HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Ngô Văn Nhận

NGHIÊN CỨU PHƯƠNG PHÁP PHÁT HIỆN TẤN CÔNG DDOS TRONG MẠNG DI ĐỘNG 5G

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI - 2025

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Ngô Văn Nhận

NGHIÊN CỨU PHƯƠNG PHÁP PHÁT HIỆN TẤN CÔNG DDOS TRONG MẠNG DI ĐỘNG 5G

CHUYÊN NGÀNH : HỆ THỐNG THÔNG TIN
MÃ SỐ: 8.48.01.04

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. NGUYỄN ĐÌNH HÓA

HÀ NỘI – 2025

LỜI CAM ĐOAN

Học viên Ngô Văn Nhận, mã học viên B23CHIS054 xin cam đoan đề án tốt nghiệp là công trình nghiên cứu của riêng học viên dưới sự hướng dẫn của TS Nguyễn Đình Hóa. Tất cả những tham khảo trong đề án tốt nghiệp bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo.

Hà Nội, ngày 03 tháng 06 năm 2025

Tác giả đề án tốt nghiệp

(Ký và ghi rõ họ tên)

Ngô Văn Nhận

LỜI CẢM ƠN

Em xin chân thành cảm ơn giảng viên hướng dẫn TS Nguyễn Đình Hóa đã giúp đỡ và định hướng cho em trong suốt quá trình học tập và thực hiện đề án tốt nghiệp.

Dưới sự hướng dẫn của TS Nguyễn Đình Hóa, em đã cố gắng hoàn thành tốt nhất có thể đề án tốt nghiệp này, tuy nhiên trong quá trình thực hiện không thể tránh được những thiếu sót, em rất mong nhận được sự góp ý của các thầy/cô trong Hội đồng để em hoàn thiện hơn đề án tốt nghiệp này.

Em xin chân thành cảm ơn!

Hà Nội, ngày 03 tháng 06 năm 2025

Học viên thực hiện

Ngô Văn Nhận

MỤC LỤC

DANH MỤC KÝ HIỆU, CHỮ VIẾT TẮT

Ký hiệu	Tiếng Anh	Tiếng Việt
DDoS	Distributed Denial-of-Service	Tấn công từ chối dịch vụ phân tán
MitM	Man-in-the-Middle	Tấn công người ở giữa
RAN	Radio Access Network	Mạng truy cập vô tuyến
D2D	Device to Device	Giao tiếp giữa thiết bị với thiết bị
MIMO	Multiple Input Multiple Output	Đầu vào đa luồng đầu ra đa luồng
mMTC	Massive machine-type communication	Truyền thông kiểu máy quy mô lớn
SDN	Software Defined Network	Mạng được xác định bằng phần mềm
NFV	Network Function Virtualization	Phương pháp ảo hóa các chức năng mạng
5G PPP	5G-Public Private Partnership	Đối tác công tư về cơ sở hạ tầng 5G, cung cấp các giải pháp, kiến trúc, công nghệ và tiêu chuẩn cho cơ sở hạ tầng truyền thông 5G
ML	Machine Learning	Học máy
SVM	Support Vector Machine	Máy vectơ hỗ trợ
IDS	Intrusion Detection System	Hệ thống phát hiện xâm nhập
AMF	Access and Mobility Management Function	Chức năng quản lý truy cập và tính di động
SMF	Session Management Function	Chức năng quản lý phiên
RFE	Recursive Feature Elimination	Loại bỏ tính năng đệ quy
SMOTE	Synthetic Minority Over-sampling Technique	Kỹ thuật lấy mẫu quá mức thiểu số tổng hợp
5G-NIDD	5G Network Intrusion Detection Dataset	Bộ dữ liệu phát hiện xâm nhập trong mạng 5G

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG DỮ LIỆU

MỞ ĐẦU

1. Lý do chọn đề tài

Trong thời đại hiện nay, khi công nghệ thông tin và truyền thông phát triển mạnh mẽ, khái niệm Mạng di động thế hệ mới, mạng 4G, 5G ngày càng trở lên quan trọng và phổ biến. Mạng di động thế hệ mới 4G, 5G mang đến rất nhiều lợi ích và là yêu cầu tất yếu trong công cuộc Chuyển đổi số và sự phát triển mạnh mẽ của khoa học công nghệ hiện đại. Mạng di động 5G mang lại rất nhiều lợi ích so với thế hệ trước (4G), tốc độ vượt trội, có thể gấp 10 lần mạng 4G, đỗ trễ thấp, hỗ trợ nhiều thiết bị... [1] Với những ưu điểm này, mạng di động 5G có thể đáp ứng các yêu cầu của 4 công nghệ đột phá sau: công nghệ điện toán đám mây (Cloud computing), dữ liệu lớn (Big data), internet vạn vật (IoT) và trí tuệ nhân tạo (AI).

Từ ngày 15/10/2024, các nhà mạng di động tại Việt Nam đã dừng cung cấp dịch vụ mạng 2G. Người dân được thúc đẩy tiếp cận thiết bị 4G và tiến tới 5G trong tương lai, có thể truy cập Internet tốc độ cao, từ đó làm quen với dịch vụ số, dịch vụ công trực tuyến. Trong thực tế, việc dừng công nghệ 2G được đánh giá mang lại rất nhiều giá trị tại Việt Nam, các nhà mạng giảm gánh nặng về hạ tầng, có thêm nguồn lực để nghiên cứu phát triển công nghệ mạng di động thế hệ mới như 5G, 6G.

Để nhanh chóng áp dụng các yêu cầu đảm bảo an toàn thông tin cho mạng di động 5G vào thực tế, tạo điều kiện thuận lợi tối đa cho việc thương mại hóa công nghệ 5G một cách an toàn, tin cậy tại Việt Nam, Bộ Khoa học và Công nghệ yêu cầu các doanh nghiệp nghiên cứu, áp dụng các yêu cầu bảo đảm an toàn thông tin cho mạng 5G trong quá trình lựa chọn thiết bị 5G; xây dựng và thiết lập mạng 5G; tiếp tục nghiên cứu, phát triển các yêu cầu đảm bảo an toàn thông tin cho mạng 5G phù hợp với tình hình thực tế và sự phát triển của khoa học và công nghệ. Do đó, việc nghiên cứu kiến trúc mạng 5G, nguy cơ đe dọa an toàn bảo mật và các giải pháp bảo mật trong mạng di động 5G là cấp thiết giúp bắt kịp xu thế công nghệ mới, sớm đưa mạng 5G vào các hoạt động thực tiễn.

Để bảo vệ mạng 5G, một số cuộc tấn công nhất định phải được giải quyết; đáng chú ý nhất là các cuộc tấn công DDoS. Một cuộc tấn công DDoS trong mạng di động hiện tại (ví dụ: 4G) chỉ có thể xâm phạm một dịch vụ. Tuy nhiên, trong mạng 5G, nếu một hacker độc hại nắm quyền kiểm soát một phần và thực hiện một cuộc tấn công DDoS, điều này có thể làm tổn hại đến các dịch vụ thuộc cùng một mạng ảo. Ngoài ra, trong mạng 5G, cuộc tấn công DDoS có thể tăng cường; các phần khác cũng có thể bị xâm phạm bởi cuộc tấn công này nếu giao thức đường hầm được chia sẻ giữa các phần 5G khác nhau. Các hệ thống IoT đang phải đối mặt với sự gia tăng bề mặt tấn công theo cấp số nhân do số lượng lớn các thiết bị IoT có thể bị xâm phạm [2]. Việc theo dõi, phát hiện sớm các truy cập bất thường sẽ giúp ngăn chặn các cuộc tấn công DDoS và hạn chế hậu quả do chúng gây ra.

2. Mục đích nghiên cứu

Mục tiêu của đề án là nghiên cứu tìm hiểu xoay quanh việc xây dựng các giải pháp kỹ thuật nhằm đảm bảo an toàn mạng trước các cuộc tấn công từ chối dịch vụ phân tán (DDoS), đặc biệt trong môi trường mạng hiện đại như mạng di động 5G. Tìm ra các biện pháp tối ưu trong việc phát hiện sớm và chính xác các cuộc tấn công DDoS, tối ưu hóa hiệu suất phát hiện trong môi trường 5G, tăng cường bảo mật cho các dịch vụ mạng, ứng dụng trí tuệ nhân tạo và học máy.

3. Đối tượng và phạm vi nghiên cứu

Đề án tập trung nghiên cứu xoay quanh các đối tượng:

- Mạng di động 5G và các quy cơ đe dọa về tấn công mạng di động; tấn công DDoS trong mạng di động 5G;

- Các phương pháp học máy, học sâu đã được công bố, có thể nhận diện các truy cập bất thường;

- Các công cụ giám sát, ngăn chặn tấn công mạng, đảm bảo an toàn bảo mật cho hệ thống mạng di động.

Phạm vi nghiên cứu:

- Nghiên cứu kiến trúc tổng quan Core network của mạng di động 5G trong phạm vi thuộc lĩnh vực hệ thống thông tin, các thiết bị mạng tham gia và mô hình hoạt động trong mạng di động;

- Các mô hình, giải pháp, thuật toán đã được công bố: KNN, SVM, MLP... sử dụng tập dữ liệu đối sánh 5G-NIDD [3].

4. Phương pháp nghiên cứu

Đề án sử dụng kết hợp các phương pháp nghiên cứu sau:

- Nghiên cứu lý thuyết: Tiến hành khảo sát, tìm hiểu các nghiên cứu về kiến trúc mạng di động nói chung và mạng di động 5G nói riêng, từ đó phát hiện, đánh giá các nguy cơ về an toàn bảo mật trong mạng di động 5G. Tìm hiểu các mô hình học máy, học sâu có thể áp dụng vào nhận diện các truy cập bất thường, cụ thể là phát hiện các cuộc tấn công DDoS.

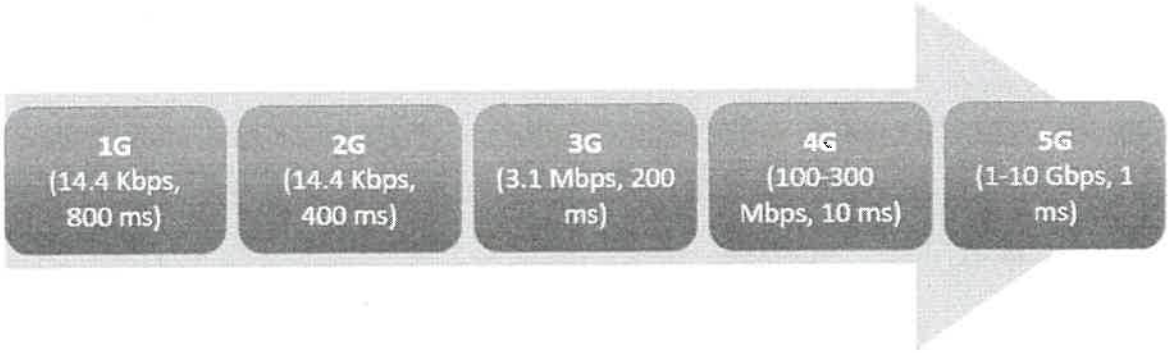
- Nghiên cứu thực nghiệm: Sử dụng các công cụ giám sát mạng để theo dõi, thu thập các gói tin lan truyền trong mạng di động private 5G, áp dụng các mô hình học máy trên tập dữ liệu đối sánh để phân lớp các gói tin bình thường và gói tin bất thường. Kết luận về việc tấn công, đánh giá kết quả thu được.

