HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Lưu Bích Hạnh

NGHIÊN CỨU GIẢI PHÁP PHÁT HIỆN XÂM NHẬP (IDS) DỰA TRÊN CÔNG NGHỆ HỌC MÁY CHO IoT GATEWAY

Chuyên ngành: Kỹ thuật viễn thông
Mã số: 8.52.02.08

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT
(*Theo định hướng ứng dụng*)

HÀ NỘI - NĂM 2022

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: **PGS.TS. LÊ HẢI CHÂU**

Phản biện 1: PGS. TS HOÀNG MẠNH THẮNG

Phản biện 2: TS. NGUYỄN NGỌC MINH

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

Vào lúc: 10 giờ 00 ngày 2 tháng 7 năm 2022

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

MỞ ĐẦU

Hiện nay, IoT đang ngày càng bùng nổ và được coi là một xu hướng mới. Bên cạnh các tiện ích mà IoT mang lại thì IoT cũng hàm chứa những mối nguy hại. Trong những năm vừa qua, IoT cũng đang dần trở thành mục tiêu hàng đầu của giới hacker trên toàn thế giới. Vì vậy, các giải pháp ứng dụng an ninh và bảo mật cho các thiết bị IoT ngày càng được quan tâm hơn. Các sản phẩm công nghệ IoT ngày càng đa dạng về chất lượng và bùng nổ về số lượng nên hệ thống phát hiện xâm nhập (IDS) hiện đang là một trong những giải pháp được quan tâm hàng đầu hiện nay nhằm bảo vệ linh hoạt, hiệu quả trước vô vàn cuộc xâm nhập trái phép trên Internet nhắm tới các thiết bị IoT. Ở đây, chúng ta có thể phát hiện ngay lập tức những hành vi truy nhập bất thường khi sử dụng kỹ thuật học máy. Thực hiện bằng cách thiết lập mô hình dựa vào những thuật toán học máy, thuật toán thống kê hoặc mạng Nơ ron nhân tạo.

Do vậy, với mục tiêu nghiên cứu, tìm hiểu và nắm bắt các giải pháp phát hiện xâm nhập hiệu quả cho các thiết bị IoT gateway, nội dung luận văn tập trung nghiên cứu, xây dựng và thử nghiệm giải pháp phát hiện xâm nhập dựa trên công nghệ học máy cho các thiết bị IoT gateway.

Luận văn được trình bày theo 03 chương với nội dung chính như sau:

- ***Chương 1 - Tổng quan về IoT, IoT gateway và các kỹ thuật phát hiện xâm nhập:*** Giới thiệu tổng quan về công nghệ IoT, khái niệm, vai trò và vị trí của thiết bị IoT gateway, đồng thời trình bày kiến trúc, thành phần và chức năng của các thành phần trong hệ thống IDS cùng khả năng ứng dụng, triển khai các hệ thống IDS trên IoT Gateway.
- ***Chương 2 - Giải pháp phát hiện xâm nhập ứng dụng học máy:*** Giới thiệu tổng quan về giải pháp phát hiện xâm nhập cho IoT gateway, đồng thời trình bày các kỹ thuật học máy cơ bản sử dụng trong phát hiện xâm nhập, mô tả chi tiết tập dữ liệu mẫu, phân tích và lựa chọn thuật toán học máy để hỗ trợ cho việc thực hiện đánh giá hiệu quả khi ứng dụng thuật toán học máy trong phát hiện xâm nhập cho IoT gateway.
- ***Chương 3 – Thử nghiệm hệ thống IDS trên IoT gateway:*** Trình bày mô hình phát hiện xâm nhập trên IoT gateway, xây dựng kiến trúc hệ thống phát hiện xâm nhập cho IoT gateway dựa trên học máy, đồng thời thiết lập thử nghiệm hệ thống IDS ứng dụng giải pháp mạng Nơ ron và thuật toán *Random Forest* từ đó đưa ra kết quả đánh giá thử nghiệm.

CHƯƠNG 1: TỔNG QUAN VỀ IoT, IoT GATEWAY VÀ KỸ THUẬT PHÁT HIỆN XÂM NHẬP

1.1 Giới thiệu chung

1.1.1 Công nghệ IoT

IoT được viết tắt bởi cụm từ Internet of Things – Công nghệ Internet vạn vật, mang ý nghĩa kết nối mọi thứ với Internet. Có thể khái quát rằng Internet of Things đề cập đến những thiết bị vật lý ở tất cả mọi nơi có khả năng kết nối với nhau, với Internet để biến mọi thứ trở nên chủ động, thông minh hơn. Hiện nay chúng ta có thể bắt gặp IoT ở khắp mọi nơi, ví dụ như xe tự lái, nhà thông minh, thiết bị đeo theo dõi sức khỏe.

1.1.2 Các thiết bị IoT gateway

Thiết bị được sử dụng để kết nối các thiết bị khác với đám mây hay trung tâm dữ liệu sẽ được gọi là IoT gateway. IoT Gateway có thể cung cấp cơ chế bảo mật bổ sung cho mạng IoT và dữ liệu được nó vận chuyển. Vì gateway quản lý thông tin di chuyển theo cả hai chiều, do đó có thể bảo vệ dữ liệu khi di chuyển lên đám mây khỏi bị đánh cắp và hạn chế các thiết bị IoT bị xâm phạm bởi các cuộc tấn công bên ngoài.

1.1.3 Các vấn đề an toàn thông tin trong IoT

Thiết bị IoT rất dễ bị tấn công mạng, do đó nếu gặp phải tấn công sẽ hình thành những lỗ hổng khi tiếp xúc nhiều thiết bị, làm hệ sinh thái bị lộ. Nhằm hỗ trợ an ninh mạng, các thiết bị IoT đều thông qua IoT Gateway như thêm một lớp bảo vệ cho hệ sinh thái.

a. Những rủi ro an ninh trong IoT
   - Chưa có một giao thức chung
   - Vấn đề bảo mật của các thiết bị và Gateway
   - Không mất dữ liệu
   - Tấn công vật lý
   - Quyền riêng tư thông tin

b. Những mối đe dọa an toàn thông tin

- Các thiết bị chỉ được bảo vệ bởi mã hóa cứng hoặc mật khẩu yếu khi kết nối với Internet.
- Các thiết bị thông minh hiện nay rất dễ gặp phải lỗ hổng bảo mật Zero-day, cơ hội để các hacker sinh sôi, nảy nở.
- Việc công khai các CVE của thiết bị IoT như Router cũng gây ảnh hưởng đến uy tín của nhà cung cấp và cả doanh nghiệp.
- Khi lập trình IoT trên Linux việc không biết những thư viện tải về dùng để code cũng sẽ là một vấn đề hết sức nguy hiểm.

