cách gọi chung cho toàn bộ các loại mã độc IoT được botmaster cài cắm vào các thiết bị IoT nạn nhân và điều khiển chúng thực hiện các hành động như tấn công từ chối dịch vụ...

Hầu hết mã độc hiện nay được sinh ra thông qua việc sao chép mã nguồn hoặc biến thể của cùng mã độc gốc.

1.2. Cấu trúc và nguyên lý hoạt động của mã độc IoT Botnet

1.2.1. Cấu trúc của mạng mã độc IoT Botnet

Mã độc IoT Botnet sử dụng cấu trúc mạng botnet tương tự với các mạng botnet thông thường, có thể chia ra thành hai dạng là mạng botnet tập trung và mạng botnet ngang hàng.

Trong cấu trúc mạng botnet tập trung, các máy chủ C&C sẽ có địa chỉ cố định. Giao thức thường được sử dụng trong trường hợp này là IRC và HTTP. Với cấu trúc mạng botnet ngang hàng, giao thức được sử dụng chủ yếu là giao thức P2P. Trong trường hợp này, việc xác định địa chỉ của máy chủ C&C sẽ khó khăn hơn.

Ngoài hai thành phần cơ bản là bot và máy chủ C&C, các mạng mã độc IoT Botnet thường có thêm bốn thành phần phụ bao gồm: máy Scanner, máy chủ Report, máy Loader, máy chủ phân phối mã độc. Các chức năng của mỗi thành phần này có thể được tổ hợp và thực hiện bởi một thành phần khác.

1.2.2. Nguyên lý hoạt động của mã độc IoT Botnet

Nguyên lý hoạt động chủ yếu của các mã độc IoT Botnet là thực hiện tấn công từ chối dịch vụ thông qua quy trình: (1) Thực hiện tấn công chiếm quyền điều khiển thiết bị; (2) Gửi các thông tin của thiết bị IoT mới về máy chủ Report; (3) Botmaster giao tiếp với máy chủ Report thông qua máy chủ C&C; (4) Botmaster gửi lệnh lây nhiễm tới máy Loader; (5) Máy loader truy cập vào thiết bị IoT, chỉ dẫn thiết bị tải và thực thi mã độc; (6) Botmaster chỉ thị các bot tấn công; (7) Các thiết bị bot tấn công máy chủ mục tiêu.

1.3. Các phương pháp phát hiện mã độc IoT Botnet

1.3.1. Phát hiện mã độc IoT Botnet dựa trên phân tích tĩnh

Phương pháp phân tích tĩnh là phương pháp phân tích, phát hiện mã độc, lỗ hổng bảo mật dựa trên những đặc trưng của các tập tin chương trình mà không cần thực thi chúng (trên thiết bị thực hoặc môi trường mô phỏng). Việc phân tích như vậy có thể thực hiện trên mã nguồn tường minh hoặc các tập tin nhị phân thực thi.

1.3.2. Phát hiện mã độc IoT Botnet dựa trên phân tích động

Phân tích động là phương pháp phân tích cách hoạt động của mã độc khi mã độc được thực thi. Bằng cách giám sát các hoạt động của mã độc, cách thức thực thi lây lan như thế nào, nó kết nối đến đâu, cài đặt những gì vào hệ thống, thay đổi thành phần nào nhằm mục đích ngăn chặn việc lây nhiễm, tạo ra các dấu hiệu nhận dạng hiệu quả.

1.3.3. Phát hiện mã độc IoT Botnet dựa trên phương pháp lai

Trong mỗi phương pháp phân tích tĩnh và phân tích động đều có những ưu điểm và hạn chế nhất định. Hướng phát hiện dựa trên phân tích tĩnh sẽ có lợi thế hơn trong việc hiểu rõ cấu trúc của mã độc. Trong khi đó, hướng phân tích động lại có thể giải quyết các kỹ thuật làm rối mã độc. Từ đó, sản sinh ra một hướng tiếp cận nữa, với mục tiêu vận dụng được ưu điểm của cả hai phương pháp trên, đó chính là phương pháp lai. Đây cũng là hướng nghiên cứu chính của luận văn.