CHƯƠNG 1: KIẾN TRÚC MẠNG DI ĐỘNG 5G VÀ AN TOÀN BẢO MẬT TRONG MẠNG DI ĐỘNG 5G

1.1. Tổng quan về mạng di động 5G

1.1.1. Kiến trúc mạng 5G

Mạng di động thế hệ thứ năm – mạng di động 5G là thế hệ tiếp theo của công nghệ truyền thông di động sau thế hệ thứ bốn (4G). Theo các nhà phát triển, mạng 5G có tốc độ nhanh hơn khoảng 100 lần so với mạng 4G hiện tại, giúp mở ra nhiều hứa hẹn về các dịch vụ mới và tiện ích. So sánh với công nghệ 3G, 4G, mạng 5G có độ trễ cực thấp (trong khoảng 5-20 ms) và có tham vọng cải tiến tối ưu, đẩy độ trễ xuống chỉ còn 1 ms [1]. Điều này có ý nghĩa rất quan trọng trong việc ứng dụng công nghệ 5G vào một số lĩnh vực đòi hỏi độ trễ của tín hiệu cực thấp như xe tự lái, máy bay tự lái, phẫu thuật từ xa… Sự phát triển của mạng 5G với hệ thống Internet vạn vật (IoT) được thiết lập để cải thiện khả năng liên lạc đáng tin cậy và kết nối ổn định. Công nghệ truy cập vô tuyến mới 5G có độ trễ thấp, tính sẵn sàng cao và tốc độ vượt trội; tất cả đều cần thiết cho các hệ thống IoT [4, 5].



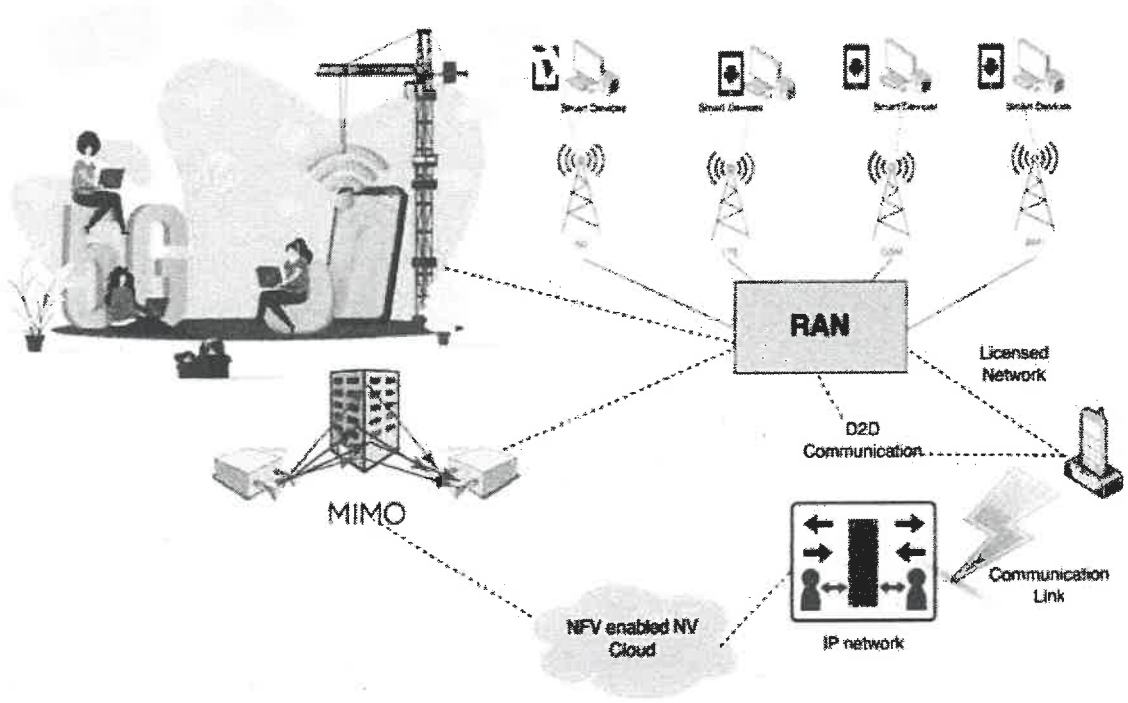
Hình 1.1: Sự phát triển từ mạng 1G đến 5G

Kiến trúc mạng 5G là một sự khác biệt căn bản so với các thế hệ trước, ở chỗ nó được đặc trưng bởi tính linh hoạt và khả năng thích ứng cao hơn. Không giống như các thế hệ mạng di động trước đó, chủ yếu được xây dựng xung quanh một cấu trúc mạng cứng nhắc, phân cấp, kiến trúc 5G được thiết kế để có tính linh hoạt cao, phân tán và được xác định bằng phần mềm. Nó kết hợp một loạt các khái niệm và kỹ

thuật, chẳng hạn như phân chia mạng, điện toán biên và SDN, cùng nhau góp phần nâng cao hiệu suất, chức năng và trải nghiệm người dùng. Thiết kế của kiến trúc mạng 5G giải quyết một số yêu cầu và thách thức do một loạt các ứng dụng mang lại, từ băng thông rộng di động tốc độ cao đến các giao tiếp cực kỳ đáng tin cậy, độ trễ thấp. Tính linh hoạt và khả năng thích ứng của cấu trúc 5G bắt nguồn từ sự chuyển đổi sang một phương pháp tiếp cận hướng đến phần mềm hơn từ phương pháp tiếp cận phần cứng truyền thống, cho phép mức độ linh hoạt, khả năng mở rộng và hiệu quả chưa từng thấy ở các thế hệ trước.

Đặc điểm cốt lõi của cấu trúc mạng 5G là việc sử dụng kiến trúc phẳng, phi tập trung thay vì kiến trúc phân cấp truyền thống, để giảm độ trễ và tối ưu hóa định tuyến lưu lượng. Việc sử dụng kiến trúc phi tập trung cũng cho phép tốc độ thông lượng cao hơn, cho phép truyền dữ liệu nhanh hơn [6]. Để đối phó với nhu cầu về tốc độ dữ liệu cao và khả năng kết nối lớn cần thiết cho các thiết bị IoT, 5G cũng bao gồm việc sử dụng các công nghệ ăng-ten tiên tiến, chẳng hạn như MIMO. Massive MIMO tăng dung lượng của một ô bằng cách sử dụng một số lượng lớn ăng-ten truyền tại trạm gốc để phục vụ một số lượng đáng kể người dùng trong cùng một tài nguyên tần số thời gian. Điều quan trọng là mạng 5G triển khai kiến trúc dựa trên đám mây, nơi các chức năng và dịch vụ có thể được khởi tạo theo cách linh hoạt và năng động. Điều này dẫn đến một sự thay đổi mô hình từ cơ sở hạ tầng phần cứng chuyên dụng sang một môi trường phần mềm linh hoạt hơn, nâng cao khả năng thích ứng và khả năng mở rộng của các dịch vụ mạng.

Hình 1.2 cho thấy kiến trúc mạng của mạng thế hệ thứ năm. Trong mạng 5G, các thiết bị như điện thoại thông minh được kết nối thông qua một phần quan trọng của cơ sở hạ tầng mạng di động được gọi là Mạng Truy cập Vô tuyến (RAN), cho phép tích hợp và cải thiện việc sử dụng mạng của các thiết bị di động. Giao tiếp giữa thiết bị với thiết bị (D2D) là một "mạng của các mạng", trong đó nhiều mạng được tích hợp cho các dịch vụ dữ liệu và giao tiếp mạng qua các công nghệ truy cập vô tuyến [7].



Hình 1.2: Kiến trúc mạng 5G

Tất cả các loại kết nối trong mạng 5G được liên kết với đầu vào đa luồng đầu ra đa luồng (MIMO), là một công nghệ để nhân công suất của liên kết vô tuyến bằng cách dựa vào các mảng ăng-ten truyền và nhận để khai thác truyền sóng đa đường. Kiến trúc mạng bao gồm ba lớp khác nhau: lớp cơ sở hạ tầng, lớp điều khiển và lớp ứng dụng. Mỗi lớp khác nhau về loại vị trí thành phần và mức độ chức năng khác nhau. Các thành phần kết nối như bộ định tuyến, bộ chuyển mạch và trạm gốc được đặt ở lớp cơ sở hạ tầng. Lớp điều khiển triển khai các thực thể ra quyết định và chức năng điều khiển mạng được tích hợp vào lớp ứng dụng. Các dịch vụ mạng được sử dụng và các ứng dụng nghiệp vụ được thực thi ở lớp ứng dụng.

Truyền thông kiểu máy quy mô lớn (mMTC) là một giao tiếp được thực hiện bởi các nền tảng máy móc hoặc phần mềm để phối hợp, cảm biến và truyền động không do con người vận hành. 5G có thể giảm độ trễ MTC xuống 1 ms giữa các thiết bị không dây. Đây là một cải tiến đáng kể so với độ trễ 50 ms và 60 ms của các công nghệ 3G và 4G tương ứng.

1.1.2. Các tính năng và công nghệ hỗ trợ

- Giao tiếp D2D (Device-to-Device):

D2D là một kỹ thuật cho phép hai thiết bị giao tiếp trực tiếp qua mạng mà không cần trạm gốc. Giao tiếp D2D có thể được sử dụng để giải quyết những khó khăn trong các mạng có mật độ cao. Trong giao tiếp D2D, mỗi thiết bị đầu cuối có tùy chọn giao tiếp trực tiếp để trao đổi dữ liệu hoặc trao đổi kết nối truy cập vô tuyến. Giao tiếp D2D có thể giúp giảm nhiễu, đặc biệt là trong các băng tần không được cấp phép. Khái niệm giao tiếp D2D không tồn tại trong mạng 4G.

- MIMO khổng lồ tiên tiến (Advance Massive MIMO):

Hiệu suất phổ tần và tốc độ dữ liệu có thể được tăng lên bằng cách sử dụng MIMO khổng lồ tiên tiến. Nhiều anten truyền và nhận có thể được sử dụng đồng thời để quản lý lượng lớn lưu lượng dữ liệu do các thiết bị tạo ra trong mạng. Trong mạng 4G, một trạm gốc có mười hai cổng để xử lý các giao tiếp di động (MIMO). Tám trong số mười hai cổng dùng để phát sóng, trong khi bốn cổng dùng để nhận. Trong 5G, con số này tăng lên khoảng 100 cổng trên mỗi trạm gốc, dẫn đến giao tiếp MIMO khổng lồ [8].

