**TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG TP.HCM**

**KHOA: HỆ THỐNG THÔNG TIN VÀ VIỄN THÁM**



**BÁO CÁO ĐỒ ÁN MÔN: BẢO MẬT MẠNG MÁY TÍNH VÀ HỆ THỐNG**

**NGHIÊN CỨU CÁC THUẬT TOÁN MÁY HỌC ỨNG DỤNG TRONG PHÁT HIỆN TẤN CÔNG DDOS**

Giảng viên hướng dẫn: **ThS. Phạm Trọng Huynh**

Sinh viên thực hiện:  **Nguyễn Xuân Tiến Dũng 0850080012**

**Lý Tô Phong Phú 0850080095**

**Lê Vũ Tử Đang 0850080062**

Lớp**: CNPM**

Khoá**: 08**

***TP. Hồ Chí Minh, tháng 4 năm 2023***

**TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG TP.HCM**

**KHOA: HỆ THỐNG THÔNG TIN VÀ VIỄN THÁM**



**BÁO CÁO ĐỒ ÁN MÔN: BẢO MẬT MẠNG MÁY TÍNH VÀ HỆ THỐNG**

**NGHIÊN CỨU CÁC THUẬT TOÁN MÁY HỌC ỨNG DỤNG TRONG PHÁT HIỆN TẤN CÔNG DDOS**

Giảng viên hướng dẫn: **ThS. Phạm Trọng Huynh**

Sinh viên thực hiện:  **Nguyễn Xuân Tiến Dũng 0850080012**

**Lý Tô Phong Phú 0850080095**

**Lê Vũ Tử Đang 0850080062**

Lớp**: CNPM**

Khoá**: 08**

***TP. Hồ Chí Minh, tháng 4 năm 2023***

**MỞ ĐẦU**

Tấn công phủ đầu bằng dữ liệu (DDoS) là một trong những hình thức tấn công mạng phổ biến nhất hiện nay. Với sức mạnh của hàng trăm hoặc hàng nghìn thiết bị đầu cuối, các kẻ tấn công có thể tạo ra lưu lượng truy cập lớn và làm quá tải hệ thống đích, dẫn đến việc gián đoạn dịch vụ hoặc thậm chí là làm cho hệ thống bị sập hoàn toàn.

Trong thời đại của kỹ thuật số và các ứng dụng mạng ngày càng phổ biến, việc phát hiện và ngăn chặn các cuộc tấn công DDoS trở nên cực kỳ cần thiết. Một trong những phương pháp tiếp cận phổ biến nhất là sử dụng các thuật toán máy học để phát hiện và ngăn chặn các cuộc tấn công này.

Trong nghiên cứu này, nhóm chúng em sẽ tìm hiểu và áp dụng các thuật toán máy học phổ biến như Random Forest, Decision Tree, Naive Bayes, Support Vector Machine, K-Nearest Neighbor để phát hiện các cuộc tấn công DDoS. Nhóm chúng em sẽ tiến hành đánh giá hiệu suất của các thuật toán này dựa trên các tập dữ liệu được thu thập từ các cuộc tấn công DDoS thực tế.

Với các nghiên cứu và kết quả đạt được, nhóm chúng em hy vọng rằng đề tài này sẽ đóng góp một phần nhỏ vào việc tăng cường bảo mật cho các hệ thống mạng và giúp cho việc phát hiện các cuộc tấn công DDoS trở nên hiệu quả hơn.

**LỜI CẢM ƠN**

Trong suốt quá trình học tập và hoàn thành đồ án, nhóm em đã nhận được sự hướng dẫn, giúp đỡ, các kiến thức chuyên môn của thầy cô và các bạn đã giúp em có thể hoàn thành đồ án một cách tốt nhất.

Với lòng kính trọng và biết ơn sâu sắc em xin được bày tỏ lời cảm ơn chân thành tới:

Ban giám hiệu trường Đại học Tài nguyên và Môi trường thành phố Hồ Chí Minh, khoa Hệ thống thông tin và viễn thám, cùng các quý Thầy Cô giáo đã truyền đạt vốn tri thức và nhiệt huyết rất quý báu cho chúng em trong suốt thời gian học tập tại trường.

ThS. Phạm Trọng Huynh đã tận tâm hướng dẫn chúng em qua từng buổi học trên lớp và những kiến thức từ thực tiễn đã góp phần giúp em hoàn thành đồ án này cũng như đã nhiệt tình hướng dẫn em trong suốt quá trình thực hiện đồ án từ lúc sơ khai cho đến khi hoàn thành.

Sau cùng, em xin kính chúc quý Thầy Cô Bộ môn Hệ Thống Thông Tin và Viễn Thám thật dồi dào sức khỏe, niềm tin để tiếp tục thực hiện sứ mệnh cao đẹp của mình là truyền đạt kiến thức cho thế hệ mai sau

**NHẬN XÉT**

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………..

**Điểm**

**MỤC LỤC**

[Chương 1. TỔNG QUAN 1](#_Toc131840227)

[1.1. Định nghĩa vấn đề 1](#_Toc131840228)

[1.2. Phạm vi 1](#_Toc131840229)

[1.3. Mục tiêu đề tài 1](#_Toc131840230)

[1.4. Tính cấp thiết của đề tài 2](#_Toc131840231)

[1.5. Lý do chọn đề tài 3](#_Toc131840232)

[Chương 2. PHÂN TÍCH ĐỀ TÀI 4](#_Toc131840233)

[2.1. Phân tích yêu cầu 4](#_Toc131840234)

[2.2. Yêu cầu chức năng 4](#_Toc131840235)

[2.3. Yêu cầu phi chức năng 5](#_Toc131840236)

[2.4. Công việc cần giải quyết 5](#_Toc131840237)

[Chương 3. THIẾT KẾ 7](#_Toc131840238)

[3.1. Giới thiệu các thuật toán máy học 7](#_Toc131840239)

[3.1.1. Random Forest 7](#_Toc131840240)

[3.1.2. Decision Tree 7](#_Toc131840241)

[3.1.3. Naive Bayes 8](#_Toc131840242)

[3.1.4. Support Vector Machine 8](#_Toc131840243)

[3.1.5. K-Nearest Neighbor 9](#_Toc131840244)

[3.2. Các loại tấn công DDOS 9](#_Toc131840245)

[3.3. Công cụ thực nghiệm 11](#_Toc131840246)

[3.3.1. Mô hình mạng SDN 11](#_Toc131840247)

[3.3.2. Mininet 12](#_Toc131840248)

[3.3.3. Virtualenv 14](#_Toc131840249)

[3.3.4. Ryu Manager 14](#_Toc131840250)

[3.3.5. Hping3 14](#_Toc131840251)

[3.4. Mô hình mạng thử nghiệm 15](#_Toc131840252)

[Chương 4. HIỆN THỰC 16](#_Toc131840253)

[4.1. Khái quát sơ lược 16](#_Toc131840254)

[4.2. Thực hiện mô phỏng 19](#_Toc131840255)

[Chương 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 24](#_Toc131840256)

[5.1. Kết quả đạt được 24](#_Toc131840257)

[5.2. Hạn chế 24](#_Toc131840258)

[5.3. Hướng phát triển 24](#_Toc131840259)

[5.4. Kết luận 25](#_Toc131840260)

[DANH MỤC TÀI LIỆU THAM KHẢO 26](#_Toc131840261)

**DANH MỤC CÁC BẢNG, SƠ ĐỒ, HÌNH**

Hình 3. 1. Mô hình mạng SDN 12

Hình 3. 2. Công cụ mininet 13

Hình 3. 3. Mô hình mạng xây dựng thử nghiệm 15

[Hình 4. 1. Mô hình SVM bằng python 16](#_Toc131810049)

[Hình 4. 2. Lênh cmd cài đặt python3.8 17](#_Toc131810050)

[Hình 4. 3. Mô hình mạng xây dựng thử nghiệm 18](#_Toc131810051)

[Hình 4. 4. Python3.8 version 18](#_Toc131810052)

[Hình 4. 5. Ryu-manager install 18](#_Toc131810053)

[Hình 4. 6. Hình mô phỏng 1 19](#_Toc131810054)

[Hình 4. 7. Hình mô phỏng 2 19](#_Toc131810055)

[Hình 4. 8. Hình mô phong 3 20](#_Toc131810056)

[Hình 4. 9. Hình mô phong 4 20](#_Toc131810057)

[Hình 4. 10. Hình mô phỏng 5 21](#_Toc131810058)

[Hình 4. 11. Hình mô phỏng 6 21](#_Toc131810059)

[Hình 4. 12. Hình mô phỏng 7 22](#_Toc131810060)

[Hình 4. 13. Hình mô phỏng 8 22](#_Toc131810061)

[Hình 4. 14. Hình mô phỏng 9 23](#_Toc131810062)

**KÍ HIỆU CÁC CỤM TỪ VIẾT TẮT**

SVM: Support vector machine

K-NN: k-nearest neighbors algorithm

ANN: Artificial neural networks

RL:   Reinforcement Learning

DT: Decision Trees

DL: Deep learning

VT: Vectorization

TCP: Transmission Control Protocol

UDP: User Datagram Protocol

HTTP: Hypertext Transfer Protocol

SYN: Synchronize

1. TỔNG QUAN
   1. **Định nghĩa vấn đề**

Vấn đề trong đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS (Distributed Denial of Service) là việc phát hiện và ngăn chặn các cuộc tấn công mạng DDoS. Đây là một trong những vấn đề quan trọng trong lĩnh vực an ninh mạng vì tấn công DDoS có thể gây ra thiệt hại nghiêm trọng đến hệ thống mạng của các tổ chức và doanh nghiệp, từ việc làm gián đoạn hoạt động đến mất dữ liệu quan trọng hoặc thậm chí là việc truy cập trái phép vào các thông tin nhạy cảm.