Có một vài cách thức thực hiện phương pháp lai khác nhau, tuy nhiên để đảm bảo tính khách quan trong hai quy trình phân tích tĩnh và động, luận văn sẽ lựa chọn cách thực hiện cả hai quy trình phân tích tĩnh và động song song, sau đó dựa trên kết quả của

hai quy trình là tập các đặc trưng tĩnh và động để tích hợp chúng lại và sử dụng thuật toán học máy để huấn luyện và phân loại mã độc.

Kết luận chương 1

Nội dung chương 1 đã trình bày một cách khái quát về mã độc IoT nói chung và IoT Botnet nói riêng. Cùng với đó là tổng quan về các hướng nghiên cứu đã được vận dụng để phân tích, phát hiện mã độc IoT Botnet.

Trong chương 1, luận văn đã đưa khái niệm, phân loại của mã độc IoT nói chung và cụ thể về IoT Botnet nói riêng. Từ đó, làm rõ cấu trúc và các nguyên lý hoạt động, lây lan của mã độc IoT Botnet. Bên cạnh đó, luận văn cũng tìm hiểu về một số hướng nghiên cứu và các phương pháp đã được vận dụng để phân tích, phát hiện mã độc IoT Botnet cùng các ưu điểm, nhược điểm của từng phương pháp phân tích động, phân tích và phân tích lai.

Kết quả nghiên cứu của chương 1 sẽ là cơ sở để luận văn lựa chọn xây dựng thử nghiệm phương pháp lai trong phát hiện mã độc IoT Botnet ở chương 2.

Chương 2: PHƯƠNG PHÁP LAI
TRONG PHÁT HIỆN MÃ ĐỘC IOT BOTNET

2.1. Xây dựng các đặc trưng tĩnh

2.1.1. Một số đặc trưng tĩnh trong phát hiện mã độc IoT Botnet

Trong phân tích tĩnh, tùy thuộc vào việc trích chọn và xử lý các đặc trưng sẽ ảnh hưởng đến độ chính xác và độ phức tạp của phương pháp phát hiện mã độc IoT Botnet. Dựa trên các nghiên cứu về phân tích tĩnh trong phát hiện mã độc IoT, có một số đặc trưng tĩnh có thể được sử dụng trong phát hiện mã độc IoT Botnet, bao gồm: mã thực thi (Opcode), các chuỗi (String), tiêu đề tập tin ELF (Executable and Linkable format), ảnh đa mức xám, đồ thị hàm gọi (FCG), đồ thị thông tin chuỗi in (PSI).

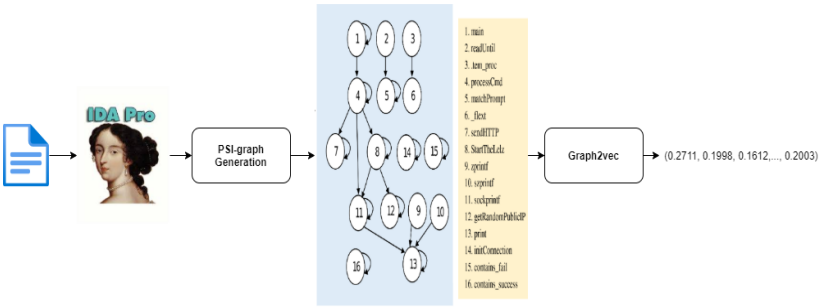
2.1.2. Đặc trưng tĩnh được chọn cho phương pháp lai

Dựa trên nguyên lý hoạt động của mã độc IoT Botnet, có thể thấy, cho dù các mã độc IoT Botnet ngày càng tiến hóa về cả số lượng biến thể lẫn chủng loại thì hoạt động của chúng vẫn tuân theo một quy luật nhất định. Đây có thể coi là đặc trưng của riêng mã độc IoT Botnet và là một hướng khá hiệu quả trong phát hiện mã độc IoT Botnet. Đồ thị PSI được xây dựng thể hiện cho một chu trình hoạt động tiêu biểu của mã độc IoT Botnet

Vì lý do đó, luận văn sẽ lựa chọn việc trích xuất các đặc trưng từ đồ thị này để làm các đặc trưng tĩnh phục vụ cho phương pháp lai

2.1.3. Xây dựng tập đặc trưng tĩnh

Sau khi đã xác định lựa chọn việc trích xuất các đặc trưng tĩnh từ đồ thị thông tin chuỗi in PSI, việc xây dựng tập đặc trưng tĩnh cho phương pháp lai sẽ được thực hiện thông qua các bước được thể hiện trong Hình 2.1.