- Mạng được xác định bằng phần mềm (Software Defined Network - SDN):

SDN là một kiến trúc động, có thể kiểm soát, thích ứng và hiệu quả về chi phí, cho phép cung cấp băng thông cao theo yêu cầu của các ứng dụng khác nhau. SDN tích hợp nhiều công nghệ mạng để làm cho mạng linh hoạt và dễ thích ứng hơn nhằm quản lý các máy chủ và cơ sở hạ tầng lưu trữ được ảo hóa. Mạng SDN là một phương pháp để xây dựng, phát triển và quản lý các mạng phân biệt mặt phẳng chuyển tiếp mạng và mặt phẳng điều khiển.

- Ảo hóa chức năng mạng (Network Function Virtualization - NFV):

NFV là một kỹ thuật mạng đang phát triển nhanh chóng cho phép thay thế các thiết bị phần cứng chuyên dụng chi phí cao, chẳng hạn như bộ định tuyến và tường lửa, bằng các công cụ mạng dựa trên phần mềm và chạy dưới dạng máy ảo trên các máy chủ tiêu chuẩn. 5G phải cho phép giao tiếp D2D sẽ tạo ra một lượng lớn dữ liệu, không khả thi để xử lý tất cả dữ liệu được tạo ra tại trung tâm dữ liệu tập trung. Hơn

nữa, các ứng dụng trong tương lai sẽ rất năng động và đòi hỏi cao, điều này yêu cầu mạng phải linh hoạt. NFV cho phép dữ liệu được xử lý theo các yêu cầu bằng cách đặt một cách chiến lược các thành phần mạng cần thiết. Kỹ thuật này tạo ra các lát mạng (network slices), cho phép điều phối các môi trường mạng được ảo hóa phù hợp với các ứng dụng khác nhau. NFV cũng có thể được sử dụng để triển khai các công cụ bảo mật được ảo hóa như tường lửa và hệ thống phát hiện xâm nhập có thể tăng giảm quy mô dựa trên mô hình lưu lượng truy cập và mức độ đe dọa.

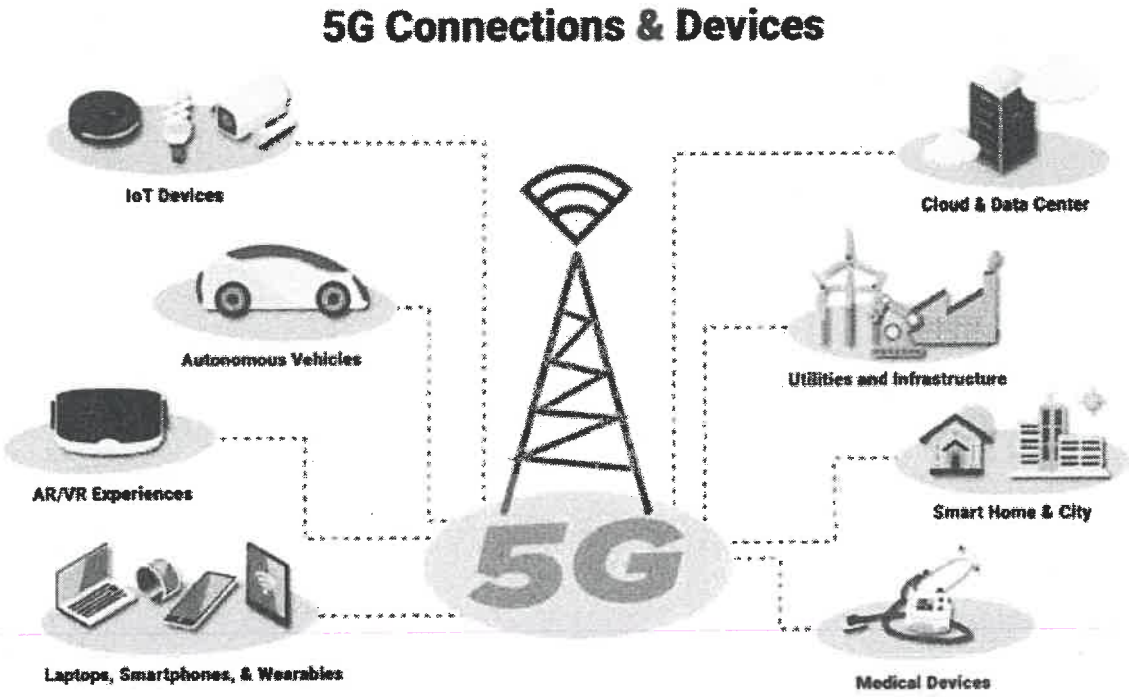
- Điện toán biên (Edge computing):

Điện toán biên là một kiến trúc công nghệ phân tán trong đó dữ liệu được xử lý ở rìa mạng gần điểm gốc. Các thiết bị được dự đoán sẽ đưa ra quyết định và phản hồi tương ứng với các tác vụ để giảm độ trễ cho phép các ứng dụng thời gian thực. Điện toán biên có tiềm năng thay đổi hoàn toàn cách thức cung cấp dịch vụ bằng cách cho phép giám sát từ xa, nhanh hơn và chính xác hơn. Trong những trường hợp này, khi quyết định nhanh chóng quan trọng hơn, điện toán biên là điều cần thiết, đặc biệt là trong các mạng 5G.

- Truyền thông kiểu máy quy mô lớn (mMTC):

Được thiết kế để hỗ trợ triển khai IoT lớn, cho phép kết nối giữa một số lượng lớn các thiết bị trên mỗi ki lô mét vuông, với mMTC, 5G có thể xử lý một mật độ thiết bị được kết nối cao hơn đáng kể so với các thế hệ trước, cho phép hoạt động liền mạch của vô số thiết bị IoT [6]. Điều này rất quan trọng đối với các ứng dụng như thành phố thông minh, nhà thông minh, giám sát môi trường và nông nghiệp, cùng những ứng dụng khác. Để quản lý kết nối thiết bị rộng lớn này, mạng 5G kết hợp quản lý thiết bị nâng cao, quản lý năng lượng và các kỹ thuật báo hiệu để đảm bảo sử dụng hiệu quả tài nguyên mạng và duy trì thời lượng pin của thiết bị.

Với các tính năng và công nghệ kể trên, mạng di động 5G có thể ứng dụng vào rất nhiều mục đích khác nhau trong thực tế. Hình 1.3 cho thấy một số ứng dụng như: Kết nối các thiết bị IoT, xe ô tô tự hành, các thiết bị di động thông minh, nhà thông minh, thành phố thông minh, các dịch vụ chăm sóc sức khỏe hiện đại...



Hình 1.3: Các ứng dụng của mạng di động 5G

Thiết kế của 5G cũng cho phép tích hợp liền mạch với các hệ thống IoT không đồng nhất, hỗ trợ các ứng dụng yêu cầu nghiêm ngặt về độ trễ như phẫu thuật từ xa, sản xuất thông minh. Các hệ thống tính toán biên di động (MEC) giúp đưa trí tuệ tính toán đến gần biên mạng hơn, cho phép mở rộng các ứng dụng AI. Mô hình phân tán này cũng giúp tăng hiệu quả xử lý, giảm tắc nghẽn backhaul và cung cấp nền tảng cho các dịch vụ thông minh và thích nghi. Nó cũng cho phép khả năng phân chia mạng động (network slicing), điều rất cần thiết để phân tách dịch vụ giữa các lĩnh vực ứng dụng.

1.2. Nguy cơ đe dọa an toàn bảo mật trong mạng di động 5G

Dù có những tiến bộ về kiến trúc, 5G cũng bộc lộ những bề mặt tấn công mới. Các giao diện lập trình động (API) và chức năng ảo hóa như NFV và SDN gia tăng độ linh hoạt nhưng đồng thời cũng là mục tiêu hấp dẫn cho kẻ tấn công. Các tín hiệu truy cập ban đầu như SSB, MIB và SIB1 không được mã hóa, dễ bị giả mạo hoặc can thiệp, tạo điều kiện cho các cuộc tấn công chiếm quyền truy cập mạng ngay từ giai

đoạn đồng bộ hóa. Ngoài ra, mô hình phân chia mạng (network slicing) cũng tiềm ẩn rủi ro khi cấu hình lát mạng không chuẩn hoặc thiếu cách ly nghiêm ngặt, tạo điều kiện cho các tấn công chéo lát (cross-slice attack), nơi tin tặc có thể xâm nhập một lát có bảo mật yếu như lát IoT rồi di chuyển sang các lát mạng khác.

Ngoài việc cung cấp hiệu suất mạng, các hệ thống IoT hỗ trợ 5G cũng cần duy trì tính bảo mật và cải thiện độ tin cậy của dịch vụ. Một cuộc tấn công được thực hiện thành công vào mạng 5G có thể dẫn đến các hoạt động không mong muốn và gây ra hậu quả nghiêm trọng. Những tin tặc đang sử dụng các chiến thuật mới để kiếm tiền từ các cuộc tấn công của chúng bằng cách kiểm soát dữ liệu nhạy cảm, yêu cầu tiền chuộc hoặc khiến mạng không hoạt động. Theo 5G PPP, các ứng dụng IoT hỗ trợ 5G dự kiến sẽ gặp phải một số vấn đề bảo mật do tính phức tạp và khả năng mở rộng của bề mặt tấn công [9, 10]. Quy mô kết nối và liên kết thiết bị ngày càng tăng cùng với việc phân chia mạng sẽ cho phép một loạt ứng dụng IoT mới. Tuy nhiên, một trong những điểm yếu về bảo mật sẽ nằm ở chính các thiết bị; chúng có thể được điều khiển từ xa để hình thành cái được gọi là mạng botnet nhằm thực hiện các cuộc tấn công bảo mật nghiêm trọng. Hầu hết các thiết bị IoT hiện tại không được phát triển với ưu tiên bảo mật [11]. Khối lượng dữ liệu do các thiết bị IoT tạo ra cũng được dự đoán sẽ tăng theo cấp số nhân; các phương pháp phát hiện xâm nhập truyền thống có thể kém tin cậy hơn và gặp nhiều thách thức hơn. Các phương pháp phát hiện tấn công bảo mật mới cần được khám phá thêm trong mạng 5G.