Các thuật toán máy học được áp dụng để phát hiện các mẫu tấn công DDoS dựa trên các đặc trưng của lưu lượng mạng. Tuy nhiên, vấn đề đặt ra là làm thế nào để xác định các đặc trưng này và áp dụng các thuật toán phân loại hiệu quả để phát hiện các tấn công DDoS. Điều này đòi hỏi các nhà nghiên cứu phải tìm hiểu sâu về các kỹ thuật phân tích dữ liệu, các thuật toán máy học và cách áp dụng chúng vào bài toán phát hiện tấn công DDoS.

* 1. Phạm vi

Phạm vi trong đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS có thể bao gồm những nội dung sau:

Tìm hiểu về tấn công DDoS: Điều này bao gồm các loại tấn công DDoS phổ biến, cách thức tấn công và cách xác định được các đặc trưng của tấn công DDoS trên lưu lượng mạng.

Tìm hiểu về các thuật toán máy học: Điều này bao gồm các thuật toán phân loại, rừng ngẫu nhiên, học sâu, mạng thần kinh nhân tạo, SVM, k-NN, ...

Xây dựng mô hình máy học: Ứng dụng các thuật toán máy học để xây dựng mô hình phân loại và phát hiện tấn công DDoS trên lưu lượng mạng.

Thử nghiệm trên dữ liệu thực tế: Áp dụng mô hình phát hiện tấn công DDoS trên dữ liệu thực tế để kiểm tra tính hiệu quả và độ chính xác của thuật toán.

Vì vậy, phạm vi trong đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS có thể bao gồm các nội dung từ cơ bản đến nâng cao, tập trung vào việc áp dụng các thuật toán máy học để phát hiện tấn công DDoS trên lưu lượng mạng.

* 1. Mục tiêu đề tài

Mục tiêu trong đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS có thể bao gồm những mục tiêu sau:

Tìm hiểu và phân tích các đặc trưng của lưu lượng mạng: Mục tiêu này là để xác định các đặc trưng quan trọng của lưu lượng mạng, bao gồm thông tin về gói tin, các thông số của giao thức, số lượng gói tin, thời gian trễ, v.v. Điều này giúp xác định các đặc trưng để sử dụng trong việc phát hiện tấn công DDoS.

Áp dụng các thuật toán máy học để phát hiện tấn công DDoS: Mục tiêu này là phát triển các thuật toán máy học như SVM, K-NN, Random Forests, mạng thần kinh nhân tạo, học sâu, v.v. để phát hiện tấn công DDoS. Sử dụng các thuật toán này có thể giúp phát hiện và loại bỏ các tấn công DDoS trong thời gian thực.

Thử nghiệm trên dữ liệu thực tế: Mục tiêu này là áp dụng các thuật toán phát hiện tấn công DDoS đã phát triển trên dữ liệu thực tế để kiểm tra tính hiệu quả và độ chính xác của thuật toán.

Tóm lại, mục tiêu của đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS là phát triển các thuật toán máy học để phát hiện tấn công DDoS trên lưu lượng mạng và đánh giá hiệu suất của chúng. Ngoài ra, đề tài cũng nhằm xác định các đặc trưng quan trọng của lưu lượng mạng để sử dụng trong việc phát hiện tấn công DDoS.

* 1. Tính cấp thiết của đề tài

Đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS là rất cần thiết vì tấn công DDoS là một trong những mối đe dọa lớn đối với hệ thống mạng và các dịch vụ trực tuyến. Tấn công DDoS có thể gây ra sự cố hệ thống, gián đoạn dịch vụ và gây tổn thất về kinh tế. Hiện nay, các kỹ thuật tấn công DDoS ngày càng tinh vi và phức tạp hơn, khiến cho việc phát hiện và ngăn chặn tấn công trở nên khó khăn.

Trong khi đó, các thuật toán máy học có thể được sử dụng để phát hiện tấn công DDoS một cách hiệu quả. Các thuật toán máy học có khả năng học từ dữ liệu và tìm ra các mẫu tấn công DDoS, giúp phát hiện và ngăn chặn tấn công trước khi chúng gây ra hậu quả nghiêm trọng. Các thuật toán máy học cũng có thể giúp tăng tính khả chịu và hiệu suất của hệ thống mạng khi đối phó với các tấn công DDoS.

Do đó, đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS là cần thiết để giúp tăng cường bảo mật cho các hệ thống mạng và dịch vụ trực tuyến. Nghiên cứu này cũng có thể đóng góp vào việc phát triển các giải pháp phòng ngừa và xử lý tấn công DDoS trong tương lai.

* 1. Lý do chọn đề tài

Đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS là một lĩnh vực nghiên cứu đang được quan tâm và phát triển rộng rãi trong cộng đồng an ninh mạng. Tấn công DDoS là một trong những mối đe dọa lớn đối với hệ thống mạng và các dịch vụ trực tuyến, vì vậy việc tìm ra giải pháp hiệu quả để phát hiện và ngăn chặn tấn công DDoS là rất cần thiết.

Trong khi đó, các thuật toán máy học được coi là một trong những công cụ tiên tiến nhất để giải quyết các vấn đề phức tạp trong lĩnh vực an ninh mạng. Các thuật toán máy học có khả năng học từ dữ liệu và tìm ra các mẫu tấn công DDoS, giúp phát hiện và ngăn chặn tấn công trước khi chúng gây ra hậu quả nghiêm trọng.

Do đó, chọn đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS là một sự lựa chọn hợp lý. Nghiên cứu này có thể đóng góp vào việc tăng cường bảo mật cho các hệ thống mạng và dịch vụ trực tuyến, đồng thời giúp phát triển các giải pháp phòng ngừa và xử lý tấn công DDoS trong tương lai. Ngoài ra, đề tài cũng mang tính ứng dụng cao, có thể áp dụng được vào thực tế để cải thiện khả năng phát hiện và ngăn chặn tấn công DDoS của các hệ thống mạng.

1. PHÂN TÍCH ĐỀ TÀI
   1. Phân tích yêu cầu

Yêu cầu đề tài là mục tiêu nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS bao gồm:

Tìm hiểu về các phương pháp phát hiện tấn công DDoS: Nghiên cứu các phương pháp phát hiện tấn công DDoS hiện có để có thể đưa ra phương pháp mới phù hợp cho đề tài nghiên cứu.

Xây dựng tập dữ liệu: Thu thập dữ liệu về các cuộc tấn công DDoS, xây dựng tập dữ liệu và chuẩn bị dữ liệu cho việc huấn luyện thuật toán máy học.

Lựa chọn thuật toán máy học: Chọn thuật toán máy học phù hợp để phát hiện tấn công DDoS trên tập dữ liệu được xây dựng. Các thuật toán máy học có thể được chọn bao gồm mạng neuron nhân tạo (ANN), học tăng cường (RL), cây quyết định (DT), học sâu (DL), học tập phiên bản (VT) và máy vector hỗ trợ (SVM).

Huấn luyện và đánh giá thuật toán: Sử dụng tập dữ liệu đã xây dựng để huấn luyện thuật toán máy học, đánh giá và so sánh hiệu quả của các thuật toán máy học khác nhau.

Cải thiện hiệu quả của thuật toán: Cải thiện hiệu quả của thuật toán máy học bằng cách điều chỉnh tham số hoặc chọn các thuật toán máy học khác để phát hiện tấn công DDoS tốt hơn.

Đánh giá hiệu quả thực tế: Đánh giá hiệu quả thực tế của thuật toán máy học bằng cách triển khai thuật toán trên hệ thống thực tế để phát hiện các cuộc tấn công DDoS.

* 1. Yêu cầu chức năng

Yêu cầu chức năng đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS bao gồm:

Phát hiện tấn công DDoS: Yêu cầu chức năng chính của đề tài là phát hiện các cuộc tấn công DDoS trên hệ thống mạng. Thuật toán máy học được sử dụng để phát hiện các cuộc tấn công này.