1.2 Hệ thống phát hiện xâm nhập (IDS)

1.2.1 Giới thiệu chung

IDS (Intrusion Detection Systems - Hệ thống phát hiện xâm nhập) là một hệ thống có nhiệm vụ theo dõi, giám sát cũng như phát hiện ra các hành vi đáng ngờ, điều này sẽ giúp ngăn chặn sự xâm nhập trái phép vào hệ thống thông tin. Mục đích của IDS là tìm ra và ngăn chặn các tấn công gây ảnh hưởng đến tính bảo mật, sự toàn vẹn thông tin của hệ thống. Hơn thế nữa, IDS còn có khả năng phân biệt giữa các cuộc tấn công từ bên ngoài với những cuộc tấn công nội bộ. Hệ thống IDS sẽ thu thập thông từ các nguồn trong hệ thống an ninh sau đó phân tích nhằm phát hiện sự xâm nhập trái phép.

1.2.2 Kiến trúc IDS

Hai hướng triển khai hệ thống IDS là tập trung và phân tán. IDS tích hợp cùng Firewall là hướng tập trung còn khi nhiều hệ thống IDS trong cùng 1 mạng lớn được kết nối với nhau là hướng phân tán. Có 2 loại hệ thống IDS cơ bản như sau:

- HIDS (Host based IDS): Phát hiện xâm nhập dựa vào việc sử dụng dữ liệu kiểm tra từ một máy trạm đơn.
- NIDS (Network based IDS): Phát hiện xâm nhập dựa vào việc sử dụng dữ liệu tại toàn bộ các lưu thông mạng và dữ liệu kiểm tra tại 1 hoặc nhiều máy trạm.

1.2.3 Thành phần chính của hệ thống phát hiện xâm nhập

IDS bao gồm 3 thành phần chính sau:

- Information collection (thành phần thu thập gói tin)
- Detection (thành phần phân tích gói tin)

- Response (thành phần phản hồi): khi gói tin được phát hiện ra là một cuộc tấn công

1.2.4 Chức năng của IDS

Những chức năng chính của IDS:

- IDS có chức năng giám sát thành phần cần bảo vệ trong hệ thống trước các hoạt động bất thường.
- IDS cần phân tích hành vi truy cập, hoạt động, sự kiện quan trọng liên quan đến thành phần được giám sát dựa vào những hành vi bất thường, tập luật, baseline.
- IDS mang đến những cảnh báo hiểm họa an toàn thông tin. Thay vì dùng những thiết lập mặc định thì cần nâng cao hơn để chống lại kẻ xâm nhập.
- IDS cũng phải thống kê và trích xuất báo cáo.

Những chức năng cơ bản của IDS:

- IDS sẽ cung cấp 1 cách nhìn tổng thể về lưu lượng mạng.
- IDS cũng giúp nhận diện những hoạt động thâm nhập hay tấn công hệ thống.
- IDS hỗ trợ kiểm tra những sự cố xảy ra trong hệ thống mạng.
- IDS còn được sử dụng để thu gom bằng chứng như log, event, flow… cho quá trình điều tra cũng như đối đầu với các sự cố bảo mật.
- IDS sẽ nhận diện được nguy cơ mất an toàn thông tin sẽ xảy ra.
- IDS cũng nhanh chóng phát hiện ra điểm yếu của hệ thống, các lỗ hổng trong chính sách bảo mật.

1.3 Phát hiện xâm nhập trong hệ thống IoT

1.3.1 Kiến trúc chung

Kiến trúc chung của IoT được chia thành năm lớp bao gồm ba miền, đó là miền ứng dụng, miền mạng và miền vật lý. Do đó, IoT có thể được tùy chỉnh để phù hợp với nhu cầu của các môi trường thông minh khác nhau.

Tầng nhận thức là tầng phần cứng bao gồm các cảm biến và các đối tượng vật lý ở các dạng khác nhau.

Tầng mạng là tầng giúp truyền thông tin từ các đối tượng vật lý hoặc cảm biến đến hệ thống xử lý qua các đường dây an toàn bằng hệ thống truyền thông.

Tầng phần mềm trung gian chịu trách nhiệm quản lý dịch vụ trên các thiết bị IoT để tạo

kết nối giữa các thiết bị IoT cung cấp cùng một dịch vụ. Hơn nữa, tầng phần mềm trung gian lưu trữ thông tin đến từ tầng mạng trong cơ sở dữ liệu để tạo điều kiện cho việc ra quyết định trên cơ sở các hoạt động xử lý thông tin.

Tầng ứng dụng chịu trách nhiệm quản lý toàn cầu các ứng dụng IoT. Tầng ứng dụng phụ thuộc vào thông tin được xử lý trong tầng phần mềm trung gian.

Tầng nghiệp vụ cũng chịu trách nhiệm quản lý toàn cầu các ứng dụng IoT cũng như quản lý dịch vụ trên các thiết bị IoT.

1.3.2 Môi trường thông minh

Thuật ngữ thông minh dùng để chỉ khả năng thu nhận và áp dụng kiến thức một cách tự chủ, thuật ngữ môi trường dùng để chỉ môi trường xung quanh.

Môi trường thông minh cung cấp một số tính năng nhất định có thể được sử dụng để nâng cao chất lượng dịch vụ (QoS) của các ứng dụng người dùng. Thông tin thời gian thực là một trong những tính năng này. Việc tích hợp môi trường thông minh và IoT mang đến những cơ hội mới liên quan đến QoS của các dịch vụ và ứng dụng. Mục tiêu của môi trường thông minh là làm cho cuộc sống của con người thoải mái và hiệu quả hơn bằng cách sử dụng các cảm biến. Bằng mạng IoT, các cảm biến có thể được theo dõi và điều khiển từ xa. Mô hình Internet of Things (IoT) gần đây đã phát triển thành một công nghệ để xây dựng môi trường thông minh. Do đó, việc bảo mật các hệ thống IoT đã trở thành mối quan tâm chính.

Vấn đề bảo mật thông tin và quyền riêng tư được coi là vấn đề chính trong bất kỳ môi trường thông minh nào trong thời đại IoT. Các lỗ hổng bảo mật trong hệ thống IoT tạo ra các mối đe dọa bảo mật ảnh hưởng đến các ứng dụng thông minh. Hệ thống phát hiện xâm nhập (IDS) là một cơ chế bảo mật hoạt động chủ yếu trong lớp mạng của hệ thống IoT.

IDS được triển khai cho hệ thống IoT sẽ có thể phân tích các gói dữ liệu và tạo phản hồi trong thời gian thực, phân tích các gói dữ liệu trong các lớp khác nhau của mạng IoT với các ngăn xếp giao thức khác nhau và thích ứng với các công nghệ khác nhau trong môi trường IoT. IDS được thiết kế cho môi trường thông minh dựa trên IoT sẽ hoạt động trong các điều kiện nghiêm ngặt về khả năng xử lý thấp, phản hồi nhanh và xử lý dữ liệu khối lượng lớn.