Mối đe dọa về an toàn bảo mật trong mạng di động 5G cần chú ý:

1.2.1. Xác thực (Authentication)

Xác thực là một khái niệm quan trọng trong bảo mật mạng 5G vì nó cho phép xác minh danh tính người dùng trong mạng. Nhiều cơ chế được sử dụng trong mạng để xác thực dữ liệu. Nó được chia thành hai phần: xác thực chính và xác thực phụ. Xác thực chính cho phép xác thực lẫn nhau của các thiết bị và mạng trong cả mạng 5G và 4G. Tuy nhiên, xác thực chính trong mạng dựa trên 5G có một số vấn đề, bao gồm kiểm soát kiến thức và việc gọi xác thực thiết bị không được hỗ trợ đầy đủ.

Những khó khăn này được giải quyết thông qua việc sử dụng Giao thức Xác thực Quyền riêng tư (5G-AKA) và các lược đồ xác thực linh hoạt. Xác thực chính tương thích với các công nghệ khác ngoài Dự án Đối tác Thế hệ thứ ba (3GPP). Xác thực phụ được sử dụng ngoài phạm vi của nhà khai thác di động và dựa trên tiêu chuẩn 3GPP. Xác thực phụ có thể thực hiện thông qua các kỹ thuật liên kết dựa trên Giao thức Xác thực Mở rộng (EAP) [8].

1.2.2. Tính bảo mật (Confidentiality)

Tính bảo mật đảm bảo rằng chỉ người được ủy quyền mới có thể truy cập thông tin về người gửi. Khóa cần thiết cho nút gốc thế hệ tiếp theo thứ cấp được tạo và cung cấp bởi nút gốc chính trước bất kỳ quá trình truyền vô tuyến an toàn nào. Tín hiệu cho kiểm soát tài nguyên vô tuyến có thể được gửi giữa nút gốc thế hệ tiếp theo thứ cấp và thiết bị người dùng. Do đó, các khóa được sử dụng để đảm bảo tính xác thực và quyền riêng tư của dữ liệu mặt phẳng người dùng (UP – User plane) và các thông điệp. Mặc dù các mạng 5G cho phép bảo vệ tính toàn vẹn cho dữ liệu, nhưng chúng không thể được triển khai trong kịch bản kết nối kép.

1.2.3. Tính sẵn sàng (Availability)

Các dịch vụ thông minh dựa trên mạng di động 5G được hưởng lợi từ các tài nguyên đám mây, giúp phát triển cơ sở hạ tầng hiệu quả về chi phí. Tuy nhiên, chúng tiềm ẩn các mối đe dọa bảo mật như tấn công mạng gây ảnh hưởng tới độ tin cậy của mạng. Các cuộc tấn công DDoS yêu cầu tài nguyên vật lý và logic ở cấp độ biên và đám mây, điều này có ảnh hưởng đến các quá trình phân chia mạng. Các cuộc tấn công gây nhiễu gây ra sự cố trên các cơ sở truy cập vô tuyến, ngăn người dùng truy cập các dịch vụ di động. Các cuộc tấn công vào tài nguyên 5G, bao gồm hệ thống hỗ trợ, mặt phẳng điều khiển và vô tuyến, có thể làm gián đoạn các dịch vụ mạng thông minh.

1.2.4. Tính chống chối bỏ (Non-repudiation)

Khả năng chứng minh tính hợp lệ và độ tin cậy của một thông điệp hoặc giao dịch và ngăn người gửi từ chối sự tham gia của họ vào giao tiếp được gọi là tính chống chối bỏ. Việc người dùng từ chối không thể bị ngăn chặn chỉ bằng xác thực.

Tuy nhiên, việc phân biệt giữa những người dùng hoặc thiết bị khác nhau là rất quan trọng để tạo ra dữ liệu an toàn, nên xác thực là cần thiết để đảm bảo tính chống chối bỏ. Các mạng 5G có thể sử dụng chữ ký số để đảm bảo tính chống chối bỏ. Chữ ký số sử dụng một kỹ thuật mật mã để tạo ra một mã duy nhất được thêm vào thông điệp, có thể được sử dụng để chứng minh tính xác thực và tính toàn vẹn của tin nhắn. Điều này gây khó khăn cho người gửi trong việc từ chối sự tham gia của họ vào giao tiếp.

1.2.5. Tính toàn vẹn (Integrity)

Một thước đo bảo mật trong mạng 5G là bảo vệ tính toàn vẹn của mặt phẳng người dùng giữa nút mạng và các thiết bị. Bảo vệ tính toàn vẹn là một tính năng tốn nhiều tài nguyên, các thiết bị IoT không thể triển khai với tỉ lệ dữ liệu cao do những hạn chế của chúng. Do đó, kiến trúc mạng thông minh dựa trên 5G phải bao gồm các giao thức đảm bảo tính toàn vẹn của mạng. Ví dụ, các mạng 5G có thể sử dụng các thuật toán mật mã như Tiêu chuẩn Mã hóa Tiên tiến (AES) và Thuật toán Băm An toàn (SHA) để tạo ra một mã duy nhất được thêm vào dữ liệu, có thể được sử dụng để xác minh tính toàn vẹn của dữ liệu. Điều này đảm bảo rằng dữ liệu không bị sửa đổi hoặc giả mạo trong quá trình truyền.

1.3. Một số hình thức tấn công

Khi mạng 5G ngày càng phát triển, các phương pháp và kỹ thuật được sử dụng bởi những kẻ tấn công khai thác các lỗ hổng cũng ngày càng tinh vi. Một số phương pháp và kỹ thuật tấn công sẽ được khái quát trong phần này. Các cuộc tấn công DDoS được dự đoán sẽ là một mối đe dọa đáng kể đối với mạng 5G. Vì mạng 5G cung cấp băng thông cao hơn đáng kể, chúng có khả năng được tận dụng để phát động các cuộc tấn công DDoS quy mô lớn hơn so với các cuộc tấn công được thấy trên các thế hệ mạng di động trước đó. Một phương pháp khác là cuộc tấn công MitM, trong đó kẻ tấn công chặn giao tiếp giữa hai bên để đánh cắp dữ liệu hoặc chèn nội dung độc hại. Sự gia tăng sự phụ thuộc vào điện toán biên và các thiết bị IoT, thường có các biện pháp bảo mật yếu hơn, làm cho mạng 5G có khả năng dễ bị tấn công MitM hơn…

1.3.1. Tấn công DoS (Denial of Service - Từ chối dịch vụ)

Trong một cuộc tấn công từ chối dịch vụ (DoS), kẻ tấn công cố gắng gửi một lượng lớn dữ liệu giả mạo nhắm vào các nút của mạng. Số lượng lớn các yêu cầu này có thể tiêu tốn năng lượng, băng thông và thời gian,..., khiến nút bị treo hoặc không thể phản hồi các yêu cầu hợp lệ. Khi nhiều máy tính trên một mạng được sử dụng để khởi tạo một cuộc tấn công từ chối dịch vụ phân tán (DDoS), tính bảo mật của một số máy tính và mạng đó có thể bị xâm phạm. Trong mạng 5G, kẻ tấn công còn khai thác các lỗ hổng trong giao thức mạng, hoặc lạm dụng các lỗ hổng bảo mật của hệ điều hành và các ứng dụng được tải xuống từ cửa hàng ứng dụng để can thiệp vào các thiết bị di động.

1.3.2. Tấn công MitM (Man-in-the-Middle)

Kẻ tấn công sẽ thực hiện một cuộc tấn công man-in-the-middle, còn được gọi là tấn công Người ở giữa, khi chúng thiết lập một điều kiện tạm thời cho phép chặn truyền dữ liệu giữa các thiết bị qua mạng để nghe nén hoặc thay đổi nội dung của việc truyền dữ liệu.

1.3.3. Tấn công chiếm quyền điều khiển (Hijacking attacks)

Nó phục vụ mục đích lãng phí tài nguyên của bộ điều khiển (tức là bão hòa mặt phẳng dữ liệu sang mặt phẳng điều khiển). Bằng cách tận dụng các tài nguyên mà bộ điều khiển cung cấp, mục tiêu của kẻ tấn công là làm chậm hoặc có thể vô hiệu hóa hoàn toàn một số khu vực của mạng.

1.3.4. Tấn công giả mạo (Spoofing attacks)

Giả mạo là một lỗ hổng cho phép kẻ tấn công chặn các tương tác hợp pháp trên mạng 5G. Các cuộc tấn công giả mạo bao gồm việc tiêm các tín hiệu sai bằng cách sử dụng danh tính giả để có được lợi thế bất hợp pháp và khởi động các cuộc tấn công độc hại khác như tấn công man-in-the-middle và tấn công từ chối dịch vụ. Các cuộc tấn công giả mạo là một rủi ro trong truyền thông không dây được sử dụng cho trong các dịch vụ thông minh dựa trên 5G do tính dễ bị tổn thương của truyền thông không dây đối với các cuộc tấn công lớp vật lý.

1.3.5. Tấn công gây nhiễu (Jamming attack)

Hiệu suất của mạng 5G có thể bị ảnh hưởng bởi các cuộc tấn công mạng độc hại, có thể khởi động chống lại mạng. Bởi vì các kênh điều khiển là cần thiết để vận hành giao diện vô tuyến một cách hiệu quả, truyền thông không dây trong các mạng dựa trên 5G dễ bị tấn công làm nhiễu giao diện vô tuyến. Bằng cách làm nhiễu các kênh điều khiển nhất định bằng các tín hiệu công suất cao, kẻ tấn công có thể gây thiệt hại cho các băng tần.

1.3.6. Tấn công trạm gốc giả mạo (Rogue base station attacks)

Tự động hóa tối ưu hóa mạng và thiết lập để quản lý mạng tối ưu đã dẫn đến sự xuất hiện của một mối đe dọa mới gọi là Trạm gốc giả mạo (RBS). Để thực hiện giám sát trái phép và bất hợp pháp đối với bất kỳ sự xáo trộn giao tiếp tiềm ẩn nào, kẻ tấn công đóng giả một trạm gốc thực. Kẻ tấn công cố gắng tìm hiểu danh tính của người dùng bằng cách theo dõi Định danh thuê bao di động quốc tế (IMSI) của thiết bị người dùng bằng cách sử dụng các trạm gốc giả.

1.3.7. Tấn công kênh bên (Side-channel attack)

Nhiều lát của mạng dựa trên 5G sử dụng cùng một cơ sở hạ tầng và tài nguyên vật lý, cho phép các cuộc tấn công kênh bên của lát mạng 5G. Khi một trình dịch có khả năng suy ra các mẫu và đặc điểm vật lý cụ thể, chẳng hạn như quản lý năng lượng, để lấy dữ liệu bí mật, loại tấn công này được gọi là tấn công kênh bên. So với các thế hệ mạng trước đó, mạng 5G dễ bị tấn công loại này hơn vì chúng dựa trên phân chia mạng, cho phép kẻ tấn công dễ dàng cô lập và đánh giá hiệu suất của một lát cụ thể.