Hình 2.1: Quy trình xây dựng tập đặc trưng tĩnh

Kết quả của quy trình này là tập các one-hot vector với chiều dài tùy ý tượng trưng cho tập các đồ thị PSI của các mẫu tệp thực thi ELF.

2.2. Xây dựng các đặc trưng động

2.2.1. Một số đặc trưng động trong phát hiện mã độc IoT Botnet

Khác với phân tích tĩnh, tính hiệu quả của phân tích động chủ yếu xoay quanh việc thực thi mã độc trong môi trường thời gian thực và giám sát hành vi của chúng. Một số loại đặc trưng động chủ yếu thường được thu thập bao gồm: các lời gọi hệ thống (System call), thông tin mở rộng của các lời gọi hệ thống (EXINFO), lưu lượng truy cập mạng và các thông tin về hiệu năng máy chủ.

2.2.2. Lựa chọn môi trường giám sát thời gian thực

Lựa chọn môi trường thời gian thực để giám sát và thu thập các hành vi đặc trưng của mã độc cũng là một yếu tố hết sức quan trọng trong quy trình phân tích động. Môi trường máy ảo hoặc sandbox được sử dụng để mô phỏng giám sát cần phải đáp ứng được một số tiêu chí nhất định.

Luận văn đã lựa chọn công cụ V-sanbox để mô phỏng. Điểm mạnh của V-sandbox nằm ở khả năng kích hoạt bổ sung các hành vi độc hại của mã độc IoT, nên dữ liệu của các lời gọi hệ thống sẽ bao gồm nhiều thông tin hơn như các lời gọi kết nối, gửi, nhận thông tin,... Dựa trên những thông tin này, luận văn sẽ tập trung khai thác dữ liệu của các lời gọi hệ thống để xây dựng tập đặc trưng động.

2.2.3. Xây dựng tập đặc trưng động

Như vậy, quy trình xây dựng tập đặc trưng động sẽ được thực hiện thông qua ba bước như được thể hiện trong Hình 2.4.

Hình 2.2: Quy trình xây dựng tập đặc trưng động

Kết quả của quy trình này cũng là một tập các one-hot vector với chiều dài tùy ý tượng trưng cho tập các đồ thị SCG của từng mẫu tệp thực thi ELF.

2.3. Phương pháp tích hợp đặc trưng

2.3.1. Lựa chọn phương pháp tích hợp đặc trưng tĩnh và động

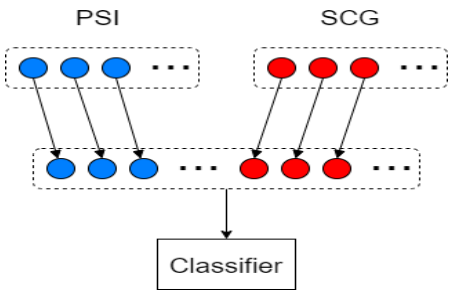
Có rất nhiều hướng tiếp cận cho phương pháp lai trong phát hiện mã độc dựa trên việc tích hợp rất nhiều các đặc trưng được trích xuất từ đặc điểm của tệp tin. Tuy nhiên, việc kết nhiều mức độ nhiễu khác nhau và sự xung đột dữ liệu khiến cho độ chính xác của phương pháp giảm xuống là điều hoàn toàn có thể xảy ra. Vì lý do đó, luận văn đã

lựa chọn việc giới hạn chỉ kết hợp hai đặc trưng của tập tin thực thi ELF là đồ thị thông tin chuỗi in PSI và đồ thị lời gọi hệ thống SCG.

Trong các hướng tích hợp vector đặc trưng, có ba hướng tiếp cận, bao gồm: Early fusion, Late fusion và Intermediate fusion. Luận văn đã lựa chọn sử dụng phương pháp Early fusion để kết hợp hai tập vector đặc trưng tĩnh và động để đảm bảo sự đơn giản về dữ liệu, tránh nhiễu và tối ưu thời gian tính toán.