1.4 Ứng dụng giải pháp phát hiện xâm nhập trên IoT gateway

1.4.1 Các kỹ thuật phát hiện xâm nhập

Có 2 kỹ thuật phát hiện xâm nhập cơ bản là: Hệ thống phát hiện xâm nhập dựa trên dấu hiệu và hệ thống phát hiện xâm nhập dựa trên bất thường.

a. Kỹ thuật phát hiện xâm nhập dựa vào dấu hiệu

   Phát hiện xâm nhập dựa trên dấu hiệu sử dụng 1 cơ sở dữ liệu gồm: chữ kí, mô hình mã độc và xâm nhập đã biết để phát hiện ra các cuộc tấn công nổi tiếng.

IDS dựa trên dấu hiệu này được thiết kế nhằm phát hiện các cuộc tấn công, xâm nhập độc hại dựa vào kiến thức trước đó.

b. Kỹ thuật phát hiện xâm nhập dựa vào sự bất thường

Đối với kỹ thuật phát hiện xâm nhập dựa vào sự bất thường, mẫu dữ liệu sẽ được tạo ra dựa trên dữ liệu của người dùng bình thường và sau đó đem so sánh với các mẫu dữ liệu đang có để phát hiện ra điều bất thường nếu có. Các hành động bất thường sẽ được phát hiện cho việc xác định những cuộc tấn công, đặc biệt là với các cuộc tấn công chưa biết. IDS dựa vào sự bất thường sẽ hoạt động bằng cách tạo ra một mô hình hành vi bình thường trong môi trường máy tính đã được cập nhật liên tục và dựa trên dữ liệu của người dùng bình thường, sau đó sử dụng mô hình này để phát hiện ra bất cứ sai sót nào so với các hành vi bình thường.

1.4.2 Ứng dụng trong IoT gateway

Luận văn sẽ tập trung mô tả 2 hướng tiếp cận hệ thống IDS là: Phân loại IDS dựa trên vị trí chiến lược và phương thức phát hiện.

a. Vị trí chiến lược

- Mô hình tập trung: Vị trí IDS tập trung sẽ được đặt trong một thành phần tập trung sử dụng để phân tích các gói tin đi qua bộ định tuyến biên nằm ở giữa miền vật lý và miền mạng.
- Mô hình phân tán: Vị trí IDS phân tán đặt mọi đối tượng vật lý, ở đây sẽ đề xuất hai kỹ thuật như chuyển dịch phụ trợ và quyết định sớm. Mục đích chính của IDS này là để giảm thiểu các tài nguyên tính toán cần thiết sử dụng trong phát hiện xâm nhập.
- Mô hình lai: Vị trí IDS lai kết hợp các khái niệm về vị trí tập trung và phân tán vào để tận dụng điểm mạnh và tránh đi phần nhược điểm. Cách tiếp cận đầu tiên là cho vị trí kết hợp tổ chức mạng thành các cụm hoặc vùng, chỉ nút chính của mỗi cụm lưu trữ một thể hiện IDS.

b. Phương thức phát hiện

- Dựa vào chữ ký: Trong các cách tiếp cận dựa vào chữ ký thì IDS phát hiện các cuộc tấn công dựa vào hành vi của hệ thống hoặc mạng, khớp với chữ ký tấn công đã được lưu ở trong cơ sở dữ liệu của IDS.
- IDS dựa trên sự bất thường: Hay còn được gọi là phát hiện dựa trên sự kiện. Kỹ thuật này sẽ giúp xác định các hoạt động độc hại bằng cách phân tích sự kiện. Đầu tiên cần xác định hành vi bình thường của mạng, nếu có bất kỳ hoạt động nào khác với hành vi bình thường thì dấu hiệu đó là một sự xâm nhập.
- Dựa vào thông số kỹ thuật: Kỹ thuật dựa vào thông số kỹ thuật sẽ hơi giống với kỹ thuật phát hiện bất thường. Nhưng trong kỹ thuật này, hành vi bình thường của mạng được xác định bằng tay vì thế nó cho tỷ lệ dương ít chính xác hơn.
- Phương pháp lai: Phương pháp này là sự kết hợp của các khái niệm: phát hiện dựa trên dấu hiệu, dựa trên đặc điểm kỹ thuật và bất thường để tối ưu hóa ưu điểm và giảm thiểu đi tác động của nhược điểm.

1.5 Kết luận Chương 1

Chương 1 giới thiệu tổng quan về công nghệ IoT, các thiết bị IoT gateway cũng như các vấn đề an toàn thông tin trong IoT. Đồng thời, nội dung chương cũng tập trung trình bày về những yếu tố cản trở và hiểm họa của IoT cũng như khái niệm, kiến trúc và chức năng các thành phần của hệ thống phát hiện xâm nhập, các kỹ thuật nhằm phát hiện xâm nhập trái phép, đưa ra quy trình chung, thành phần, mô hình của các giải pháp.

CHƯƠNG 2: GIẢI PHÁP PHÁT HIỆN XÂM NHẬP ỨNG DỤNG HỌC MÁY

2.1 Giới thiệu chung

Dưới sự phát triển của công nghệ học máy, những giải pháp phát hiện xâm nhập hiện đã và đang được chú trọng nghiên cứu, phát triển cũng như ứng dụng vào thực tế nhằm đóng góp và cải thiện tỷ lệ phát hiện, tính chính xác cũng như giảm đi tối đa số lượng cảnh báo nhầm. Có thể thấy rằng bài toán phân biệt các hành vi truy nhập là bình thường hay bất thường và dùng đến những tài nguyên của hệ thống là một bài toán điển hình trong kỹ thuật học máy. Dựa vào những đặc điểm của từng hành vi thu thập được, hệ thống học máy sẽ dựng lên mô hình tự động phân loại để cho ra kết quả chính xác nhất.

2.2 Một số kỹ thuật học máy sử dụng trong phát hiện xâm nhập

2.2.1 K-Nearest Neighbors

K-NN được xem như một phương pháp để phân lớp các đối tượng theo khoảng cách gần nhất giữa đối tượng cần xếp lớp (*Query point*) với tất cả những đối tượng khác trong dữ liệu đào tạo. KNN được sử dụng để áp dụng vào hai loại của bài toán học có giám sát, đó là phân lớp, hồi quy. Có thể thấy, kết quả dự đoán của một điểm dữ liệu mới sẽ được chỉ ra trực tiếp từ k điểm dữ liệu gần nhất tại tập dữ liệu huấn luyện. Mô tả thuật toán K-NN:

- Xác định giá trị tham số K (số neighbors gần nhất).
- Tính khoảng cách giữa đối tượng cần phân lớp (*Query Point*) với tất cả những đối tượng trong training data (thường sử dụng khoảng cách *Euclidean*).
- Xếp thứ tự khoảng cách tăng dần và xác định K láng giềng gần nhất với *Query Point*.
- Lấy toàn bộ các lớp của K láng giềng gần nhất đã xác định.
- Việc xác định lớp cho *Query Point* sẽ phụ thuộc vào các lớp của láng giềng gần nhất.