Xác định loại tấn công DDoS: Ngoài việc phát hiện tấn công DDoS, đề tài cần có khả năng xác định loại tấn công DDoS. Các loại tấn công DDoS có thể bao gồm tấn công SYN Flood, tấn công UDP Flood, tấn công HTTP Flood, tấn công DNS Amplification, tấn công Smurf, tấn công Ping Flood và tấn công Slowloris.

Đưa ra cảnh báo: Khi phát hiện các cuộc tấn công DDoS, đề tài cần có khả năng đưa ra cảnh báo cho người quản trị hệ thống để người đó có thể xử lý vấn đề kịp thời.

Xử lý dữ liệu lớn: Hệ thống cần có khả năng xử lý dữ liệu lớn và phân tích các cuộc tấn công DDoS trên mạng lớn. Việc xử lý dữ liệu lớn cần được thực hiện một cách nhanh chóng và hiệu quả để giảm thiểu tác động của tấn công DDoS lên hệ thống.

Tích hợp vào hệ thống mạng: Thuật toán máy học cần được tích hợp vào hệ thống mạng để có thể phát hiện các cuộc tấn công DDoS và đưa ra cảnh báo kịp thời. Điều này đòi hỏi đề tài phải đáp ứng được yêu cầu về tích hợp hệ thống.

Tính linh hoạt: Đề tài cần đáp ứng yêu cầu về tính linh hoạt khi sử dụng các thuật toán máy học khác nhau để phát hiện tấn công DDoS. Hệ thống cần có khả năng sử dụng và thay đổi các thuật toán máy học tùy thuộc vào tính chất của các cuộc tấn công DDoS cụ thể.

* 1. Yêu cầu phi chức năng

Tính khả diễn giải: Các thuật toán máy học được áp dụng cần có tính khả diễn giải để người dùng có thể hiểu được quá trình phát hiện và đưa ra các quyết định phù hợp.

Hiệu suất và tốc độ xử lý: Các thuật toán phát hiện tấn công DDoS cần có tốc độ xử lý nhanh để có thể phát hiện các tấn công trong thời gian thực.

Độ chính xác: Các thuật toán cần đảm bảo độ chính xác cao để không phát hiện nhầm các gói tin hợp lệ là tấn công DDoS.

Khả năng học tập: Các thuật toán cần có khả năng học tập và tự động cập nhật để có thể phát hiện các loại tấn công mới.

Tính sẵn sàng và ổn định: Hệ thống cần có tính sẵn sàng cao và ổn định để đảm bảo phát hiện các tấn công DDoS trong mọi tình huống và không bị gián đoạn khi có sự cố xảy ra.

Tính bảo mật: Các thuật toán cần đảm bảo tính bảo mật để ngăn chặn các tấn công từ bên ngoài và tránh các trường hợp bị hack hoặc xâm nhập.

* 1. Công việc cần giải quyết

Các công việc cần giải quyết trong đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS có thể được phân chia thành các giai đoạn chính sau:

Tìm hiểu về các loại tấn công DDoS: Tìm hiểu các loại tấn công DDoS phổ biến như TCP SYN Flood, UDP Flood, HTTP Flood, và các biến thể khác để xác định các đặc điểm của chúng và cách thức tấn công.

Tìm hiểu về các thuật toán máy học: Tìm hiểu các thuật toán máy học phổ biến và các kỹ thuật khác như học sâu (deep learning), học tăng cường (reinforcement learning), học bán giám sát (semi-supervised learning), để đánh giá khả năng ứng dụng của chúng trong việc phát hiện tấn công DDoS.

Xây dựng tập dữ liệu huấn luyện: Thu thập dữ liệu về các tấn công DDoS từ các nguồn khác nhau và xây dựng tập dữ liệu huấn luyện cho các thuật toán máy học. Tập dữ liệu này phải được đánh giá kỹ càng để đảm bảo tính đầy đủ và đại diện.

Tiền xử lý dữ liệu: Xử lý dữ liệu để loại bỏ các nhiễu, chuẩn hóa dữ liệu, và trích xuất các đặc trưng có ý nghĩa cho việc phát hiện tấn công DDoS.

Xây dựng mô hình phát hiện: Áp dụng các thuật toán máy học đã tìm hiểu để xây dựng mô hình phát hiện tấn công DDoS. Mô hình này sẽ được huấn luyện trên tập dữ liệu đã được xử lý.

Đánh giá và tinh chỉnh mô hình: Đánh giá hiệu suất của mô hình trên các tập dữ liệu kiểm tra và điều chỉnh các siêu tham số để cải thiện hiệu suất của mô hình.

Đánh giá và so sánh hiệu suất với các phương pháp khác: Đánh giá hiệu suất của mô hình so với các phương pháp phát hiện tấn công DDoS khác, bao gồm các phương pháp truyền thống và các phương pháp dựa trên máy học khác.

1. THIẾT KẾ
   1. Giới thiệu các thuật toán máy học
      1. Random Forest

Random Forest là một thuật toán máy học kết hợp giữa việc áp dụng quyết định cây (Decision Tree) và kỹ thuật ensemble learning. Trong Random Forest, ta tạo ra nhiều cây quyết định (Decision Trees) bằng cách lấy ngẫu nhiên một phần của tập dữ liệu huấn luyện, sau đó kết hợp kết quả của các cây quyết định này để đưa ra dự đoán cuối cùng.

Các cây quyết định được xây dựng dựa trên tập dữ liệu huấn luyện được chọn ngẫu nhiên từ tập dữ liệu ban đầu. Khi phân loại một mẫu mới, các cây quyết định trong Random Forest sẽ đưa ra dự đoán và kết quả được xác định bằng cách lấy ý kiến đa số của các cây quyết định.

Random Forest có thể được sử dụng trong các bài toán phân loại, hồi quy và phát hiện bất thường. Trong đề tài nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDoS, Random Forest có thể được sử dụng để phân loại các gói tin mạng dựa trên các đặc trưng của chúng, giúp phát hiện các tấn công DDoS.

* + 1. Decision Tree

Decision Tree là một thuật toán máy học được sử dụng trong việc phân loại và dự đoán dữ liệu. Thuật toán này sử dụng cây quyết định để xây dựng mô hình dự đoán. Mô hình được xây dựng dựa trên việc phân tách các điểm dữ liệu thành các nhóm dựa trên các câu hỏi phân loại (được gọi là nút quyết định) và các điều kiện phân loại (được gọi là cạnh).

Cây quyết định được xây dựng bằng cách đặt các câu hỏi phân loại liên tiếp cho dữ liệu đầu vào và sử dụng các câu hỏi này để tìm ra cách phân loại tốt nhất cho mỗi điểm dữ liệu. Các câu hỏi phân loại được xây dựng bằng cách phân tích các thuộc tính của dữ liệu đầu vào, như số lượng truy cập, địa chỉ IP hoặc số lượng gói tin.

Một cây quyết định được xây dựng theo cách sao cho các nút quyết định phân chia dữ liệu thành các nhóm sao cho các điểm dữ liệu trong cùng một nhóm có sự giống nhau nhất định. Sau khi cây quyết định được xây dựng, mô hình có thể được sử dụng để phân loại các điểm dữ liệu mới dựa trên câu hỏi phân loại đã được xây dựng trong cây.

* + 1. Naive Bayes

Naive Bayes là một thuật toán máy học dựa trên lý thuyết xác suất. Nó được sử dụng để phân loại và dự đoán các giá trị mới dựa trên các bộ dữ liệu đã biết trước đó. Thuật toán Naïve Bayes giả định rằng các đặc trưng của dữ liệu là độc lập với nhau và có cùng đóng góp vào kết quả cuối cùng.

Để sử dụng thuật toán Naïve Bayes trong việc phát hiện tấn công DDoS, ta sử dụng các đặc trưng của lưu lượng mạng (như địa chỉ IP nguồn, đích, giao thức, cổng, thời gian, số lượng gói tin) làm đầu vào để phân loại lưu lượng mạng vào các nhóm tấn công DDoS hoặc không phải.

Thuật toán Naïve Bayes cần phải được huấn luyện trên các dữ liệu lưu lượng mạng được đánh nhãn để có thể xác định các xác suất của các lớp tấn công DDoS và không phải. Sau đó, nó có thể được sử dụng để dự đoán lớp của các lưu lượng mạng mới dựa trên các đặc trưng của chúng.

Một trong những ưu điểm của thuật toán Naïve Bayes là tốc độ tính toán nhanh và phù hợp cho các tập dữ liệu lớn. Tuy nhiên, thuật toán Naïve Bayes có giả định đặc trưng độc lập, do đó nó có thể không hoạt động tốt trên các bộ dữ liệu có các đặc trưng phụ thuộc lẫn nhau.