2.3.1. Xây dựng tập các đặc trưng lai giữa tĩnh và động

Tập các đặc trưng lai giữa tĩnh và động được xây dựng sử dụng phương pháp early fusion để kết hợp tập vector đặc trưng tĩnh trích xuất từ đồ thị thông tin chuỗi in PSI và tập vector đặc trưng động trích xuất từ đồ thị lời gọi hệ thống SCG. Quy trình thực hiện được mô tả như trong Hình 2.5



Hình 2.3: Xây dựng tập vector đặc trưng lai giữa tĩnh và động

Vector đặc trưng thu được sau quá trình kết hợp sẽ có dạng: $v_{fused} = [v_{psi}, v_{scg}]$

2.4. Các thuật toán phân loại mã độc

Hiện nay, có rất nhiều nghiên cứu sử dụng các kỹ thuật học máy trong phân loại mã độc. Trong đó, các thuật toán như thuật toán Cây quyết định, thuật toán k-láng giềng gần nhất, Support Vector Machines và Random Forest thường sẽ cho ra độ chính xác cao. Nguyên nhân bởi bài toán phát hiện mã độc là bài toán phân loại hai lớp. Do đó, các thuật toán với nguyên tắc phân chia các ví dụ có nhãn khác nhau thường sẽ có hiệu quả tốt hơn. Vì vậy, luận văn đã lựa chọn sử dụng bốn thuật toán này cho bài toán phân loại mã độc với tập vector đặc trưng lai giữa tĩnh và động.

2.4.1. Cây quyết định (DT)

Cây quyết định nhận đầu vào là một bộ giá trị đặc trưng mô tả một đối tượng hay một tình huống và trả về một giá trị rời rạc. Để xác định nhãn phân loại cho một mẫu nào đó, ta cho mẫu chuyển động từ gốc cây về phía nút lá. Tại mỗi nút, đặc trưng tương ứng với nút được kiểm tra, tùy theo giá trị của thuộc tính đó mà mẫu được chuyển xuống

nhánh tương ứng bên dưới. Quá trình này lặp lại cho đến khi mẫu tới được nút lá và được nhận nhãn phân loại là nhãn của nút lá tương ứng.

2.4.2. K-láng giềng gần nhất (k-NN)

Các bộ phân loại k-láng giềng gần nhất dựa trên việc học bằng sự giống nhau. Khi có một mẫu chưa biết cho trước thì bộ phân loại k-láng giềng gần sẽ tìm kiếm trong không gian mẫu k mẫu huấn luyện gần mẫu chưa biết đó nhất. Mẫu chưa biết được phân vào lớp phổ biến nhất trong số k láng giềng gần nhất của nó.

2.4.3. Support Vector Machines (SVM)

Nguyên tắc của SVM là tìm một siêu phẳng phù hợp với lề cực đại, tức là khoảng cách từ siêu phẳng tới những ví dụ có nhãn khác nhau là lớn nhất. SVM là một thuật toán khá mạnh và thường được dùng trong các bài toán phân loại. Tuy nhiên, thuật toán này yêu cầu khả năng tính toán khá cao, khá nhạy với các dữ liệu nhiễu và thường gặp phải vấn đề quá vừa dữ liệu.

2.4.4. Random Forest (RF)

Random Forest (RF) là thuật toán phân loại bằng cách xây dựng nhiều cây quyết định từ tập dữ liệu. Thuật toán Ramdom Forest có thể xử lý một lượng lớn các đặc trưng trong tập dữ liệu. Ngoài ra, trong quá trình xây dựng các cây, chúng cũng tạo ra sự ước lượng không mang tính khuynh hướng của các lỗi tổng quát. Thêm vào đó, thuật toán còn có thể ước tính khá tốt những dữ liệu bị thiếu.

Kết luận chương 2

Nội dung chương 2 của luận văn đã trình bày kết quả của quá trình nghiên cứu, tìm hiểu hướng tiếp cận dựa trên phương pháp lai trong phát hiện mã độc IoT Botnet. Hướng tiếp cận của phương pháp lai mà luận văn đã chọn là tiến hành phân tích tĩnh và động song song, sau đó kết hợp các đặc trưng thu được thành một tập đặc trưng lai dùng cho phát hiện mã độc IoT Botnet.