- Trong các bài toán phân lớp, kết quả đầu ra sẽ là lớp mà dữ liệu thuộc về và phụ thuộc vào việc bình chọn của k điểm gần nhất. Đồng thời, trong bài toán hồi quy thì đầu ra của một điểm dữ liệu sẽ bằng trung bình đầu ra của k điểm gần nhất.

Những phương pháp đo khoảng cách giữa các điểm để tìm ra điểm gần nhất phổ biến bao gồm khoảng cách *Hamming*, khoảng cách *Manhattan*, khoảng cách *Minkowski*.

2.2.2 SVM

SVM (*Support Vector Machine*) là thuật được sử dụng nhiều trong các bài toán phân lớp. Ý tưởng nằm ở việc tìm ra một siêu mặt phẳng phân chia các lớp tối ưu nhất. Cụm từ '*support vector*' nhằm chỉ ra rằng các điểm nằm gần siêu mặt phẳng nhất nếu bị xóa đi có thể khiến vị trí của siêu mặt phẳng bị ảnh hưởng. Đồng thời ta cũng có giá trị biên (*margin*) là khoảng cách giữa *support vector* và siêu mặt phẳng.

Siêu mặt phẳng càng nằm xa các lớp chứng tỏ dự đoán càng chính xác. Do đó mặc dù có thể tìm được rất nhiều siêu mặt phẳng cho mỗi bài toán nhưng việc tìm kiếm được một siêu mặt phẳng để biên lớn nhất vẫn là mục tiêu của SVM. Thuật toán SVM thường sẽ cho ra những kết quả khá chính xác.

2.2.3 Naive Bayes

Naive Bayes là thuật toán phân lớp dựa vào định lý *Bayes* về lý thuyết xác suất. Đây là thuật toán có thể được sử dụng cho các bài toán nhị phân hoặc phân lớp nhiều lớp. Thuật toán *Naive Bayes* sẽ xử lý từng đặc trưng một cách độc lập, nó sẽ tính xác suất của mỗi đặc trưng trước và đưa ra dự đoán dựa vào định lý *Bayes*. Thuật toán Naive Bayes có ưu điểm đơn giản và dễ hiểu. Thuật toán sẽ thích hợp với những tập dữ liệu có nhiều đặc trưng khác nhau do dự đoán cần phụ thuộc vào xác suất của các đặc trưng đó.

2.2.4 J48 Decision Tree

Thuật toán cây quyết định J48 (*J48 Decision Tree*): Đây là một cây phân cấp có cấu trúc được sử dụng để phân lớp các đối tượng dựa vào dãy các luật. Nếu đưa dữ liệu cho các đối tượng bao gồm những thuộc tính cùng với lớp của nó, lúc này *J48 Decision Tree* sẽ sinh ra các luật nhằm dự đoán lớp của các đối tượng chưa biết. Điều mà thuật toán *J48 Decision Tree* muốn đạt được là kết quả chính xác nhất với số lần lựa chọn ít nhất. Thuật toán cây quyết định J48 xử lý tốt các tập dữ liệu lớn cũng như

có nhiều dữ liệu nhiễu, theo dõi quá trình lựa chọn một cách tường minh, do đó cây quyết định trở thành thuật toán phổ biến.

2.2.5 Random Forest

Thuật toán *Random Forest* khá thông dụng và được dùng nhiều trong học máy. Điểm đặc biệt là *Random Forest* hầu như không cần xử lý dữ liệu hay lập mô hình trước đó nhưng vẫn mang lại kết quả tương đối chính xác. Thuật toán xây dựng trên tính ngẫu nhiên và được tạo ra bởi nhiều cây quyết định.

Thuật toán *Random Forest* sẽ coi mỗi cây quyết định giống như một cử tri bỏ phiếu độc lập. Khi đó cuối cuộc bầu cử, câu trả lời nhận được nhiều bình chọn nhất từ các cây quyết định sẽ là câu trả lời được lựa chọn. *Random Forest* có một cách khác để chắc chắn rằng tất cả các cây quyết định sẽ không cho cùng câu trả lời, đó là chọn ngẫu nhiên các quan sát.

2.2.6 Mạng Nơ Ron

a. Khái niệm

Mạng nơ ron nhân tạo (ANN) được tạo nên từ một số lượng lớn các phần tử nơ ron liên kết với nhau. Tại đây, mỗi nơ ron sẽ tính tổng những giá trị đầu vào với mỗi trọng số đã học được, sau đó chuyển kết quả cho hàm trả về một giá trị. Chức năng logistic sẽ là một lựa chọn phổ biến cho chức năng kích hoạt. Trong đó, bố cục của các nơ ron cũng phụ thuộc vào kiến trúc mạng.

b. Kiến trúc

Các mạng nơ ron truyền thống sẽ được chia thành ba loại kiến trúc mạng khác nhau là mạng cấp dữ liệu một lớp, mạng cấp dữ liệu đa lớp và mạng hồi quy.

ANN chuyển tiếp sẽ được xây dựng dựa vào một hoặc nhiều lớp tế bào nơ ron kết nối cùng với các lớp tế bào nơ ron sau, không có kết nối với các lớp trước. Đồng thời đối với mạng chuyển tiếp nguồn cấp một lớp thì chỉ có lớp đầu ra của các nút thực hiện toàn bộ các tính toán. Đối với trường hợp những mạng chuyển tiếp nguồn cấp đa lớp thì các lớp nằm giữa lớp đầu ra và nguồn được gọi là các lớp ẩn.

c. Quá trình xử lý

Kết quả đầu ra của một ANN là giải pháp cho một vấn đề cụ thể nào đó. Trọng số liên kết thể hiện độ cần thiết của dữ liệu đầu vào với tiến trình xử lý thông tin. Việc thay đổi những trọng số của dữ liệu đầu vào để có được kết quả mong muốn chính là quá trình học của ANN. Mạng nơ ron nhân tạo được huấn luyện theo hai kỹ thuật cơ bản đó là học có giám sát và học không giám sát.

d. Phương thức huấn luyện

Mạng nơ ron sẽ có ba cách huấn luyện chính là: Huấn luyện theo gói, huấn luyện ngẫu nhiên và huấn luyện trực tuyến. Riêng đối với huấn luyện trực tuyến thì các trọng số của mạng sẽ được cập nhật ngay lập tức sau khi một mẫu đầu vào được đưa vào mạng. Còn huấn luyện ngẫu nhiên cũng khá giống với huấn luyện trực tuyến nhưng ở đây việc chọn các mẫu đầu vào để đưa vào mạng từ tập huấn luyện sẽ được thực hiện ngẫu nhiên. Đối với huấn luyện theo gói thì tất cả các mẫu đầu vào sẽ được đưa vào mạng cùng lúc, sau đó cập nhật các trọng số mạng đồng thời. Trong quá trình huấn luyện mạng, thuật ngữ "*epoch*" được dùng để thể hiện quá trình.