* + 1. Support Vector Machine

Support Vector Machine (SVM) là một thuật toán học máy giám sát được sử dụng cho các bài toán phân loại và hồi quy. Nó hoạt động bằng cách tìm ra một đường ranh giới (hyperplane) phân chia tốt nhất các điểm dữ liệu trong các lớp khác nhau.

Đối với bài toán phân loại, SVM cố gắng tìm một đường ranh giới sao cho khoảng cách từ những điểm gần ranh giới nhất đến đường này là lớn nhất. Khoảng cách này được gọi là margin. SVM cố gắng tìm ra đường ranh giới có margin lớn nhất để giảm thiểu sự lỗi của bài toán phân loại.

Nếu dữ liệu không thể được phân loại bằng một đường ranh giới duy nhất, SVM sử dụng một kỹ thuật gọi là kernel trick để chuyển đổi dữ liệu vào một không gian cao hơn, nơi các lớp dữ liệu có thể được phân loại bằng một đường ranh giới tuyến tính. Kỹ thuật này giúp SVM xử lý được các bài toán phân loại phi tuyến tính.

SVM là một trong những thuật toán phân loại hiệu quả nhất khi dữ liệu có số chiều cao và có tính tuyến tính. Tuy nhiên, đòi hỏi thời gian huấn luyện và dự đoán lớn hơn so với một số thuật toán khác, đặc biệt là khi kích thước của dữ liệu lớn.

* + 1. K-Nearest Neighbor

K-Nearest Neighbor (KNN) là một thuật toán học có giám sát trong máy học, được sử dụng trong các bài toán phân loại và dự đoán. KNN dựa trên nguyên tắc rằng các điểm dữ liệu cùng nhãn (label) thường có sự tương đồng lớn về mặt khoảng cách trong không gian đặc trưng.

KNN được gọi là thuật toán "lười" (lazy algorithm) vì nó không tiến hành bất kỳ tính toán nào để xây dựng mô hình từ dữ liệu đào tạo, mà chỉ lưu trữ các điểm dữ liệu này để sử dụng cho việc dự đoán các điểm mới. Khi có một điểm dữ liệu mới cần được phân loại, thuật toán sẽ tính khoảng cách từ điểm mới đến tất cả các điểm đã được lưu trữ trong tập dữ liệu, sau đó chọn ra k điểm gần nhất (k là một tham số được người dùng định nghĩa trước) để tính toán dự đoán.

Trong bài toán phát hiện tấn công DDoS, KNN có thể được sử dụng để phân loại các gói tin mạng vào các lớp khác nhau dựa trên các đặc trưng của chúng, như địa chỉ IP nguồn, địa chỉ IP đích, giao thức mạng, cổng nguồn, cổng đích, kích thước gói tin, thời gian giữa các gói tin liên tiếp và các đặc trưng khác. KNN có thể được sử dụng để xác định liệu một gói tin có phải là tấn công DDoS hay không, bằng cách so sánh khoảng cách của gói tin đó với các gói tin đã được phân loại trước đó và xác định lớp phù hợp nhất dựa trên k gói tin gần nhất.

* 1. Các loại tấn công DDOS

DDoS (Distributed Denial of Service) là một hình thức tấn công mạng được thực hiện bằng cách tạo ra lưu lượng truy cập không mong muốn đến một hệ thống hoặc mạng nhằm gây ra quá tải, làm gián đoạn hoặc ngừng hoạt động của hệ thống đó. DDoS có nghĩa là phân tán, điều này có nghĩa là các máy chủ tấn công không được tập trung tại một địa điểm, mà chúng được phân tán trên nhiều hệ thống khác nhau.

Có nhiều loại tấn công DDoS khác nhau, bao gồm:

1. Tấn công ICMP (Ping) Flooding là một loại tấn công DDoS được thực hiện bằng cách tạo ra một lượng lớn yêu cầu ping (ICMP Echo Request) gửi đến một địa chỉ IP nhất định, nhằm làm quá tải và làm cho máy chủ trở nên chậm chạp hoặc ngừng hoạt động. Khi một lượng lớn ICMP Echo Request được gửi đi, các máy chủ hoặc thiết bị đích phải xử lý số lượng lớn yêu cầu đó, gây ra một khối lượng lớn lưu lượng trên mạng. Do đó, các tài nguyên mạng sẽ bị đánh sập và dẫn đến một tình trạng mạng chậm hoặc không hoạt động.

2. Tấn công UDP (User Datagram Protocol) Flooding là một loại tấn công DDoS mà tập trung vào các port sử dụng giao thức UDP. Tấn công này được thực hiện bằng cách gửi một lượng lớn yêu cầu UDP tới các port mục tiêu, làm cho các port đó trở nên quá tải và không còn khả năng xử lý các yêu cầu tiếp theo. Khi máy chủ đích không thể xử lý yêu cầu UDP, các gói tin sẽ bị mất hoặc bị lọc và tài nguyên mạng sẽ bị quá tải.

Các tấn công UDP Flooding thường sử dụng các giao thức như DNS, SNMP và NTP để tấn công. Ví dụ, trong tấn công DNS Amplification, kẻ tấn công sẽ sử dụng các máy chủ DNS để phản hồi các yêu cầu UDP với lượng dữ liệu lớn hơn, khiến cho các máy chủ đích bị quá tải và bị đánh sập.

3. Tấn công SYN (Synchronize) Flooding là một trong những loại tấn công DDoS phổ biến nhất. Khi một máy tính muốn thiết lập một kết nối TCP với một máy tính khác, nó gửi một gói tin SYN để yêu cầu thiết lập kết nối. Máy tính nhận được yêu cầu sẽ gửi một gói tin SYN-ACK để xác nhận yêu cầu, và máy tính yêu cầu sẽ gửi một gói tin ACK để xác nhận lại. Trong tấn công SYN flooding, kẻ tấn công gửi một lượng lớn các gói tin SYN giả mạo đến máy tính đích, khiến nó phải đợi và tiêu tốn tài nguyên của nó để đợi các gói tin xác nhận. Nếu lượng gói tin SYN giả mạo đủ lớn, máy tính đích sẽ không còn đủ tài nguyên để tiếp tục kết nối, khiến cho nó không thể phục vụ yêu cầu của các máy tính khác và dẫn đến tình trạng quá tải hoặc sập hệ thống.

4. Tấn công HTTP Flood là một loại tấn công DDoS tập trung vào việc tấn công các máy chủ web bằng cách gửi các yêu cầu HTTP giả mạo từ nhiều nguồn khác nhau. Mục đích của tấn công này là làm cho máy chủ web không thể xử lý được tất cả các yêu cầu và khiến nó bị quá tải, từ đó dẫn đến việc trang web trở nên không thể truy cập được cho người dùng.

Tấn công HTTP Flood thường sử dụng botnet để thực hiện, với các bot được kiểm soát từ một máy chủ chủ để. Những bot này sẽ gửi yêu cầu giả mạo đến máy chủ web, giả vờ như chúng là các yêu cầu hợp lệ từ các nguồn khác nhau, để khiến máy chủ web quá tải và không thể phục vụ yêu cầu của người dùng thực sự.

5. Tấn công Slowloris là một loại tấn công khai thác lỗ hổng trên giao thức HTTP để gây ngập dòng kết nối tới một trang web hoặc máy chủ web. Tấn công này được thực hiện bằng cách giữ các kết nối mở với máy chủ bằng cách gửi các yêu cầu HTTP không hoàn chỉnh hoặc chậm chạp.

Thay vì gửi yêu cầu hoàn chỉnh đến máy chủ, Slowloris gửi các yêu cầu HTTP dần dần, mỗi yêu cầu chỉ chứa một phần của thông tin. Như vậy, các kết nối sẽ được giữ mở trong thời gian dài, khiến máy chủ không thể phục vụ các yêu cầu mới từ các kết nối khác. Với đủ lượng kết nối, tấn công Slowloris có thể gây ra tình trạng quá tải và làm cho máy chủ không hoạt động được.

6. Tấn công Application-layer là một dạng tấn công DDoS tập trung vào lớp ứng dụng của một hệ thống, thường là các ứng dụng web. Đây là loại tấn công khó phát hiện hơn so với các tấn công khác, vì chúng tập trung vào các lỗ hổng ứng dụng thay vì các lỗ hổng hạ tầng.

Một số ví dụ về tấn công Application-layer bao gồm:

SQL injection: tấn công này tập trung vào việc chèn các truy vấn SQL độc hại vào các trang web, với mục đích lấy thông tin từ cơ sở dữ liệu hoặc thực hiện các hành động độc hại khác.

Cross-site scripting (XSS): tấn công này cho phép tin tặc chèn mã độc vào trang web, với mục đích lấy thông tin từ người dùng hoặc thực hiện các hành động khác.

Distributed Reflected Denial of Service (DRDoS): tấn công này tập trung vào việc tận dụng các lỗ hổng ứng dụng để phát tán các gói tin tấn công đến các máy chủ bị lợi dụng để tạo ra một tấn công DDoS.