Để tích hợp tập các vector đặc trưng tĩnh và động thành tập các vector đặc trưng lai, luận văn đã sử dụng phương pháp tích hợp đặc trưng Early fusion. Phương pháp kết hợp đặc trưng này sẽ đảm bảo sự đơn giản về dữ liệu, tránh nhiễu và tối ưu thời gian tính toán đối với tập vector đặc trưng lai.

Mô hình phương pháp lai thử nghiệm được xây dựng trong chương 2 sẽ là cơ sở để tiến hành thực nghiệm và đánh giá hiệu quả của phương pháp này được tiến hành trong chương 3.

Chương 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ

3.1. Xây dựng tập dữ liệu

3.1.1. Phương pháp thu thập các mẫu mã độc IoT Botnet và lành tính

Tập dữ liệu được thu thập từ các nguồn như sau:

- Đối với các mẫu mã độc IoT: có hai nguồn thu thập chính là từ nhóm dự án IoTPOT và từ VirusShare.

- Đối với các mẫu lành tính: trích xuất từ các thiết bị IoT SOHO, tải về từ trang web của các hãng sản xuất, thu thập của một số hãng khác từ OpenWRT.

Sau khi thu thập, các tệp tin sẽ được kiểm tra để đảm bảo độ chính xác của các nhãn mã độc và lành tính. Các mẫu bị trùng lặp, hư hại hoặc không phải tập tin ELF sẽ bị loại bỏ để đảm bảo tính khách quan. Các mẫu không thể dịch ngược hoặc không thể thực thi trong sandbox cũng sẽ bị loại bỏ để đảm bảo tính thống nhất của tập dữ liệu.

3.1.2. Mô tả tập dữ liệu

Như vậy, sau khi thực hiện các phương pháp trên, tập dữ liệu thu được bao gồm 6520 mẫu, trong đó có 4644 mẫu mã độc IoT Botnet và 1876 mẫu tập tin IoT lành tính.

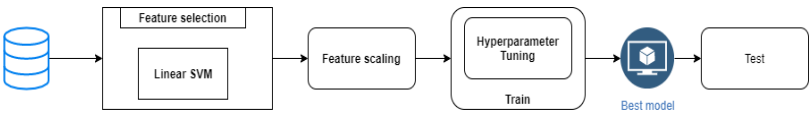
Tập dữ liệu sẽ được chia ra thành hai tập con phục vụ cho việc huấn luyện và kiểm tra với tỷ lệ phân chia lần lượt là 70% và 30%.

3.2. Phương pháp đánh giá và các độ đo sử dụng

3.2.1. Phương pháp đánh giá

Bên cạnh việc giá độ chính xác của phương pháp lai trong phát hiện mã độc, để so sánh sự nổi trội của phương pháp lai, luận văn cũng sẽ thực hiện các thực nghiệm tương tự với hai tập vector đặc trưng tĩnh vã động riêng biệt.

Để xử lý vấn đề thiếu cân bằng giữa các mẫu mã độc và lành tính trong tập dữ liệu, các bộ phân loại sẽ được đánh trọng số cho nhãn. Quá trình thực nghiệm được thể hiện trong Hình 3.1.



Hình 3.1: Thực nghiệm phân loại mã độc

Tập vector đặc trưng lai sẽ được đưa vào trích chọn đặc trưng để giảm bớt kích thước dữ liệu, sau đó được chuẩn hóa để tập dữ liệu có giá trị kỳ vọng bằng không và độ lệch chuẩn bằng 1. Cuối cùng tập dữ liệu chuẩn hóa sẽ được đưa vào bộ phân loại để huấn luyện và chọn ra mô hình phân loại tốt nhất để kiểm tra.

3.2.2. Các độ đo sử dụng để đánh giá

Tương tự như các nghiên cứu về phân loại mã độc, luận văn cũng sử dụng các độ đo như Accuracy, TPR, FPR, Precision, F1 để đánh giá hiệu quả phân loại.

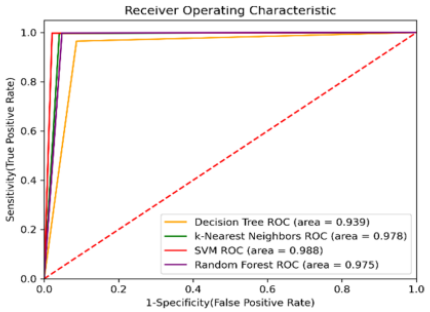
3.3. Kết quả thực nghiệm

Kết quả thực nghiệm của phương pháp lai được thể hiện trong Bảng 3.2.