2.3 Thuật toán học máy trên IoT gateway

2.3.1 Phân tích và lựa chọn mạng Nơ ron

Mạng Nơ ron là một mô hình học máy phổ biến, nét đặc trưng của mạng Nơ ron là khả năng học. Mạng Nơ ron có thể gần đúng mối quan hệ tương quan phức tạp giữa các yếu tố đầu vào và đầu ra của các quá trình cần nghiên cứu, sau khi đã học được thì việc kiểm tra độc lập thường sẽ cho ra kết quả tốt. Đồng thời, khi đã học xong, mạng Nơ ron nhân tạo có thể tính toán kết quả đầu ra tương ứng với bộ số liệu đầu vào mới. Về mặt cấu trúc, mạng Nơ ron nhân tạo là một hệ thống gồm nhiều phần tử xử lý đơn giản cùng hoạt động song song. Tính năng này của ANN cho phép nó có thể được áp dụng để giải các bài toán lớn.

Mạng Nơ ron được sử dụng để giải quyết nhiều bài toán thuộc nhiều lĩnh vực của các ngành khác nhau. Điển hình nhóm ứng dụng mà mạng Nơ ron đã được áp dụng rất có hiệu quả là bài toán phân lớp: Loại bài toán này đòi hỏi giải quyết vấn đề phân loại các đối tượng quan sát được thành các nhóm dựa trên các đặc điểm của các nhóm đối tượng đó. Đây là dạng bài toán cơ sở của rất nhiều bài toán trong thực tế: nhận dạng

chữ viết, tiếng nói, phân loại gen, phân loại chất lượng sản phẩm,… Do đó luận văn sẽ triển khai thử nghiệm mạng Nơ ron trên bộ dữ liệu UNSW-NB15.

2.3.2 Phân tích và lựa chọn thuật toán Random Forest

Khi so sánh với các thuật toán học có giám sát hiện giờ như *Boosting, Baging, Nearest neighbors*, SVM, Mạng nơ ron, C45... Có thể thấy thuật toán *Random Forest* (RF) cho độ chính xác phân lớp cao hơn. *Random Forest* phân loại hiệu quả các cuộc tấn công vì là một bộ phân loại đồng bộ và hoạt động tốt so với các phân loại truyền thống khác. Vì *Random Forest* là một thành viên trong họ thuật toán *Decision Tree* vậy nên tư tưởng chính của RF là tạo ra nhiều cây quyết định từ dataset, mỗi cây quyết định sẽ dự đoán một kết quả và kết quả nào được nhiều cây quyết định dự đoán nhất thì đó sẽ trở thành kết quả cuối cùng.

Một nhóm nghiên cứu khác đã chỉ ra kết quả của thuật toán *Random Forest* hiệu quả hơn SVM, *Naive Bayes, Decision Tree* trên bộ dữ liệu UNSW-NB15 với những chỉ số vượt trội như sau:

Bảng 2.1: Kết quả thử nghiệm của các thuật toán

Methods	**Accuracy**	**Sensitivity**	**Specificity**	**Training Time**	**Prediction Time**
Random Forest	**97.49**	**93.53**	**97.75**	**5.69**	**0.08**
SVM	92.28	92.13	91.15	38.91	0.20
Naive Bayes	74.19	92.16	67.82	2.25	0.18
Decision Tree	95.82	92.52	97.10	4.80	0.13

2.4 Phân tích và lựa chọn tập dữ liệu mẫu UNSW-NB15

Luận văn sẽ ứng dụng tập dữ liệu UNSW-NB15, đây là tập dữ liệu có được sự kết hợp của dữ liệu mạng bình thường và các phương thức tấn công hiện đại. Các gói tin mạng thô trong bộ dữ liệu UNSW-NB15 sẽ được xây dựng nhờ công cụ IXIA PerfectStorm ở trong Phòng thí nghiệm Cyber Range của Trung tâm An ninh mạng

(ACCS) - Australia để tạo thành hỗn hợp các hoạt động bình thường trong thực tế cũng như tổng hợp các hành vi tấn công mới.

Với tổng số bản ghi dữ liệu là 2 triệu, và 540.044 bản ghi được lưu trữ trong bốn tệp CSV. Cụ thể ở đây là UNSW-NB15_1.csv, UNSW-NB15_2.csv, UNSW-NB15_3.csv và UNSW-NB15_4.csv. Mọi tính năng được mô tả trong tập tin NUSW-NB15_features.csv. Ở đây các tính năng sẽ có nhiều loại khác nhau: *Integer, Float, Binary, Nominal và Timestamp*.

2.5 Kết luận chương 2

Nội dung Chương 2 tập trung trình bày các kỹ thuật học máy cơ bản sử dụng trong các hệ thống phát hiện xâm nhập. Đồng thời, trong Chương 2 các thông tin, phân tích và lựa chọn thuật toán học máy, tập dữ liệu mẫu cũng được mô tả chi tiết để hỗ trợ cho việc thực hiện thử nghiệm và đánh giá hiệu quả khi ứng dụng thuật toán học máy trong phát hiện xâm nhập trên IoT gateway.

CHƯƠNG 3: THỬ NGHIỆM HỆ THỐNG IDS TRÊN CÁC THIẾT BỊ IoT GATEWAY

3.1 Mô hình phát hiện xâm nhập trên IoT gateway

Mô hình phát hiện xâm nhập cho các thiết bị IoT gateway dựa trên học máy phải thực hiện phân tích chuyên sâu về lưu lượng mạng, gồm một số thành phần như tiền xử lý dữ liệu, xếp hạng, lựa chọn tính năng, phân lớp học máy và nhận dạng tấn công.

Bước 1: Tiền xử lý dữ liệu.

Bước 2: Xếp hạng và lựa chọn tính năng khắc phục dữ liệu đầu vào để tránh sự chênh lệch dẫn tới sự sai lệch kết quả.

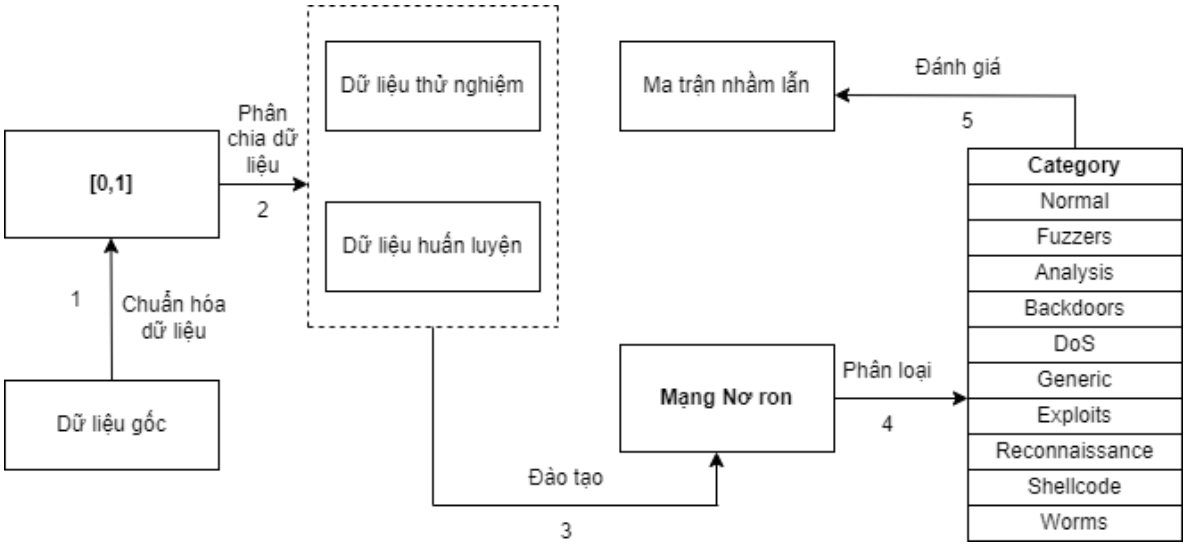
Bước 3: Áp dụng học máy phân loại dữ liệu tấn công hay dữ liệu bình thường.