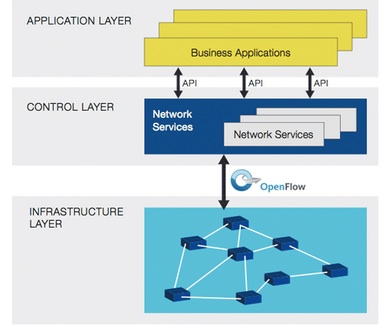
* 1. Công cụ thực nghiệm
     1. Mô hình mạng SDN

Software-Defined Network (SDN) là một kiến trúc mạng linh hoạt và điều khiển tập trung có thể lập trình được (programmed). Nó giúp cho người quản trị có thể dễ dàng quản lý được toàn bộ hệ thống mạng và triển khai và thi hành các chính sách dich vụ khác nhau.

Thông qua các giao diện mở APIs được cung cấp sẵn, SDN cho phép lập trình các hành vi mạng theo cách kiểm soát tập trung. Bằng cách mở ra một hạ tầng mạng truyền thống và triển khai SDN control layer, chúng ta có thể quản lý toàn bộ hệ thống mạng và các thiết bị một cách đồng thời mà không cần quan tâm đến mức độ phức tạp của hạ tầng bên dưới.

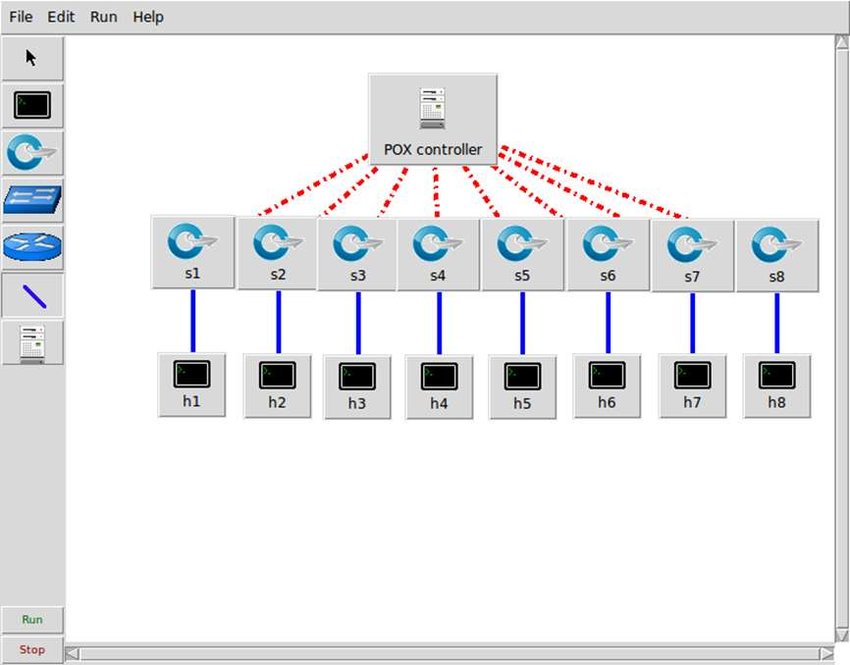
Kiến trúc và các thành phần cơ bản

* Application layyer (SDN App)
* Control plane
* Data plane



Hình 3. . Mô hình mạng SDN

* + 1. Mininet



Hình 3. . Công cụ mininet

Mininet là một công cụ giả lập mạng, bao gồm tập hợp các hosts đầu cuối, các switches, routers và các liên kết trên một Linux kernel. Mininet sử dụng công nghệ ảo hóa (ở mức đơn giản) để tạo nên hệ thống mạng hoàn chỉnh, chạy chung trên cùng một kernel, hệ thống và user code.

Các host ảo, switch, liên kết và các controller trên mininet là các thực thể thực sự, được giả lập dưới dạng phần mềm thay vì phần cứng. Một host mininet có thể thực hiện ssh vào đó, chạy bất kì phần mềm nào đã cài trên hệ thống linux (môi trường mà mininet đang chạy). Các phần mềm này có thể gửi gói tin thông các ethernet interface của mininet với tốc độ liên kết và trễ đặt trước.

Mininet cho phép tạo topo mạng nhanh chóng, tùy chỉnh được topo mạng, chạy được các phần mềm thực sự như web servers, TCP monitoring, Wireshark; tùy chỉnh được việc chuyển tiếp gói tin. Mininet cũng dễ dàng sử dụng và không yêu cầu cấu hình đặc biệt gì về phần cứng để chạy: mininet có thể cài trên laptop, server, VM, cloud (linux).

* + 1. Virtualenv

Virtualenv là một phần mềm viết bằng Python, giúp tạo ra các "môi trường ảo" để tách biệt các môi trường của các project khác nhau.

Mỗi "môi trường ảo" về bản chất chỉ là một thư mục, với một vài thay đổi khiến cho khi ta dùng pip để cài package, các package sẽ được cài vào thư mục này. Điều này giải quyết các vấn đề nói trên:

tách biệt các môi trường giữa các project, 2 project có thể dùng 2 phiên bản khác nhau của cùng 1 phần mềm.

người dùng có thể tạo môi trường ảo trong thư mục $HOME của mình, không cần chạy sudo - không ảnh hưởng đến hệ thống.

* + 1. Ryu Manager

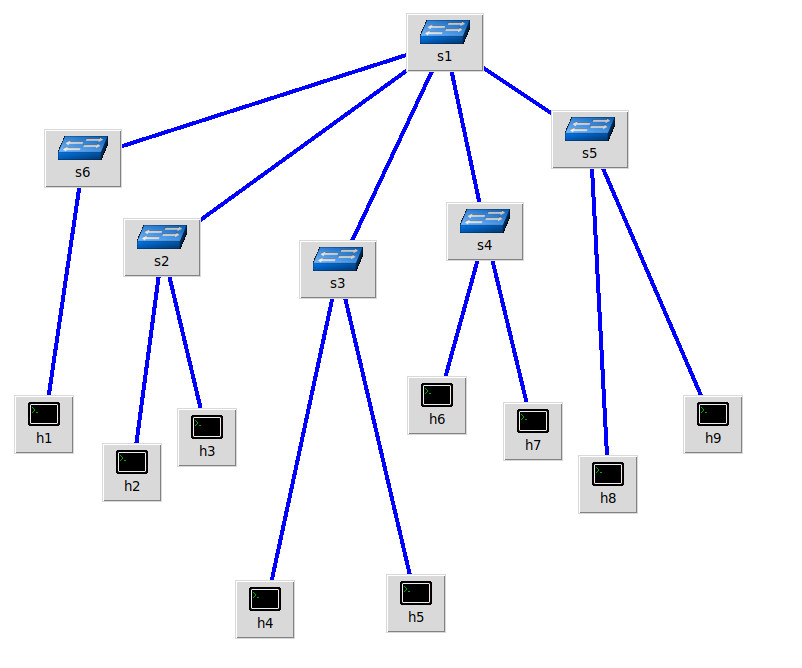
Ryu Manager là một phần mềm điều khiển mạng mã nguồn mở được phát triển bởi dự án Ryu. Nó cho phép quản trị mạng tùy chỉnh và điều khiển mạng bằng cách sử dụng các giao thức mạng như OpenFlow. Ryu Manager được viết bằng Python và có thể được sử dụng để phát triển các ứng dụng điều khiển mạng, bao gồm cả các ứng dụng SDN (Software-Defined Networking). Nó cũng cung cấp các tính năng như lập lịch và quản lý lưu lượng mạng để giúp tối ưu hóa hiệu suất mạng.

* + 1. Hping3

hping3 là một công cụ phân tích và kiểm tra gói tin mạng mã nguồn mở, được sử dụng để kiểm tra tính bảo mật của hệ thống mạng. Hping3 cung cấp nhiều tính năng như tạo gói tin tùy chỉnh, phân tích cấu trúc gói tin, xác định các cổng mạng đang mở, kiểm tra các cơ chế bảo mật của hệ thống mạng, v.v. Nó được sử dụng rộng rãi trong việc kiểm tra tính bảo mật của các hệ thống mạng và là một trong những công cụ phân tích mạng được ưa chuộng nhất.