1.4. Các phương pháp bảo mật hiện có

Khi chúng ta chứng kiến sự tăng trưởng theo cấp số nhân trong việc triển khai mạng 5G, điều bắt buộc là phải giải quyết các mối quan tâm liên quan tới bảo mật. Các vấn đề này bao gồm các lỗ hổng kỹ thuật, sự mơ hồ về quy định và các mối đe dọa tiềm tàng từ một nhóm đa dạng các tác nhân độc hại. Để củng cố bảo mật của mạng 5G, vô số các giải pháp hiện tại và tiềm năng đang được khám phá. Các giải pháp này có nhiều lớp và hoạt động trong các khía cạnh khác nhau của mạng, chẳng

hạn như các cấp độ kỹ thuật, tổ chức, quy định và người dùng. Các giải pháp kỹ thuật tạo thành nền tảng của bảo mật 5G, vì chúng trực tiếp giải quyết các mối đe dọa và lỗ hổng khác nhau.

1.4.1. ENDER

ENDER là viết tắt của hệ thống tiền quyết định (*pre-dEcisioN*), quyết định nâng cao (*advance Decision*), và học hỏi (*lEaRning*) [8]. Hai phương pháp lý thuyết quyết định được sử dụng trong chiến lược CPS để xác định lưu lượng tấn công trên đám mây nhằm giảm thiểu các cuộc tấn công lưu lượng HX-DoS và SIPDAS. Tiếp theo là một quy trình tương tự như của một hệ thống phát hiện xâm nhập tiêu chuẩn. Do đó, nó có khả năng xác định và đánh dấu một tin nhắn tấn công. Khi một tin nhắn tấn công SIPDAS hoặc HX-DoS được phát hiện, tin nhắn đó được đánh dấu bằng một bit duy nhất. Các thuật toán sẽ xóa những tin nhắn này khỏi hệ thống.

1.4.2. Thuật toán dựa trên học máy (Machine learning)

Việc triển khai một hệ thống xác thực dựa trên học máy có thể làm giảm các mối đe dọa giả mạo trong mạng 5G. Mô hình xác thực được củng cố bằng cách áp dụng một đặc điểm hai chiều cho phương pháp một chiều vì phương pháp hai chiều hiệu quả hơn trong việc xác định những kẻ tấn công. Bộ phân loại được trình bày có khả năng xác thực nhanh chóng vì nó được huấn luyện ngoại tuyến.

1.4.3. SDN-5G

Để làm cho mạng trở nên an toàn hơn nữa, kiến trúc bảo mật SDN-5G đã được phát triển. Phương pháp này dựa trên một khóa bí mật được đồng bộ hóa. Khóa này được tạo ra bởi một thuật toán mã hóa, sau đó chúng được lưu trữ trong các hệ thống hỗ trợ và thiết bị 5G. Mạng sẽ có thể biết liệu một cuộc tấn công có đến từ một cuộc tấn công giả mạo IP hay không vì khóa bí mật thay đổi theo thời gian. Bên cạnh đó, cuộc tấn công sẽ được phát hiện và ngăn chặn vì hệ thống hỗ trợ không thể cập nhật khóa bí mật được lưu trữ trên thiết bị. Liên quan đến kiểu tấn công phát lại, khả năng điều này xảy ra trong mạng 5G theo thiết kế này gần như bằng 0, vì khóa bí mật sẽ thay đổi với mỗi lần truyền, khiến tin nhắn được phát lại với một khóa đã lỗi thời.

1.4.4. Khung học sâu (Deep learning framework)

Bằng cách sử dụng các thuật toán học sâu dựa trên AI, các cuộc tấn công gây nhiễu được giảm thiểu. Phương pháp học sâu được đề xuất sử dụng một mạng lưới các thiết bị không dây không đồng nhất được kết nối với nhau để thu thập dữ liệu cần thiết. Việc mở rộng khám phá các mối đe dọa mạng đòi hỏi phải sử dụng rộng rãi kỹ thuật học không giám sát.

1.4.5. Phương pháp NFV

Ảo hóa phân tán cải thiện đáng kể tính bảo mật của mạng; nó cũng tăng tính linh hoạt và độ tin cậy của mạng và giải quyết nhiều vấn đề tấn công ở nhiều tầng của mạng dựa trên 5G, bao gồm giả mạo dữ liệu, nghe lén, v.v.. Truyền dữ liệu được mã hóa ngăn chặn các cuộc tấn công giả mạo dữ liệu trong quá trình truyền.

1.4.6. SDN-guard

SDN là một kiến trúc đột phá cho quản lý mạng tập trung cho hệ thống phân chia mạng. Điều này cho phép những cải tiến trong khả năng lập trình mạng để đáp ứng một loạt các ứng dụng ngày càng tăng. *SDN-guard* được giới thiệu như một phương pháp để giải quyết các vấn đề liên quan đến bảo mật và cung cấp câu trả lời để bảo vệ các ứng dụng internet xúc giác sử dụng hệ thống mờ chống lại các cuộc tấn công MitM và DoS.

1.4.7. SDN-SC

Đây là một kiến trúc cho mạng xác định bằng phần mềm (SDN) cung cấp bảo mật cho mạng 5G. Được áp dụng để xử lý các thách thức bảo mật trong mạng lõi 5G. Do đó, trí thông minh tập trung theo logic, tính trừu tượng và khả năng lập trình của SDN mang lại những lợi ích lớn cho việc giải quyết các vấn đề bảo mật mạng di động. Do khả năng linh hoạt của kiến trúc này, nó có thể đáp ứng nhiều yêu cầu bảo mật của người dùng và triển khai dịch vụ nhanh nhất có thể.

1.4.8. Học Q-network sâu (Deep Q-network learning)

Trí tuệ nhân tạo có tiềm năng mở ra các giao thức bảo mật mới cho mạng. Đáng chú ý nhất là Học tăng cường, trong đó hệ thống giám sát độ phức tạp bảo mật của mạng và sau đó học hỏi từ việc giám sát đó bằng cách sử dụng thuật toán học *Q-network sâu.*

1.5. Học máy và bảo mật 5G

1.5.1. Vai trò của Học máy trong việc tăng cường bảo mật 5G

Việc sử dụng các thuật toán Học máy (ML) cung cấp một giải pháp tiên tiến và hiệu quả để tăng cường và giảm các thách thức bảo mật liên quan đến sự ra đời của mạng 5G. Các thuộc tính vốn có của ML, chẳng hạn như khả năng thích ứng, khả năng dự đoán và xử lý dữ liệu quy mô lớn, cung cấp các giải pháp đầy hứa hẹn để xử lý quy mô và độ phức tạp chưa từng có của hệ sinh thái 5G. Sự ra đời của mạng 5G đã mang lại một sự thay đổi mô hình trong thế giới viễn thông, hứa hẹn khả năng kết nối tốc độ cao, độ trễ thấp và tích hợp liền mạch hàng tỷ thiết bị trên toàn cầu. Tuy nhiên, sự tiến bộ to lớn này cũng mang lại một loạt các thách thức bảo mật mới đòi hỏi các giải pháp sáng tạo. Các thuật toán ML có khả năng xử lý khối lượng lớn dữ liệu được tạo bởi mạng 5G, trích xuất thông tin chi tiết có giá trị về các mối đe dọa có thể xảy ra và cung cấp các biện pháp bảo mật chủ động để ngăn chặn các cuộc tấn công tiềm ẩn. ML đã nổi lên như một công cụ mạnh mẽ trong bối cảnh này, cung cấp các khả năng mạnh mẽ để cải thiện khuôn khổ bảo mật trong mạng 5G.

Học máy, với khả năng nhận dạng mẫu, phát hiện bất thường và mô hình hóa dự đoán, đóng một vai trò quan trọng trong việc tăng cường bộ máy bảo mật của mạng 5G [6]. Vai trò của ML trong bảo mật 5G là vô cùng phong phú và đa dạng, bao gồm các lĩnh vực như phát hiện xâm nhập, bảo vệ quyền riêng tư, định tuyến an toàn và tình báo mối đe dọa, cùng những lĩnh vực khác. Nó có tiềm năng biến đổi các khuôn khổ bảo mật thông thường, thích ứng thành các hệ thống chủ động, có khả năng ngăn chặn thông minh các cuộc tấn công mạng trước khi chúng có thể gây ra thiệt hại đáng kể.

Vô số các kỹ thuật học máy có liên quan đáng kể trong lĩnh vực bảo mật 5G, cung cấp các khả năng độc đáo cho các khía cạnh khác nhau của quản lý bảo mật mạng. Các kỹ thuật học có giám sát như Máy vectơ hỗ trợ (SVM) và Cây quyết định thường được sử dụng để phát hiện xâm nhập trong mạng 5G [6]. Các thuật toán này học từ dữ liệu đào tạo được gắn nhãn để phân loại các hoạt động mạng là bình thường hoặc độc hại. Các thuật toán học không giám sát như phân cụm K-means, Hồi quy tuyến tính, bộ phân loại được giám sát và phân cụm phân cấp có giá trị để phát hiện bất thường, xác định các mẫu bất thường trong dữ liệu mạng có thể biểu thị vi phạm bảo mật. Các thuật toán này không yêu cầu dữ liệu được gắn nhãn, làm cho chúng trở thành các công cụ linh hoạt để khám phá các mối đe dọa chưa biết.

1.5.2. Ứng dụng của Học máy trong bảo mật 5G

Học máy, với các khả năng phân tích nâng cao, đóng vai trò quan trọng trong các ứng dụng khác nhau trong bảo mật 5G. Các ứng dụng này trải rộng trên các lĩnh vực đa dạng như phát hiện xâm nhập, giảm thiểu phần mềm độc hại và mô hình hóa bảo mật dự đoán. Trong ngữ cảnh phát hiện xâm nhập, các thuật toán học máy phân tích dữ liệu lưu lượng mạng để xác định các mẫu lệch khỏi chuẩn mực thông thường. Các kỹ thuật như SVM và Cây quyết định giúp tăng cường khả năng phản ứng và độ chính xác của hệ thống bảo mật, do đó giảm thiểu rủi ro tấn công mạng thành công.

Học máy cũng đóng vai trò quan trọng trong việc chống lại phần mềm độc hại. Nó tạo điều kiện cho việc xác định và phân loại phần mềm có hại dựa trên các đặc điểm hành vi và các tính năng nhị phân của chúng. Ví dụ, các phương pháp học sâu như CNN có hiệu quả cao trong việc trích xuất và học các tính năng phức tạp từ dữ liệu quy mô lớn, cho phép phát hiện hiệu quả phần mềm độc hại zero-day [6]. Mô hình hóa bảo mật dự đoán là một lĩnh vực khác mà học máy có tiềm năng đáng kể. Điều này cho phép các biện pháp phòng thủ chủ động, do đó cải thiện khả năng phục hồi của mạng 5G trước các mối đe dọa mạng tinh vi.