Bước 4: Xác định kiểu tấn công dựa vào mạng Nơ ron.

Mô hình để phân loại các mẫu tập dữ liệu UNSW-NB15 được phát triển bằng phương pháp sử dụng một mạng nơ ron ngẫu nhiên và chuyển tiếp nguồn cấp dữ liệu. Tại đây, hệ thống sẽ sử dụng thuật toán *Random Forest* để phân loại đâu là dữ liệu bình thường và đâu là dữ liệu độc hại. Và từ đây dữ liệu tấn công sẽ tiếp tục được sử dụng để huấn luyện mạng nơ ron để phân loại thành những loại tấn công khác nhau.

3.2 Kiến trúc phát hiện xâm nhập cho IoT gateway dựa trên học máy

3.2.1 Kiến trúc giải pháp IDS sử dụng mạng Nơ ron

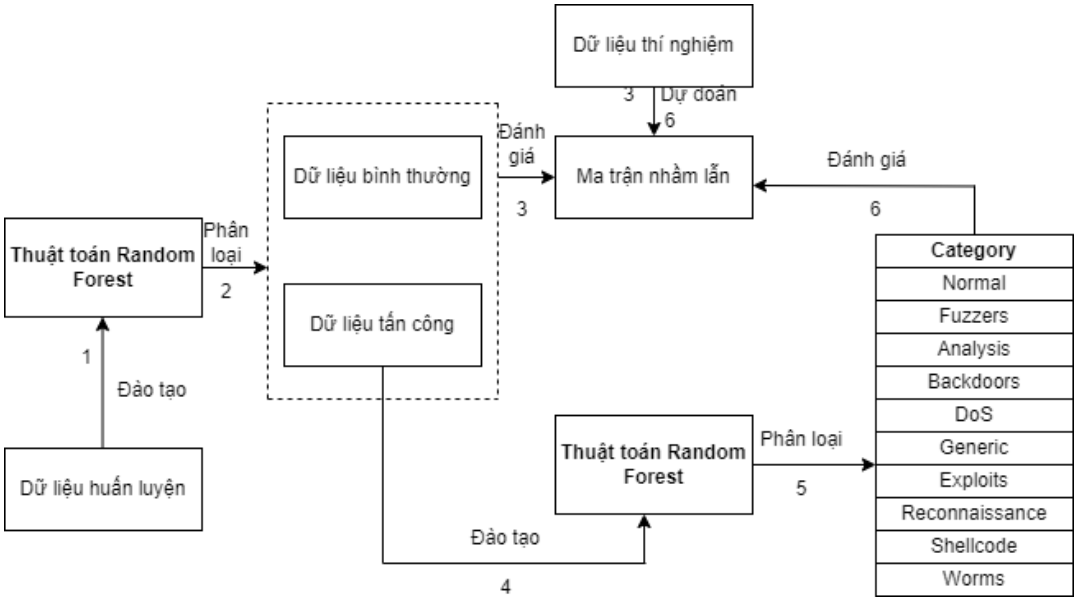


Hình 3.1: Kiến trúc tổng thể khi dùng mạng Nơ ron

Các bước thực hiện như sau:

- Bước 1: Dữ liệu gốc làm input đầu vào cho thuật toán RF, chuẩn hóa dữ liệu về dạng [0, 1].
- Bước 2: Phân chia dữ liệu thành dữ liệu thử nghiệm và dữ liệu huấn luyện.
- Bước 3: Mạng Nơ ron tiến hành đào tạo.
- Bước 4: Mạng Nơ ron phân loại dữ liệu tấn công thành 10 danh mục bao gồm: *Normal, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, Worms*.
- Bước 5: Cuối cùng ma trận nhầm lẫn sẽ đánh giá khả năng phân loại của mạng Nơ ron.

3.2.2 Kiến trúc giải pháp IDS sử dụng Random Forest



Hình 3.2: Kiến trúc tổng thể khi dùng Random Forest

Phân tích các bước thực hiện:

- Bước 1: Dữ liệu huấn luyện chính là input đầu vào cho thuật toán RF.
- Bước 2: Thuật toán RF phân loại dữ liệu thành 2 nhóm là: dữ liệu bình thường và dữ liệu tấn công.
- Bước 3: Dữ liệu thử nghiệm và kết quả phân loại ở bước 2 được Ma trận nhầm lẫn đánh giá dự đoán của thuật toán *Random Forest*.
- Bước 4: Tại bước này RF sử dụng dữ liệu tấn công để làm dữ liệu huấn luyện.

- Bước 5: RF sẽ tiếp tục phân loại dữ liệu tấn công thành 10 danh mục bao gồm: *Normal, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode* và *Worms*.

- Bước 6: Ma trận nhầm lẫn tiếp tục đánh giá khả năng phân loại của RF ở bước 5 dựa vào bộ dữ liệu thử nghiệm.

3.3 Thiết lập thử nghiệm phát hiện xâm nhập dựa trên thuật toán Random Forest và mạng Nơ ron

Quá trình thực hiện khi sử dụng mạng Nơ ron như sau:

- Bước 1: Tiền xử lý dữ liệu: Phân loại các thuộc tính → DictVectorizer các thuộc tính nominal → Gộp với các thuộc tính còn lại → Normalize tạo thành 294 tính năng.
- Bước 2: Thực hiện chia dữ liệu để đào tạo và kiểm tra bộ dữ liệu.
- Bước 3: Tiếp tục huấn luyện mạng Nơ ron với dữ liệu huấn luyện nhằm dự đoán, gán nhãn cho dữ liệu là dữ liệu tấn công hay dữ liệu bình thường.
- Bước 4: Dự đoán danh mục tấn công cùng với dữ liệu thử nghiệm.
- Bước 5: Áp dụng các số liệu hiệu suất để đo lường mức độ hệ thống tổng quát hóa dữ liệu.