Bảng 3.1: Kết quả thực nghiệm với phương pháp lai

(%)	DT	k-NN	SVM	RF
ACC	94.99	98.57	99.18	98.52
ROC AUC	93.89	97.78	98.79	97.53
FPR	8.70	4.09	2.13	4.80
Precision	96.48	98.37	99.14	98.10
TPR	96.48	99.64	99.71	99.86
F1	96.48	99.00	99.43	98.97

Mối quan hệ giữa tỷ lệ dương đúng và dương sai được biểu diễn thông qua đường cong ROC tương ứng với từng bộ phân loại như trong Hình 3.2



Hình 3.2: Đường cong ROC của các bộ phân loại

Kết quả so sánh của phương pháp lai với hai phương pháp phân tích tĩnh và động được thể hiện trong Bảng 3.3.

Bảng 3.2: So sánh phương pháp phân tích động, phân tích tĩnh và phương pháp lai

(%)	PSI	SCG	Lai
ACC	96.83	98.41	99.18
ROC AUC	96.66	97.75	98.79
FPR	3.73	3.71	2.13
Precision	98.47	98.61	99.14
TPR	97.06	99.2	99.71

F1	97.76	98.91	99.43

3.1. Đánh giá và so sánh

Theo Bảng 3.1, cả bốn bộ phân loại đều cho tỷ lệ chính xác cao với tỷ lệ dương sai thấp. Trong đó, mô hình cho ra kết quả cao nhất là Support Vector Machines. Như vậy, phương pháp nghiên cứu có thể phát hiện mã độc IoT Botnet với độ chính xác cao.

Hình 3.2 cho thấy bộ phân loại không bị bias do số lượng lớn các mẫu mã độc trong tập dữ liệu khả năng phân loại hai nhãn mã độc và lành tính khá tốt.

Đối với thực nghiệm riêng lẻ với giữa hai phương pháp phân tích tĩnh và động, kết quả cũng cho thấy bộ phân loại Support Vector Machines cho ra kết quả tốt nhất đối với cả phương pháp tĩnh và động. Bên cạnh đó, dựa trên dữ liệu từ Bảng 3.3, có thể thấy phương pháp lai cho ra kết quả cao tốt hơn so với việc sử dụng riêng lẻ từng tập đặc trưng tĩnh và động cho thực nghiệm.

Bảng 3.3: So sánh với một số phương pháp khác

Phương pháp	ACC (%)	AUC (%)	Precision (%)	TPR (%)	F1 (%)
HaddadPajouh	96,35	96,28	97,74	96,53	97,13
Su	95,13	95,23	96,9	94,67	95,77
Luận văn	99,18	98,79	99,14	99,71	99,43

So sánh mô hình phương pháp lai đã nghiên cứu với các mô hình phát hiện mã độc IoT Botnet trong một số nghiên cứu khác cũng cho kết quả khả quan theo Bảng 3.4, thể hiện sự hiệu quả và cải tiến của phương pháp lai so với các phương pháp tĩnh và động đơn thuần.

Kết luận chương 3

Trong chương 3, luận văn đã trình bày về quá trình thử nghiệm và đánh giá mô hình phương pháp lai đã trình bày ở chương 2, bao gồm xây dựng tập dữ liệu, lựa chọn phương pháp đánh giá, tiến hành thử nghiệm, và cuối cùng là so sánh và đưa ra kết luận đánh giá.

Kết quả đánh giá cho thấy phương pháp lai thử nghiệm có độ chính xác khá cao và tỷ lệ dương sai thấp, có khả năng phân loại mã độc và lành tính khá tốt. Đồng thời, kết quả này cũng thể hiện được tính hiệu quả và cải tiến của phương pháp lai thử nghiệm trong luận văn.