* 1. Mô hình mạng thử nghiệm

 Xây dựng mô hình mạng SDN bằng công cụ Mininet để thử nghiệm thuật toán phát hiện DDOS như sau:



Hình 3. . Mô hình mạng xây dựng thử nghiệm

1. HIỆN THỰC
   1. Khái quát sơ lược

Thực hiện việc thử nghiệm phát hiện DDOS bằng thuật toán máy học sử dụng các công cụ như sau:

Mô hình đề xuất phát hiện tấn công DDOS

Mô hình SVM (Support Vector Machine) được đề xuất để phát hiện tấn công DDoS vì nó có nhiều ưu điểm như sau:

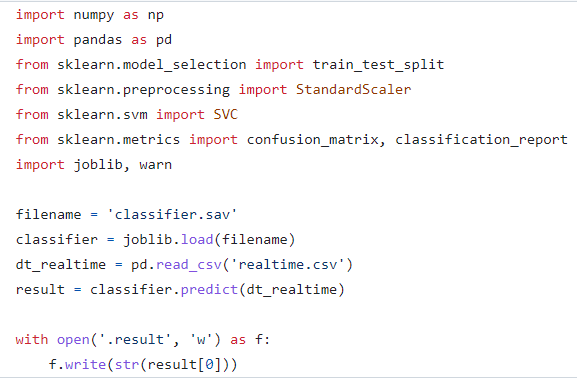
Hiệu quả tính toán: SVM sử dụng phương pháp tối ưu hóa toán học để xác định đường ranh giới giữa các lớp dữ liệu. Điều này làm cho SVM trở thành một phương pháp học máy nhanh và hiệu quả tính toán.

Độ chính xác cao: SVM có độ chính xác cao trong việc phân loại các mẫu dữ liệu, đặc biệt là trong trường hợp dữ liệu là tách biệt tuyến tính.

Tính tổng quát cao: SVM có khả năng xử lý dữ liệu phi tuyến tính, giúp nó trở thành một phương pháp học máy tổng quát và có thể được sử dụng trong nhiều ứng dụng khác nhau.

Khả năng xử lý dữ liệu lớn: SVM có khả năng xử lý dữ liệu lớn, vì nó chỉ phải lưu trữ các vector hỗ trợ (support vectors) thay vì toàn bộ tập dữ liệu.

Xây dựng mô hình SVM bằng python:



Hình 4. . Mô hình SVM bằng python

Đoạn code này được sử dụng để sử dụng mô hình máy học để dự đoán kết quả dữ liệu thời gian thực từ tệp "realtime.csv" và lưu trữ kết quả dự đoán vào tệp ".result".

Cụ thể, code này sử dụng thư viện NumPy, pandas và Scikit-learn để đọc tệp dữ liệu và tạo mô hình học máy SVC (Support Vector Machine) với các tham số được tối ưu hóa. Sau đó, mô hình được tải từ tệp đã được huấn luyện trước đó ('classifier.sav') bằng cách sử dụng hàm joblib.load ().

Sau đó, dữ liệu thời gian thực được đọc từ tệp 'realtime.csv' và được sử dụng để dự đoán kết quả sử dụng phương thức predict () của mô hình đã được tải.

Kết quả dự đoán được lưu trữ vào tệp '.result' bằng cách sử dụng hàm write () để ghi vào tệp với định dạng chuỗi.

Chú thích các tệp thực thi:

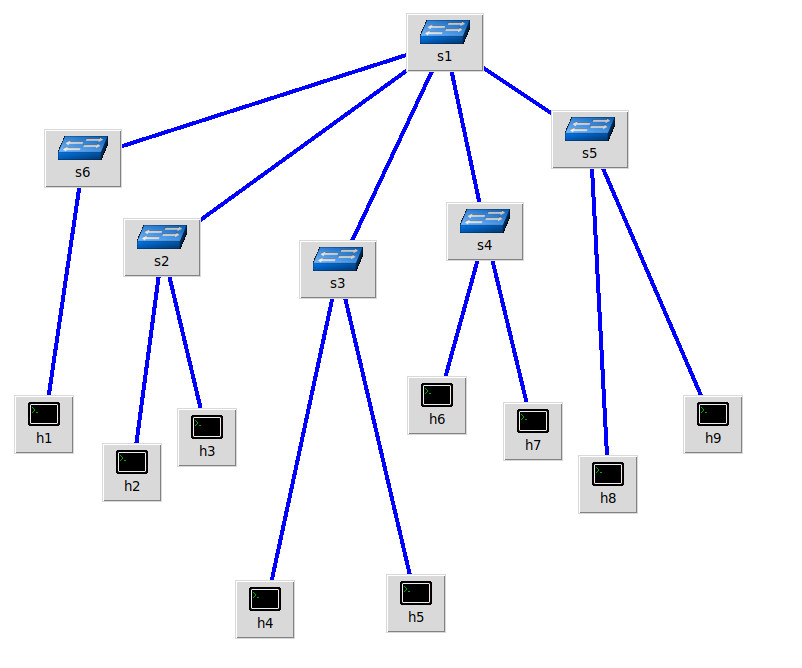
* .result: thể hiện kết quả phân loại từ mô hình, đúng hay sai rằng hệ thống đang bị tấn công DDoS.
* gentraffic.sh: tạo ra lưu lượng truy cập bình thường.
* warn.py: bỏ qua cảnh báo do không dùng nữa.
* topo.py: cấu trúc liên kết mininet.
* realtime.csv: tệp csv chứa 5 giá trị đặc trưng.
* inspector: gọi mô hình để phân loại các giá trị đặc trưng đã cho.
* customCtrl.py: bộ điều khiển Ryu tùy chỉnh.
* computeTuples.py: tính toán 5 đặc trưng từ dữ liệu thô.
* collect.sh: thu thập các bản ghi trong các flow table từ openflow switch để xử lý và trích xuất dữ liệu thô.

Cài đặt môi trường ảo bằng công cụ venv của python



Hình 4. . Lênh cmd cài đặt python3.8

Xây dựng hệ thống mạng ảo bằng công cụ Mininet



Hình 4. . Mô hình mạng xây dựng thử nghiệm

* Chạy file dữ liệu mạng ổn định từ h2 port 10.0.0.2 tới h1 port 10.0.0.1 thông qua router s2 và s6.
* Thực hiện tấn công DDOS từ cổng mạng h2 với cổng port 10.0.0.2 thông qua router s2 tấn công tới router s6 vào h1 với cổng port 10.0.0.1

Ngôn ngữ lập trình: Python3.8



Hình 4. . Python3.8 version

Sử dụng công cụ ryu-manager để bắt các gói tin



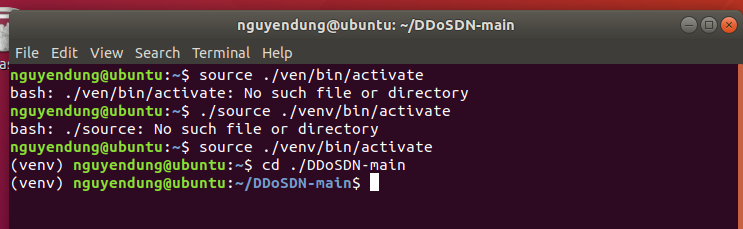
Hình 4. . Ryu-manager install

Môi trường máy áo VMWare với phiên bản linux 18.04

Tấn công DDOS bằng công cụ hping3

* Thực hiện lệnh tấn công thông qua công cụ hping3
* Sử dụng cơ chế tấn công syn flood.
  1. Thực hiện mô phỏng

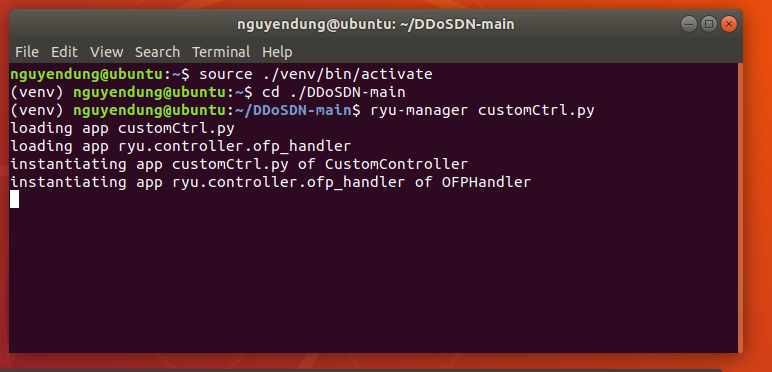
Bước 1: mở cmd trong linux và truy cập vào môi trường ảo venv của python



Hình 4. . Hình mô phỏng 1

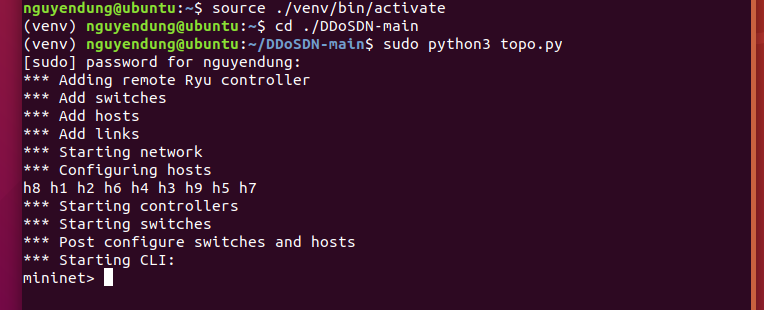
+ Thực hiện mở 3 cửa sổ tương tư và truy cập vào môi trường ảo venv

Bước 2: Cửa sổ thứ 2 ta mở tệp customCtrl.py bằng ryu-manager để thực thi ryu controller bắt các gói tin kết nối của mô hình mạng SDN.