Quá trình thực hiện khi sử dụng thuật toán Random Forest như sau:

- Bước 1: Tiền xử lý dữ liệu: Phân loại các thuộc tính → DictVectorizer các thuộc tính nominal → Gộp với các thuộc tính còn lại → Normalize tạo thành 294 tính năng.
- Bước 2: Thực hiện chia dữ liệu để đào tạo và kiểm tra bộ dữ liệu.
- Bước 3: Tiếp tục huấn luyện thuật toán Random Forest với dữ liệu huấn luyện nhằm dự đoán, gán nhãn cho dữ liệu là dữ liệu tấn công hay dữ liệu bình thường mà không cần kết hợp sử dụng tính năng.
- Bước 4: Sử dụng thuật toán Random Forest để tạo các tính năng cho dữ liệu thử nghiệm.
- Bước 5: Dự đoán danh mục tấn công cùng với dữ liệu thử nghiệm (với các tính năng được tạo ra ở bước 4).

- Bước 6: Tiếp theo sẽ áp dụng các số liệu hiệu suất để đo lường mức độ hệ thống tổng quát hóa dữ liệu.

Dữ liệu sẽ được đọc từ các tệp CSV với Pandas. Số lượng lớn dữ liệu mạng cũng là một thách thức không nhỏ trong mỗi bước xử lý.

3.3.1 Tiền xử lí dữ liệu

Tiền xử lý được chia thành hai giai đoạn. Trong đó giai đoạn đầu tiên được mô tả bên dưới (tệp pre_process.py) như sau:

- Đọc dữ liệu.
- Chia dữ liệu theo các loại tính năng.
- Chuyển đổi dữ liệu thành từng loại cho phù hợp.
- Thay thế NaN, bằng 0 xóa điểm dữ liệu.
- Sau đó cắt dữ liệu danh mục từ các khoảng trắng thừa, đặt chữ thường, vector hóa.
- Bình thường hóa dữ liệu giữa [0, 1].
- Lưu vào HDF5.

3.3.2 Phân tách dữ liệu

Giai đoạn thứ hai (tệp create_data_sets.py) tiến hành phân tách dữ liệu thành các phần huấn luyện và thử nghiệm. Tổng số mẫu sử dụng: 2.539.739 mẫu. Dữ liệu có thể được chia để có được 70% cho huấn luyện (1.777.812 mẫu) và 30% (761.927 mẫu) cho thử nghiệm với mỗi danh mục, vì các danh mục cũng sẽ rất mất cân bằng. Được mô tả như sau:

- Đọc dữ liệu đã xử lý trước từ HDF5.
- Chia dữ liệu thành các tập dữ liệu thử nghiệm và huấn luyện không chồng chéo.
- Tìm các tính năng quan trọng cho cả hai mô hình.
- Chọn các tính năng quan trọng nhất.
- Lưu tập dữ liệu vào HDF5.

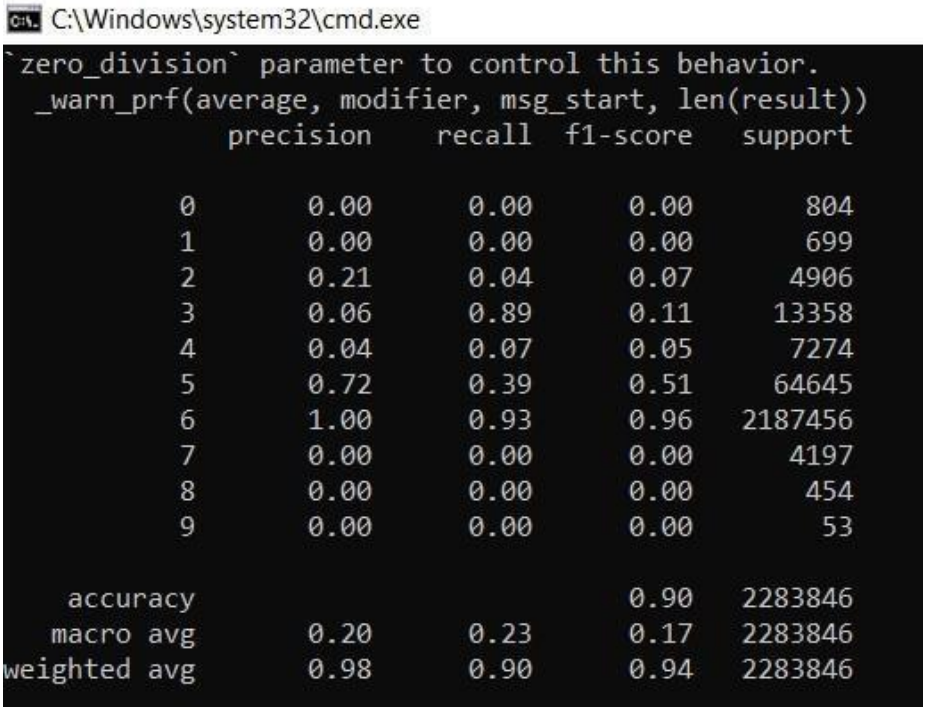
Tiếp theo đến bước giảm tính năng. Dữ liệu có điểm này 294 tính năng sau khi vector hóa. ExtraTreesClassifier đã được sử dụng để chọn 10 tính năng quan trọng nhất cho phân loại tấn công hay bình thường (Random Forest) và 25 tính năng quan trọng nhất

cho mạng Nơ ron. Hình 3.7 và 3.8 sẽ minh họa cho các tính năng nào có tầm quan trọng trong cả hai nhiệm vụ phân loại.

Chia từng loại dữ liệu thành dữ liệu huấn luyện và thử nghiệm cho cả thuật toán *Random Forest* và mạng Nơ ron (NN). Giá trị điểm dữ liệu chỉ nhằm mục đích minh họa.