KẾT LUẬN VÀ KIẾN NGHỊ

Sự phát triển mạnh mẽ của mạng Internet và đặc biệt thời gian gần đây là Internet vạn vật (IoT - Internet of Things) đã mang lại các tiện ích phong phú cho người dùng như trong truyền nhận dữ liệu, liên lạc, quản lý, theo dõi sức khoẻ, thiết bị tự hành hay cơ sở hạ tầng thông minh. Phân tích và phát hiện mã độc IoT Botnet trên các thiết bị IoT là một hướng quan trọng về mặt lý thuyết cũng như ứng dụng, là cơ sở nền tảng để mở rộng để nghiên cứu cho tất cả các loại mã độc trên thiết bị IoT, và là một trong những nội dung quan trọng để phòng chống nguy cơ tấn công mạng trong hoạt động đảm bảo an ninh mạng, góp phần bảo vệ an ninh quốc gia, trật tự an toàn xã hội.

Để xác định phạm vi nghiên cứu, luận văn tiến hành nghiên cứu các phương pháp phân tích, phát hiện mã độc Botnet trên thiết bị IoT, đồng thời tìm hiểu hướng tiếp cận sử dụng phương pháp lai trong phát hiện mã độc IoT Botnet. Mô hình của phương pháp lai đã được tiến hành thực nghiệm với tập dữ liệu các mẫu IoT Botnet được thu thập trên thực tế. Kết quả đánh giá đã cho thấy rõ ưu thế của phương pháp lai so với các phương pháp tĩnh và động đơn thuần.

Kết quả của luận văn là bước đầu cho việc mở rộng hướng phát hiện mã độc IoT Botnet dựa trên phương pháp lai, kết hợp các ưu điểm tốt nhất của các phương pháp phân tích tĩnh và phân tích động. Nội dung nghiên cứu của luận văn được trình bày và công bố tại Kỷ yếu hội nghị quốc tế, cụ thể là*"IoT Botnet Detection Based on the Integration of Static and Dynamic Vector Features"* tại Hội nghị quốc tế lần thứ tám về Truyền thông và Điện tử (ICCE 2020) và "Toward an approach using graph-theoretic for IoT botnet detection" tại Hội nghị quốc tế về máy tính, mạng và Internet vạn vật (CNIOT 2021)

Tuy nhiên, luận văn vẫn còn một số hạn chế trong xây dựng tập đặc trưng lai. Cụ thể là những khó khăn trong việc dịch ngược một số mẫu mã độc IoT Botnet để trích xuất đồ thị PSI. Bên cạnh đó, do công cụ V-sandbox mới chỉ hỗ trợ 5 loại nền tảng kiến trúc và danh sách các chỉ thị của máy chủ C&C mô phỏng vẫn còn hạn chế nên chưa thể mô phỏng được hết các mẫu mã độc. Ngoài ra, phương pháp tích hợp các đặc trưng tĩnh và động còn khá đơn giản và chỉ phù hợp khi tích hợp số ít các loại vector đặc trưng. Trong trường hợp cần tích hợp nhiều loại đặc trưng cả tĩnh và động thì phương pháp tích hợp này có thể sẽ không hiệu quả.

Dựa trên kết quả nghiên cứu, luận văn đưa ra một số kiến nghị cho các hướng phát triển trong tương lai như sau:

- Tiếp tục cải thiện phương pháp tích hợp các đặc trưng tĩnh và động, để có thể tích hợp được nhiều đặc trưng mạnh hơn, góp phần cải thiện độ chính xác của phương pháp phát hiện mã độc IoT Botnet nói riêng và mã độc IoT nói chung.

- Nghiên cứu các phương pháp phòng thủ, bảo vệ các thiết bị IoT khỏi bị lây nhiễm bởi mã độc IoT Botnet nói chung và mã độc IoT nói chung. Từ đó xây dựng các bộ công cụ kiểm định, đánh giá độ an ninh, an toàn của thiết bị IoT trước khi đưa vào sử dụng trong các hệ thống thông tin, nhất là các hệ thống thông tin quan trọng về an ninh quốc gia.

- Nghiên cứu các phương pháp phòng, chống hoạt động tấn công mạng sử dụng mã độc IoT Botnet của các cá nhân, tổ chức. Xây dựng quy trình ứng phó với phương thức tấn công mạng này để xác định nguồn gốc, loại trừ hành vi tấn công mạng và thu thập các tài liệu chứng cứ phục vụ quá trình điều tra, truy tố.