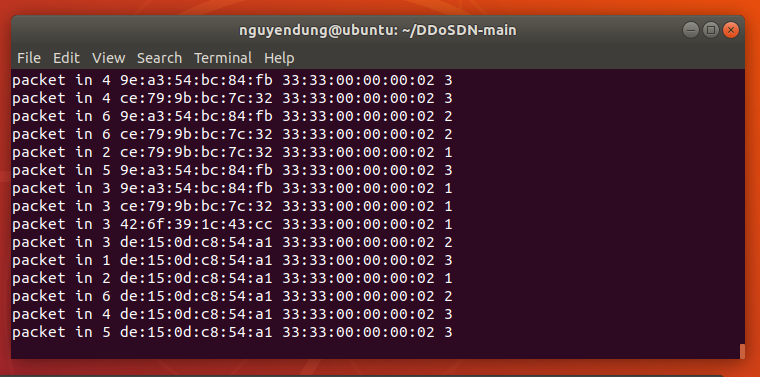
Hình 4. . Hình mô phỏng 2

Bước 3: Cửa sổ thứ 1 chạy file cấu hình mininet bằng python để khổi động mô hình mạng SDN.



Hình 4. . Hình mô phong 3

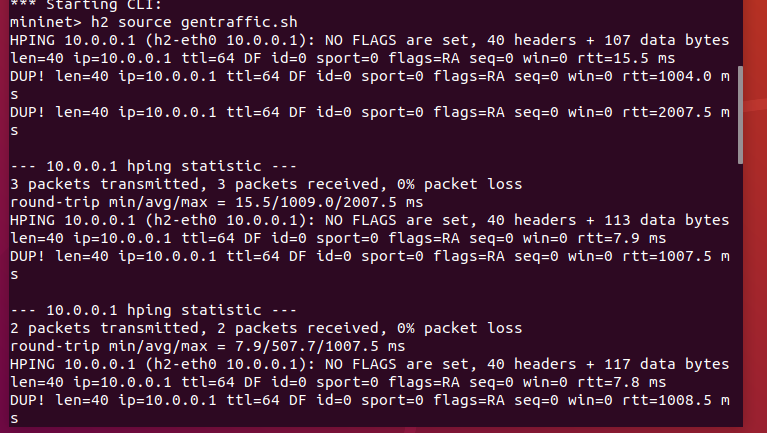
Kết quả bên cửa sổ thứ 2 của ryu controller dò được các kết nối của mạng



Hình 4. . Hình mô phong 4

Bước 4: thực hiện gửi dữ liệu mạng ổn định bằng cách chạy file gentraffic.sh trong file là các lệnh ping tới cổng h1 với port 10.0.0.1 từ cửa sổ 1 đồng thời khởi chạy mô hình SVM bằng file collect.sh là file với nhiệm vụ lấy dữ liệu và đưa xuống mô hình xử lý từ cửa sổ 3.

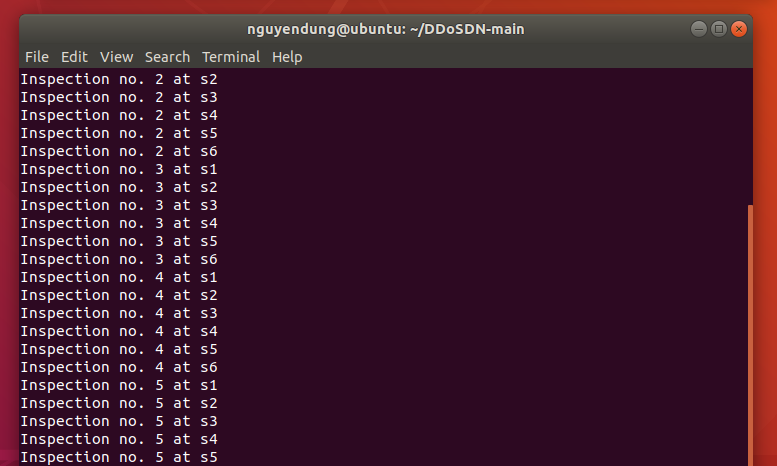
* Cửa sổ 1



Hình 4. . Hình mô phỏng 5

* Cửa sổ 3

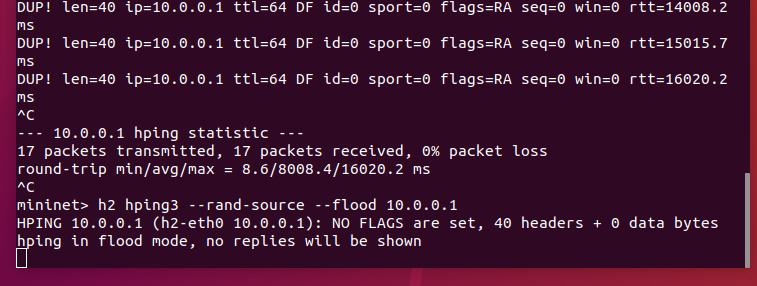
+Sau khi chạy file sẽ được cửa sổ thứ 3 hiển thị các truy cập tới các cổng



Hình 4. . Hình mô phỏng 6

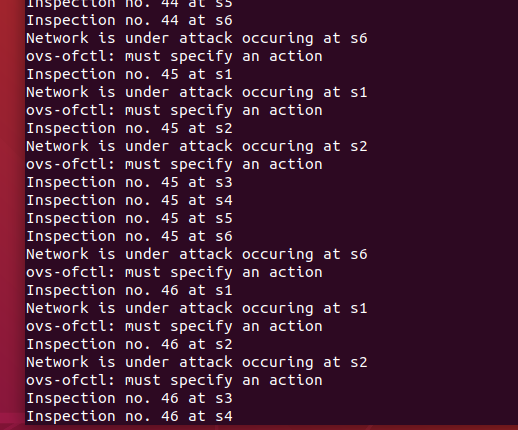
Bước 5: thực hiện tấn công DDOS bằng hping3 với cơ chế tấn công là syn flood mục tiêu tấn công là cổng h1 với port 10.0.0.1.

* Cửa sổ thứ 1: tấn công syn flood DDOS với lệnh sau



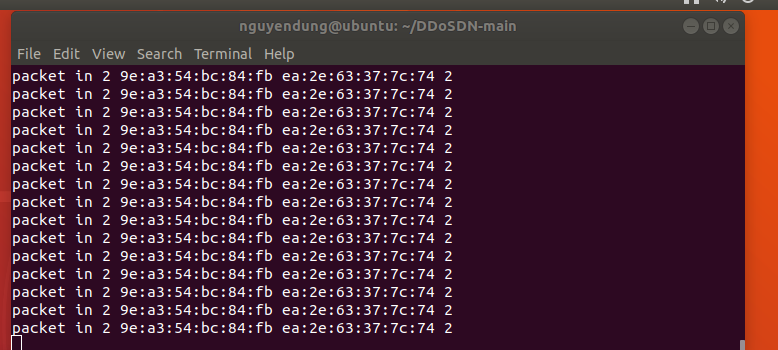
Hình 4. . Hình mô phỏng 7

Cửa sổ thứ 3: lập tức ta nhận được các thông báo là “under attack occuring at s6, s1, s2”



Hình 4. . Hình mô phỏng 8

* Cửa sổ thứ 2: Bắt được hàng loạt gói tin với lưu lượng mạnh truyền tải nhanh chóng so với tệp dữ liệu truyền tải mạng ổn định.



Hình 4. . Hình mô phỏng 9

* Mục tiêu tấn công DDOS là h1 và sử dụng h2 là chung gian điều khiển gừi hàng loạt các gói tin tới h1 thông qua các router s6, s2 ,s1 sau khi thực hiện huấn luyện các tệp dữ liệu mạng ổn định từ mô hình SVM và sau đó là tấn công bằng công cụ hping3 thông qua h2 để mô hình SVM có thể phân loại được tệp dữ liệu ổn định và đang bị tấn công ngay lập tức hệ thống sẽ thông báo vị trí cổng tấn công trên lưu lượng mạng trong đó ryu controller đóng vai trò là cầu nối gửi các gói tin thông qua hệ thống mạng SDN.

1. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN
   1. Kết quả đạt được

Nghiên cứu và xây dựng thành công thuật toán máy để phát hiện tấn công DDOS trên mô hình mạng SDN.

Mô hình thuật toán SVM có khả năng phân biệt được hai tệp dữ liệu từ dữ liệu mạng ổn định và dữ liệu mạng bị tấn công.

Hệ thống hiển thị và tìm kiếm được vị trí mạng bị tấn công trên mô hình SDN.