Một số công trình nghiên cứu đã được công bố về việc sử dụng các thuật toán học máy trong bảo mật 5G:

- **"A Comprehensive Study on the Role of Machine Learning in 5G Security: Challenges, Technologies, and Solutions"** (Fakhouri, H.N.; Alawadi, S.; Awaysheh, F.M.; Hani, 2023) [6]: Nghiên cứu này tổng hợp vai trò của Học máy trong việc tăng cường bảo mật 5G. Nó đề cập đến việc sử dụng các kỹ thuật học có giám sát như Máy vectơ hỗ trợ (SVM) và Cây quyết định để phát hiện xâm nhập, học từ dữ liệu đào tạo được gắn nhãn để phân loại các hoạt động mạng là bình thường hoặc độc hại. Các thuật toán học không giám sát như phân cụm K-means, Hồi quy tuyến tính, bộ phân loại được giám sát và phân cụm phân cấp được sử dụng để phát hiện bất thường, xác định các mẫu bất thường trong dữ liệu mạng có thể biểu thị vi phạm bảo mật mà không yêu cầu dữ liệu được gắn nhãn. Học máy cũng tạo điều kiện cho việc xác định và phân loại phần mềm độc hại dựa trên các đặc điểm hành vi và các tính năng nhị phân của chúng.

Ưu điểm: Khả năng thích ứng, dự đoán và xử lý dữ liệu quy mô lớn: Cung cấp các giải pháp để xử lý quy mô và độ phức tạp của hệ sinh thái 5G. Có khả năng xử lý khối lượng lớn dữ liệu được tạo bởi mạng 5G, trích xuất thông tin chi tiết có giá trị về các mối đe dọa có thể xảy ra và cung cấp các biện pháp bảo mật chủ động để ngăn chặn các cuộc tấn công tiềm ẩn.

Nhược điểm: Hiệu quả của các kỹ thuật ML phụ thuộc vào chất lượng và tính toàn diện của dữ liệu có sẵn.

- **"Cyber-Attack Detection and Mitigation Using SVM for 5G Network"** (Sulaiman Yousef Alshunaifi, Shailendra Mishra and Mohammed Alshehri, 2022) [12]: Nghiên cứu này sử dụng thuật toán Máy vectơ hỗ trợ (SVM) để phát hiện các cuộc tấn công DDoS trong mạng 5G. SVM phân loại lưu lượng truy cập giữa các gói tin bình thường và bất thường, đồng thời các cuộc tấn công có thể được phân loại dựa trên các đặc điểm lưu lượng truy cập như tốc độ cổng nguồn và độ lệch chuẩn của các gói tin luồng.

Ưu điểm: Tỷ lệ phát hiện các cuộc tấn công từ chối dịch vụ phân tán (DDoS) sử dụng thuật toán SVM nằm trong khoảng từ 93% đến 98%. Thuật toán tiêu tốn ít bộ nhớ vì chỉ sử dụng các điểm trong tập hỗ trợ để dự báo trong hàm quyết định.

Nhược điểm: Việc chọn hàm kernel phù hợp và tinh chỉnh các tham số của thuật toán SVM đóng vai trò quan trọng, ảnh hưởng trực tiếp đến hiệu suất của mô hình. Lựa chọn kernel không phù hợp có thể làm giảm đáng kể hiệu quả của mô hình. Việc triển khai SVM trong môi trường mạng 5G với lượng dữ liệu cực lớn và yêu cầu xử lý thời gian thực đặt ra nhiều thách thức về khối lượng dữ liệu, thời gian huấn luyện, tính đa dạng của lưu lượng mạng.

1.5.3. Tối ưu hóa giao thức bảo mật với Học máy

ML cũng có thể được tận dụng để tối ưu hóa hoạt động của các giao thức bảo mật trong mạng 5G. Mục tiêu của tối ưu hóa giao thức bảo mật là tăng cường hiệu suất bảo mật của mạng mà không tiêu tốn quá nhiều tài nguyên tính toán hoặc ảnh hưởng đến hiệu quả hoạt động của mạng. Một lĩnh vực quan trọng mà ML có thể được sử dụng là trong việc tối ưu hóa các giao thức quản lý khóa, một nền tảng của giao tiếp an toàn trong mạng 5G. Các thuật toán ML có thể học và dự đoán thời gian tối ưu cho vòng quay khóa, do đó tăng cường bảo mật của giao tiếp được mã hóa trong khi giảm thiểu chi phí liên quan đến các thay đổi khóa thường xuyên.

Ngoài ra, ML cũng có thể hỗ trợ tối ưu hóa các cấu hình và chính sách bảo mật. Với sự phức tạp của mạng 5G, việc quản lý và tối ưu hóa các cài đặt bảo mật có thể là một nhiệm vụ khó khăn. Các thuật toán ML có thể học hỏi từ các sự cố trong quá khứ, phân tích tác động của các cấu hình khác nhau và đề xuất các cài đặt tối ưu để tăng cường tư thế bảo mật của mạng. Hiệu quả của các kỹ thuật ML phụ thuộc vào chất lượng và tính toàn diện của dữ liệu có sẵn.

1.6. Kết luận chương

Chương 1 của đề án tập trung vào việc trình bày kiến trúc mạng di động 5G và các vấn đề liên quan đến an toàn bảo mật trong môi trường mạng này. Mạng 5G được thiết kế với độ trễ cực thấp, tốc độ truyền dữ liệu cao, khả năng kết nối hàng tỷ thiết bị và tính linh hoạt vượt trội nhờ các công nghệ như SDN, NFV, MIMO, điện toán biên và phân chia mạng. Tuy nhiên, chính sự phức tạp và mở rộng của kiến trúc này đã tạo ra nhiều bề mặt tấn công mới, đặc biệt là trong các giao thức chưa mã hóa

và các lát mạng cấu hình không chuẩn. Các nguy cơ bảo mật được phân tích bao gồm xác thực, tính bảo mật, tính toàn vẹn, tính sẵn sàng và chống chối bỏ, với nhiều điểm yếu có thể bị khai thác bởi tin tặc. Chương cũng liệt kê các hình thức tấn công phổ biến trong mạng 5G như DDoS, MitM, giả mạo, chiếm quyền điều khiển, gây nhiễu, trạm gốc giả mạo và tấn công kênh bên. Để đối phó với các mối đe dọa này, nhiều giải pháp bảo mật đã được đề xuất như ENDER, SDN-guard, SDN-SC, học sâu và học tăng cường. Cuối cùng, chương giới thiệu các giải pháp bảo mật hiện có và vai trò của học máy trong việc tăng cường an toàn, hứa hẹn mang lại nhiều giải pháp cho bảo mật mạng di động 5G.

CHƯƠNG 2: PHÁT HIỆN TẤN CÔNG DDOS TRONG MẠNG DI ĐỘNG 5G

2.1. Tấn công DDoS mạng di động 5G

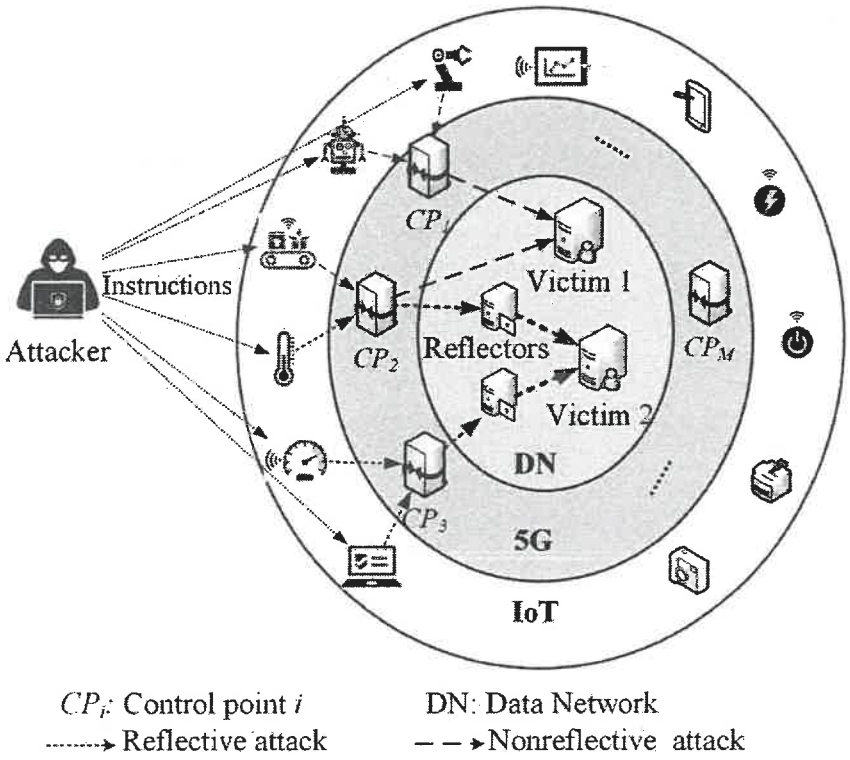
DoS (*Denial-of-Service*) hay tấn công từ chối dịch vụ là hình thức tấn công mạng mà kẻ tấn công tìm cách ngăn cản người dùng máy tính sử dụng mạng truy cập đến máy chủ. Hình thức tấn công DoS phổ biến chính là tin tặc sẽ gửi một lưu lượng lớn truy cập vào hệ thống máy chủ làm cạn kiệt tài nguyên của nạn nhân. Có nhiều phương pháp tấn công từ chối dịch vụ như SYN Flood, UDP Flood, TCP Flood, Slowloris, Ping Flood, Ip NULL,…

Một cuộc tấn công 5G DoS nhằm mục đích làm gián đoạn chức năng của một thành phần cụ thể của mạng 5G, chẳng hạn như các trạm gốc hoặc các chức năng mạng ảo lõi, bằng cách áp đảo nó bằng lưu lượng hoặc yêu cầu quá mức, khiến các dịch vụ tạm thời hoặc vĩnh viễn không khả dụng. Một cuộc tấn công 5G DDoS là một hình thức phân tán của phương pháp này, trong đó nhiều thiết bị mạng bị xâm nhập bằng số lượng lớn lưu lượng độc hại một cách đồng thời và nhanh hơn. Mạng 5G đặc biệt dễ bị tác động bởi các cuộc tấn công này do sự hỗ trợ của chúng đối với việc kết nối cùng lúc nhiều thiết bị bởi băng thông mở rộng. Ngoài ra, với kiến trúc di động phân tán thế hệ mới, kết hợp các tính năng nâng cao như phân chia mạng và điện toán biên, tiềm ẩn nhiều bề mặt tấn công mới.