3.4 Đánh giá kết quả thử nghiệm

3.4.1 Kết quả khi sử dụng mạng Nơ ron

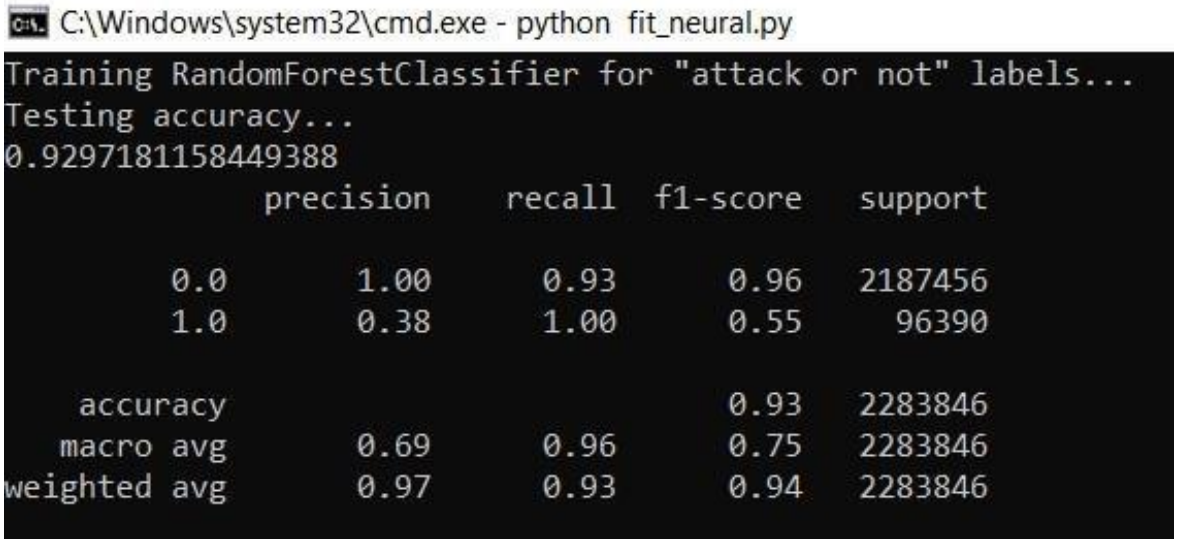


Hình 3.3: Kết quả phân loại tấn công (mạng Nơ ron)

Kết quả cho thấy mạng Nơ ron hoạt động khá ổn với dữ liệu này, độ chính xác 0, 90 cho dữ liệu cuộc tấn công. Các lớp tấn công được thể hiện bằng các số 0 hay 9. Số 6 là một điểm dữ liệu bình thường của điểm số và phần còn lại của các số là các loại tấn công khác nhau.

Lớp 6 hầu như luôn được dự đoán chính xác. Precision là 1.0 và Recall là 0,93. Điều này cho thấy rằng lớp 6 (dữ liệu bình thường) có các tính năng riêng biệt trong tập dữ liệu.

3.4.2 Kết quả khi dùng thuật toán Random Forest



Hình 3.4: Kết quả phân loại tấn công (Random Forest)

Qua kết quả ở trên ta có thể thấy Random Forest Classifier hoạt động khá tốt với dữ liệu này. Điểm số cũng sẽ được cải thiện hơn nữa sau khi giảm tính năng ở cùng một phân loại. Điểm Recall là 0.93 cho lớp 0 và 1.00 cho lớp 1. Đồng thời, Precision cho lớp 1 là 0,38 thấp hơn lớp 0 là 1.00.

- TP (True Positive): 96261
- FP (False Positive): 129
- TN (True Negative): 2027072
- FN (False Negative): 160384

Sau quy trình xử lý tập dữ liệu mẫu để thực hiện đánh giá hiệu quả khi sử dụng thuật toán học máy Random Forest có thể thấy rằng thuật toán Random Forest hoạt động hiệu quả khá tốt trong phân loại tấn công.

3.5 Kết luận chương 3

Chương 3 trình bày ứng dụng triển khai giải pháp phát hiện xâm nhập dựa trên mạng Nơ ron và thuật toán Random Forest. Trong chương 3 đã nêu ra mô hình phát hiện xâm nhập trong IoT gateway, các kiến trúc hệ thống phát hiện xâm nhập, tiến hành thiết lập thử nghiệm và đưa ra kết quả đánh giá cuối cùng. Các kết quả thử nghiệm cho thấy hệ thống IDS ứng dụng giải pháp mạng Nơ ron có thể đạt độ chính xác 90% và ứng dụng thuật toán *Random Forest* thì có được độ chính xác lên đến 93%.

KẾT LUẬN

Hệ thống phát hiện xâm nhập hiện đang là một trong những giải pháp được quan tâm hàng đầu hiện nay nhằm bảo vệ linh hoạt, hiệu quả trước vô vàn cuộc xâm nhập trái phép trên Internet nhắm tới các thiết bị IoT. Mặc dù còn gặp nhiều thách thức do các nguy cơ tấn công bảo mật đều khá phức tạp và khó có thể đoán trước được, các hệ thống phát hiện xâm nhập ứng dụng kỹ thuật học máy đã và đang cho thấy nhiều tiềm năng và thu hút được nhiều sự quan tâm, đầu tư và nghiên cứu. Trong tình hình đó, việc nghiên cứu, tìm hiểu và nắm bắt các giải pháp phát hiện xâm nhập hiệu quả cho các thiết bị IoT gateway dựa trên công nghệ học máy là rất cấp thiết.

Trong khuôn khổ luận văn này, học viên tập trung nghiên cứu vấn đề an toàn thông tin cho các thiết bị IoT gateway và các kỹ thuật phát hiện xâm nhập ứng dụng kỹ thuật học máy trong IoT, trên cơ sở đó, xây dựng và thử nghiệm giải pháp phát hiện xâm nhập sử dụng công nghệ học máy trong kịch bản ứng dụng cho các thiết bị IoT gateway. Các nội dung chính đạt được trong luận văn bao gồm:

- Nghiên cứu tổng quan về Internet of things, các thiết bị IoT Gateway, các kỹ thuật mà một hệ thống IDS truyền thống sử dụng để phát hiện xâm nhập cũng như lý thuyết về các thuật toán học máy ứng dụng trong phát hiện xâm nhập: KNN, SVM, *Naive Bayes, J48 Decision Tree*. Đặc biệt là thuật toán *Random Forest* và mạng Nơ ron.
- Nghiên cứu về thuật toán học máy ứng dụng tiếp cận trong phát hiện xâm nhập, đưa ra những giải pháp phát hiện xâm nhập ứng dụng cho IoT Gateway cũng như đề xuất mô hình ứng dụng học máy trong phát hiện xâm nhập. Đồng thời nghiên cứu về mô hình, kiến trúc của *Random Forest* và mạng Nơ ron nhằm áp dụng vào hệ thống phát hiện xâm nhập giúp cải thiện tỷ lệ phát hiện chính xác cũng như giảm thiểu tỷ lệ cảnh báo nhầm của một hệ thống IDS thông thường.
- Ứng dụng triển khai thử nghiệm giải pháp phát hiện xâm nhập sử dụng mạng Nơ ron và thuật toán *Random Forest* trên tập dữ liệu UNSW-NB15. Thông qua các bước xây dựng mô hình kiến trúc tổng thể và thiết lập thử nghiệm đã cho thấy hệ

thống IDS ứng dụng giải pháp mạng Nơ ron có thể đạt độ chính xác tương đối khoảng 90% và nếu ứng dụng thuật toán *Random Forest* thì có thể đạt độ chính xác lên đến 93%.

Tuy nhiên, vì thời gian nghiên cứu ngắn cũng như phạm vi của lĩnh vực nghiên cứu rộng, nên luận văn này vẫn còn nhiều vấn đề chưa đề cập và chưa giải quyết triệt để. Trong thời gian tới, em sẽ tiếp tục nghiên cứu thử nghiệm kết hợp với các giải pháp phù hợp vào hệ thống phát hiện xâm nhập và cố gắng đưa ứng dụng vào thực tiễn.