* 1. Hạn chế

Hệ thống chỉ ở mức thử nghiệm chưa thể áp dụng vào mô hình mạng thực tế.

Mô hình SVM còn nhiều khuyết điểm như:

* Nhạy cảm với nhiễu dữ liệu: SVM có thể bị ảnh hưởng bởi nhiễu dữ liệu và dữ liệu ngoại lai, làm giảm độ chính xác của mô hình.
* Khó cài đặt: SVM có nhiều tham số cần được thiết lập, và việc cài đặt các tham số này có thể khó khăn đối với những người mới bắt đầu với phương pháp học máy.
* Không phù hợp cho dữ liệu không tách biệt tuyến tính: Trong trường hợp dữ liệu không tách biệt tuyến tính, SVM có thể không hoạt động hiệu quả và cần được kết hợp với các phương pháp học máy khác để tăng độ chính xác.

Mô hình mạng SDN trong nghiên cứu còn thô sơ chưa có sự phức tạp và tương ứng với các mô hình mạng ngoài thực tế.

* 1. Hướng phát triển

Mở rộng phạm vi nghiên cứu: Nghiên cứu sâu hơn các loại tấn công DDoS khác nhau và các phương pháp tấn công mới để hiểu rõ hơn về cách thức hoạt động của chúng. Đồng thời, xây dựng các thuật toán máy học phù hợp để phát hiện các loại tấn công này.

Sử dụng các kỹ thuật mới: Nghiên cứu các kỹ thuật mới trong lĩnh vực machine learning như deep learning, reinforcement learning, và transfer learning để cải thiện độ chính xác của các thuật toán phát hiện tấn công DDoS.

Sử dụng các công cụ hỗ trợ nghiên cứu: Sử dụng các công cụ phân tích dữ liệu, trực quan hóa dữ liệu và các công cụ hỗ trợ machine learning như Python, Tensorflow, Keras, và các thư viện mã nguồn mở khác để tăng hiệu quả nghiên cứu.

Nghiên cứu cách phát hiện tấn công DDoS trên các mạng lớn: Nghiên cứu các phương pháp phát hiện tấn công DDoS trên các mạng lớn và phát triển các thuật toán máy học phù hợp để đáp ứng nhu cầu phát hiện tấn công DDoS trên các mạng lớn.

Kết hợp các phương pháp phát hiện: Kết hợp các thuật toán máy học khác nhau để cải thiện độ chính xác và giảm sai số khi phát hiện tấn công DDoS.

Nghiên cứu các phương pháp phòng chống tấn công DDoS: Nghiên cứu và phát triển các phương pháp phòng chống tấn công DDoS dựa trên các thuật toán máy học.

Ứng dụng thực tế: Áp dụng các thuật toán máy học phát hiện tấn công DDoS vào các hệ thống thực tế, đánh giá hiệu quả của các thuật toán và đưa ra các giải pháp phòng chống tấn công DDoS phù hợp.

* 1. Kết luận

Qua quá trình thực hành môn Bảo Mật Mạng Máy Tính và Hệ Thống và thực hành trên đề tài “**Nghiên cứu các thuật toán máy học ứng dụng trong phát hiện tấn công DDOS**”. Đã giúp cho nhóm em có thêm kiến thức cũng như việc thực hành thực tế trong quá trình bảo mật hệ thống với các cách thức tấn công DDOS nguy hiểm và phức tạp ngoài ra còn được nghiên cứu và tìm hiểu thêm về các thuật toán máy học vô cùng hiện đại và tiên tiến với các thuật toán học máy phức tạp và cực kỳ mạnh mẽ trong việc phát hiện tấn công DDOS, biết và tìm hiểu thêm về quá trình hoạt động cũng như việc triển khai trí tuệ nhân tạo áp dụng vào việc phát hiển tấn công DDOS trên hệ thống mô hình mạng SDN.

DANH MỤC TÀI LIỆU THAM KHẢO

**Tiếng việt**

[1] Nghiên cứu phương pháp phát hiện sớm xâm nhập bất thường mạng DDOS dựa trên các thuật toán học máy, Lê Hoàng Hiệp, Lê Xuân Hiếu, Hồ Thị Tuyền và Dương Thị Quý, Trường Đại học Công nghệ thông tin và Truyền thông, ĐH Thái Nguyên, 2022.

[9] Hệ thống phát hiện xâm nhập hai tầng cho các mạng IoT sử dụng máy học, Minh Tuấn Thái, Hoàng Hảo Phạm, Thanh Nam Trần, 2022.

[5] So sánh các thuật toán học máy trong phát hiện tấn công DdoS, Nguyễn Thị Khánh Trâm và TS. Đoàn Trung Sơn, Đại học Công Nghệ, Đại học quốc gia Hà Nội, 2022.

[2] Mô hình phát hiện tấn công DDOS sử dụng Machine learning, Võ Hồ Thu Sang, Nguyễn Đức Nhuận và Phan Hoàng Hải, Trường đại học Sư phạm, Đại học Huế,2021.

[8] Đánh giá các kỹ thuật học máy trong bài toán phát hiện tập tin thực thi độc hại, Nguyễn Văn Căn, Lê Văn Giang, Đoàn Ngọc Tú, Lương Khắc Định, Trường Đại học Kỹ Thuật, Hậu cần CAND, 2020.

[6] Một cách tiếp cận ứng dụng học sâu trong bài toán phát hiện tấn công máy chủ website, Đoàn Ngọc Tú, Nguyễn Văn Căn, Hoàng Việt Long, Lê Văn Giang, Lương Khác Định. Trường Đại học Hạ Long Uông Bí, Quảng Ninh, 2020.

[3] Đánh giá một số thuật toán học máy ứng dụng trong triển khai hệ thống phát hiện tấn công website, Mai Cường Thọ và Đỗ Văn Tuấn, khoa công nghệ thông tin trường Đại học Nha Trang, 2019.

[7] Phát Hiện BOTNET Dựa Trên Học Máy, Nguyễn Trọng Hưng, Hoàng Xuân Dậu, Vũ Xuân Hạnh, Khoa Công nghệ và An ninh thông tin, Học Viện An Ninh Nhân Dân, 2018.

[4] Ứng dụng kỹ thuật học máy vào phát hiện mã độc, Võ Văn Trường, Trường Đại học công nghệ, Đại học Quốc Gia Hà Nội, 2016.

[10] Nghiên cứu phát hiện tấn công DDOS dựa trên IP ENTROPY, Trần Quang Hưng, Học Viện Công Nghệ Bưu Chính Viễn Thông, 2016.

**Tiếng anh**

[1] Saied, et al (2015). Detection of known and unknown DDoS attacks using Artificial Neural Networks, Neurocomputing http://dx.doi.org/10.1016/j.neucom.2015.04.101i.

[2] Andrew W. Moore and ndrew W. Moore (2005). Internet Traffic Classification Using Bayesian Analysis Techniques; SIGMETRICS’05. 0.9 0.92 0.94 0.96 0.98 1 COR\_RF PCA\_RF COM\_RF Accuracy Precision Recall F1-Score AUC MÔ HÌNH PHÁT HIỆN TẤN CÔNG DDOS SỬ DỤNG MACHINE LEARNING 171.

[3] Jungtaek Seo1, Cheolho Lee1, Taeshik Shon2, Kyu-Hyung Cho2 (2005), A New DDoS Detection Model Using Multiple SVMs and TRA, International Federation for Information Processing, LNCS 3823, pp. 976 – 985.

[4] Kokila RT', Thamarai Selvi, Kannan Govindarajan (2014). DDoS Detection and Analysis in SDN-based Environment Using Support Vector Machine Classifier; Sixth Interational Conference on Advanced Computingv(ICoAC).

[5] LuanPM –Adminvietnam (2015), Kiến thức về DDOS, https://adminvietnam.org/kienthuc-ve-ddos-phan-2-phan-loai/1031/.

[6] Manjula Suresh and R. Anitha (2011). Evaluating Machine Learning Algorithms for Detecting DDoS Attacks; Springer-Verlag Berlin Heidelberg, CNSA 2011, CCIS 196, pp. 441–452.

[7] Marwane Zekri and Youssef Saadi (2017). DDoS attack detection using machine learning techniques in cloud computing environments; IEEE.

[8] Mohamed Idhammad, Karim Afdel and Mustapha Belouch (2018). Semi-supervised machine learning, Springer Science+Business Media, LLC, part of Springer Nature.

[9] Mohamed Idhammad, Karim Afdel(2017). DoS Detection Method based on Artificial Neural Networks, (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 8, No. 4.

[10] Naveen Bindraa, and Manu Sood (2019). Detecting DDoS Attacks Using Machine Learning Techniques and Contemporary Intrusion Detection Dataset, Automatic Control and Computer Sciences, Vol. 53, No. 5, pp. 419–428.