2.1.1. Mạng Botnet IoT

Một số lượng lớn các thiết bị IoT dự kiến kết nối qua đặt ra một thách thức đáng kể trong nghiên cứu bảo mật mạng di động 5G. Kẻ tấn công khai thác các thiết bị IoT được kết nối qua 5G để tiêm mã độc hại, tạo ra các botnet lớn có thể được điều khiển từ xa thông qua máy chủ Điều khiển và chỉ huy (C&C) [13]. Do việc cài đặt chức năng bảo mật cấp cao cho các thiết bị IoT cấp thấp là khá khó khăn hoặc không khả thi, nhiều thiết bị IoT có bảo mật tích hợp tối thiểu với mật khẩu yếu và bảo vệ cũ. Vì vậy, chúng có nhiều khả năng bị khai thác. Một trong những thành phần thường xuyên bị nhắm mục tiêu nhất cho loại tấn công này là tài nguyên vô tuyến của RAN.

Botnet tràn ngập RAN bằng cách yêu cầu truy cập đồng thời quá mức vào tài nguyên không dây. Với khả năng mở rộng mạnh mẽ của 5G, có thể hỗ trợ hàng tỷ thiết bị, càng làm tăng tác động tiềm tàng của các lỗ hổng như vậy đối với an toàn bảo mật.

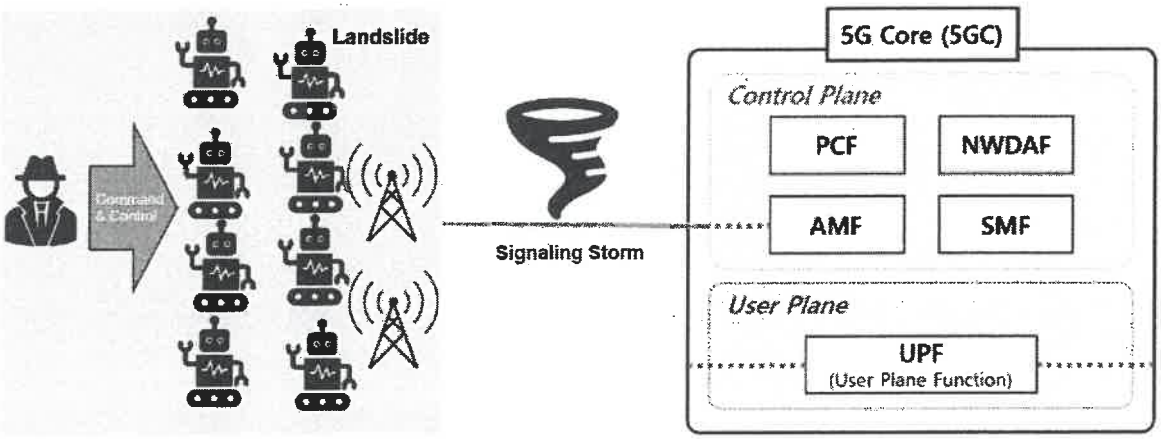


Hình 2.1: Tấn công DDoS trong mạng 5G sử dụng botnet IoT

2.1.2. Bão tín hiệu (Signaling Storm)

Bão tín hiệu 5G là một loại tấn công Từ chối dịch vụ phân tán (DDoS) khai thác *Mặt phẳng điều khiển* của mạng 5G bằng cách áp đảo nó với một lượng lớn các thông báo tín hiệu. Không giống như các cuộc tấn công truyền thống tràn ngập mặt phẳng dữ liệu bằng lưu lượng truy cập, các cơn bão tín hiệu nhắm vào các quy trình chịu trách nhiệm quản lý kết nối, tính di động và kiểm soát phiên, tạo ra tắc nghẽn tài nguyên trong mạng lõi. Khi các thiết bị mạng bị xâm nhập hoặc được cấu hình kém, liên tục gửi các yêu cầu kết nối, thiết lập phiên hoặc các tín hiệu liên quan đến tính di động, nó buộc mạng phải xử lý các tác vụ này liên tục. Điều này có thể dẫn đến tắc nghẽn đáng kể trong các thành phần cốt lõi như Chức năng quản lý truy cập và tính di động (AMF) hoặc Chức năng quản lý phiên (SMF), làm cạn kiệt tài nguyên tính toán và làm chậm các chức năng mạng hợp pháp. Việc cung cấp các tính năng nâng

cao trong 5G, chẳng hạn như phân chia mạng, ảo hóa thông qua Ảo hóa chức năng mạng (NFV) và khả năng kết nối IoT lớn, tạo ra các bề mặt tấn công bổ sung cho các cơn bão tín hiệu.



Hình 2.2: Bão tín hiệu tấn công mặt phẳng điều khiển của 5G Core

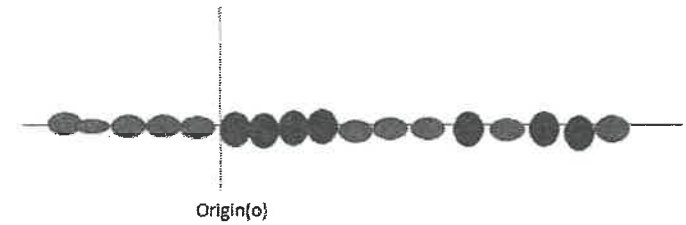
2.2. SVM trong phát hiện tấn công DDoS

2.2.1. Cơ bản về thuật toán SVM

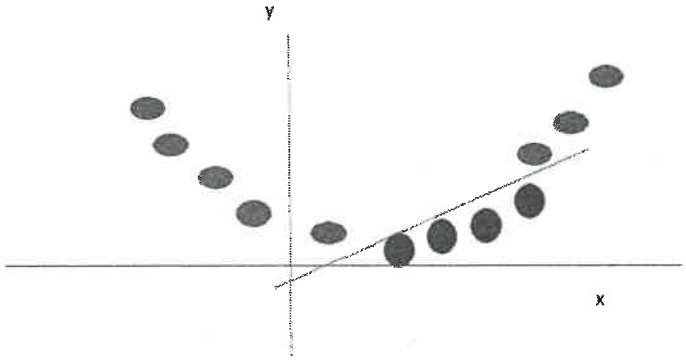
Với sự phát triển vượt trội của mạng di động 5G, nảy sinh nhu cầu cấp thiết trong việc phát triển các kỹ thuật học máy để phát hiện tấn công mạng. Các phương pháp phân loại như Rừng ngẫu nhiên (RF), K-láng giềng gần nhất (KNN) và Hồi quy Logistic (LR) có thể là những công cụ hữu ích để phát hiện một cuộc tấn công từ chối dịch vụ phân tán (DDoS).

Máy vectơ hỗ trợ (SVM) là một thuật toán học máy có giám sát, nó có thể sử dụng trong cả các tác vụ phân loại hoặc hồi quy. Tuy nhiên, trên thực tế SVM thường được sử dụng cho việc phân loại. SVM hướng đến việc tìm siêu mặt phẳng (hyper-plane) tối ưu trong không gian n-chiều để phân tách các điểm dữ liệu thành các lớp khác nhau. Thuật toán tối đa hóa biên độ giữa các điểm gần nhất của các lớp khác nhau. Trong đó, siêu mặt phẳng là ranh giới quyết định phân tách các lớp khác nhau trong không gian đặc điểm, được biểu diễn bằng phương trình $\mathbf{wx + b = 0}$ trong phân loại tuyến tính. Các vectơ hỗ trợ là những điểm dữ liệu gần nhất trỏ đến siêu mặt phẳng, chúng rất quan trọng để xác định siêu phẳng và biên trong SVM. Khoảng cách

giữa siêu mặt phẳng tới các vectơ hỗ trợ được gọi là biên độ (margin). SVM hướng đến mục tiêu tối đa hóa khoảng cách này để có hiệu suất phân loại tốt nhất. Ngoài phân loại tuyến tính, SVM cũng có thể sử dụng các hạt nhân (kernel) để thực hiện phân loại phi tuyến tính. Kernel là một hàm ánh xạ các điểm dữ liệu vào không gian có nhiều chiều hơn mà không cần tính toán rõ ràng tọa độ trong không gian đó. Điều này cho phép SVM hoạt động hiệu quả với dữ liệu phi tuyến tính bằng cách thực hiện ánh xạ ngầm định.



Bộ dữ liệu 1D gốc để phân loại



Ánh xạ dữ liệu 1D sang 2D để có thể tách biệt hai lớp

Hình 2.3: SVM sử dụng Kernel trong phân loại phi tuyến tính

Thuật toán này có ưu điểm là hoạt động tốt đối với những mẫu dữ liệu có kích thước lớn và thường mang lại kết quả vượt trội so với các thuật toán khác trong học có giám sát. Một số ưu điểm của SVM có thể kể đến như

- Đây là thuật toán hoạt động hiệu quả với không gian nhiều chiều (*high dimensional spaces*);

- Thuật toán tiêu tốn ít bộ nhớ vì chỉ sử dụng các điểm trong *tập hỗ trợ* để dự báo trong *hàm quyết định*;

- Có thể tạo ra nhiều *hàm quyết định* từ những hàm kernel khác nhau. Thậm chí sử dụng đúng kernel có thể giúp cải thiện thuật toán lên đáng kể.

2.2.2. Áp dụng vào phát hiện tấn công DDoS

Trong khuôn khổ phát hiện các cuộc tấn công DDoS, SVM phân loại lưu lượng truy cập giữa các gói tin bình thường và bất thường, đồng thời các cuộc tấn công có thể được phân loại dựa trên các đặc điểm lưu lượng truy cập như tốc độ cổng nguồn và độ lệch chuẩn của các gói tin luồng. SVM đã được sử dụng vì một lý do rất đơn giản nhưng hiển nhiên, đó là trên thực tế nó có cơ chế phát triển khi nhắc đến các thuật toán và các kỹ thuật dự đoán tĩnh của nó vượt trội trên hầu hết các khuôn khổ, chưa kể đến việc nó có thể đào tạo và học các thuật toán nhanh chóng và thông minh như thế nào trong cả năng lực tuyến tính và phi tuyến tính. SVM phổ biến hơn các kỹ thuật khác vì nó có thể tách không gian tuyến tính rất nhanh chóng khỏi không gian được cho là phi tuyến tính [12].

Tỷ lệ phát hiện các cuộc tấn công từ chối dịch vụ phân tán (DDoS) sử dụng thuật toán SVM nằm trong khoảng từ 93% đến 98% [12]. Vì vậy, xuất phát từ mức độ hiệu quả cao của các cuộc tấn công DDoS và khả năng cản trở các chức năng mạng thông thường của chúng là động lực chính để bắt đầu nghiên cứu này bằng cách sử dụng SVM. 5G tạo ra mức độ đe dọa bảo mật cao hơn chủ yếu là do có thêm các vectơ mà kẻ tấn công có thể lợi dụng. Một yếu tố khác cần xem xét ngữ cảnh này là khả năng của 5G trong việc cho phép nhiều thiết bị được kết nối với nhau, được gọi là Internet vạn vật (IoT). Vấn đề tổng thể là tìm cách tối đa hóa lợi ích của 5G và mặt khác giảm thiểu những rủi ro không mong muốn. Với các vấn đề an ninh mạng ngày càng tăng, phát sinh từ việc sử dụng mạng 5G, các phương pháp mới cần được áp dụng để giải quyết những vấn đề này.

Để phát hiện các cuộc tấn công mạng bằng việc áp dụng các thuật toán Học máy, cụ thể trong trường hợp này là kỹ thuật SVM, có thể phân ra thành hai giai đoạn thực hiện chính:

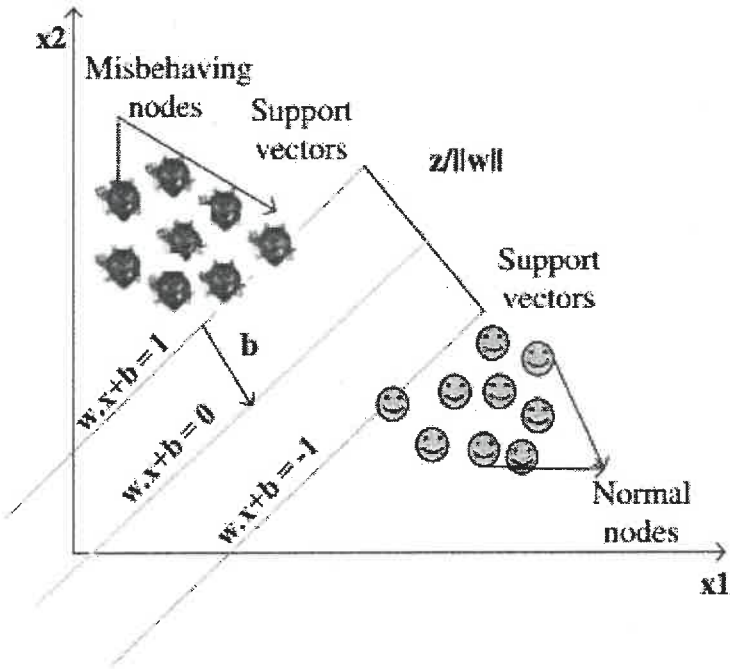
- Thu thập dữ liệu:

Từ môi trường mạng, các luồng gói tin khác nhau được thu thập vào các thời điểm khác nhau bằng những công cụ giám sát mạng. Từ đó, những thông tin như kích

thước gói tin, loại gói tin, thông tin máy chủ và giao thức… được trích xuất. Khi các luồng gói tin được thu thập, chúng được lưu trữ trong cơ sở dữ liệu.

- Phân tích gói tin để phát hiện tấn công:

Bước này đóng vai trò quan trọng trong việc xác định xem luồng gói tin đầu vào có phải là một cuộc tấn công hay không. Do tính chất động của các nút di động, luồng gói tin đầu vào thay đổi về kích thước và ngữ cảnh. Trong bước này, quá trình phát hiện xâm nhập diễn ra, phân tích các đặc điểm của gói tin và thực hiện hành động thích hợp. Các đặc tính gói tin khác nhau như địa chỉ IP của người gửi, địa chỉ IP đích, loại giao thức, số lượng gói tin đã gửi, thời gian liên lạc, loại gói tin và kích thước gói tin được sử dụng cho mục đích phân loại. Bằng cách áp dụng thuật toán SVM, có thể phân loại luồng gói tin ban đầu thành nhóm các gói tin thông thường và nhóm các gói tin bất thường.



Hình 2.4: Phát hiện tấn công sử dụng SVM

2.2.3. Ảnh hưởng của việc chọn Kernel và Tinh chỉnh tham số

Việc lựa chọn hàm kernel phù hợp và tinh chỉnh các tham số của thuật toán SVM đóng vai trò quan trọng, ảnh hưởng trực tiếp đến hiệu suất của mô hình, đặc biệt trong bài toán phát hiện tấn công DDoS.

Lựa chọn Kernel:

Như đã đề cập, SVM có thể sử dụng nhiều loại kernel khác nhau (ví dụ: Linear, Polynomial, Radial Basis Function - RBF, Sigmoid) để xử lý các bài toán phân loại tuyến tính và phi tuyến tính:

+ **Kernel Linear:** Phù hợp khi dữ liệu có thể phân tách tuyến tính. Đây là lựa chọn đơn giản và tính toán nhanh nhất;

+ **Kernel Polynomial:** Hiệu quả với dữ liệu không phân tách tuyến tính. Tham số độ (degree) của polynomial cần được lựa chọn cẩn thận; độ quá cao có thể dẫn đến overfitting;

+ **Kernel RBF - Radial Basic Function (Gaussian):** Là lựa chọn phổ biến và mạnh mẽ, có khả năng xử lý các mối quan hệ phức tạp trong dữ liệu. Kernel RBF có tham số gamma (γ) cần được tinh chỉnh. Giá trị γ lớn có thể dẫn đến overfitting, trong khi giá trị nhỏ có thể làm mô hình quá đơn giản (underfitting);

Việc lựa chọn kernel nào phụ thuộc vào đặc điểm của từng bộ dữ liệu. Thử nghiệm với các kernel khác nhau và đánh giá hiệu suất là cần thiết. Các nghiên cứu chỉ ra rằng việc lựa chọn kernel không phù hợp có thể làm giảm đáng kể hiệu quả của mô hình.

Tinh chỉnh tham số SVM:

Tham số C (Regularization parameter - Tham số điều chuẩn):

+ Tham số C quyết định sự đánh đổi giữa việc tối đa hóa khoảng cách lề (margin) và việc giảm thiểu lỗi phân loại trên tập huấn luyện;

+ Giá trị C lớn: Mô hình sẽ cố gắng phân loại chính xác tất cả các điểm dữ liệu huấn luyện, có thể dẫn đến lề hẹp hơn và nguy cơ overfitting (mô hình hoạt động tốt trên dữ liệu huấn luyện nhưng kém trên dữ liệu mới;

+ Giá trị C nhỏ: Mô hình cho phép một số điểm dữ liệu huấn luyện bị phân loại sai để đạt được lề rộng hơn, giúp tăng khả năng tổng quát hóa nhưng có thể dẫn đến underfitting (mô hình không đủ phức tạp để nắm bắt quy luật của dữ liệu).

Tham số Gamma (γ) (Đối với Kernel RBF):

+ Tham số γ xác định mức độ ảnh hưởng của một điểm dữ liệu huấn luyện;

+ Giá trị γ lớn: Ảnh hưởng của mỗi điểm dữ liệu chỉ giới hạn trong phạm vi hẹp, mô hình có thể trở nên quá nhạy cảm với nhiễu và dẫn đến overfitting;

+ Giá trị γ nhỏ: Ảnh hưởng của mỗi điểm dữ liệu lan rộng hơn, mô hình có thể trở nên quá tổng quát và không nắm bắt được các chi tiết quan trọng, dẫn đến underfitting.

Các tham số khác: Tùy thuộc vào loại kernel được chọn (ví dụ: degree cho kernel Polynomial, coef0 cho kernel Polynomial và Sigmoid).

Phương pháp tinh chỉnh: Các kỹ thuật như Grid Search kết hợp với Cross-Validation (Kiểm định chéo) thường được sử dụng để tìm ra tổ hợp tham số tối ưu nhất cho mô hình SVM. Quá trình này bao gồm việc thử nghiệm nhiều giá trị khác nhau cho từng tham số và chọn tổ hợp mang lại hiệu suất tốt nhất trên tập dữ liệu kiểm định.

2.3. Áp dụng SVM vào thực tế với dữ liệu cực lớn của mạng 5G

Việc triển khai SVM để phát hiện tấn công DDoS trong môi trường mạng 5G, với đặc thù là lượng dữ liệu cực lớn và yêu cầu xử lý theo thời gian thực, đặt ra nhiều thách thức đáng kể:

2.3.1. Thách thức

+ **Khối lượng dữ liệu (Big Data):** Mạng 5G tạo ra một lượng dữ liệu khổng lồ với tốc độ rất cao. Việc thu thập, lưu trữ và xử lý luồng dữ liệu này để huấn luyện và vận hành mô hình SVM là một thách thức lớn. Thời gian huấn luyện mô hình SVM có thể rất lâu với các bộ dữ liệu lớn;

+ **Yêu cầu xử lý thời gian thực:** Hệ thống phát hiện tấn công DDoS cần phải đưa ra cảnh báo gần như ngay lập tức để có biện pháp đối phó kịp thời. Độ trễ trong

quá trình phát hiện có thể làm giảm hiệu quả của hệ thống. Các minh chứng thực tế cũng chỉ ra "Độ trễ có thể tăng khi xử lý đồng thời nhiều luồng dữ liệu lớn", đây là một hạn chế cần lưu ý đối với mạng 5G;

+ **Tính đa dạng và động của lưu lượng mạng 5G:** Lưu lượng mạng 5G rất đa dạng, bao gồm nhiều loại dịch vụ và ứng dụng khác nhau (eMBB, URLLC, mMTC). Các mẫu tấn công DDoS cũng ngày càng tinh vi và biến đổi liên tục. Điều này đòi hỏi mô hình SVM phải có khả năng thích ứng và học hỏi nhanh chóng;

+ **Tài nguyên tính toán:** Việc huấn luyện và triển khai các mô hình SVM, đặc biệt với các kernel phức tạp trên dữ liệu lớn, đòi hỏi tài nguyên tính toán đáng kể (CPU, bộ nhớ). Điều này có thể là một rào cản, nhất là đối với các thiết bị biên hoặc các thành phần mạng có tài nguyên hạn chế;

+ **Chọn lọc đặc trưng (Feature Selection):** Với số lượng lớn các đặc trưng tiềm năng từ dữ liệu mạng 5G, việc chọn ra các đặc trưng quan trọng và hiệu quả cho việc phát hiện DDoS là rất quan trọng để giảm độ phức tạp tính toán và cải thiện độ chính xác của mô hình.

2.3.2. Giải pháp và hướng tiếp cận

+ **Kỹ thuật giảm chiều dữ liệu:** Sử dụng các kỹ thuật như Phân tích thành phần chính (PCA), Kernel PCA (KPCA) để giảm số lượng đặc trưng mà vẫn giữ lại được thông tin quan trọng, từ đó có thể giảm thời gian huấn luyện và độ phức tạp của mô hình;

+ **Học tăng cường (Incremental Learning) và Học trực tuyến (Online Learning):** Cho phép mô hình SVM cập nhật và học hỏi từ các luồng dữ liệu mới mà không cần phải huấn luyện lại toàn bộ mô hình từ đầu. Điều này rất quan trọng để thích ứng với sự thay đổi của các mẫu tấn công và lưu lượng mạng.

+ **Kiến trúc phân tán và tính toán song song:** Xử lý dữ liệu và huấn luyện mô hình trên các hệ thống phân tán (ví dụ: sử dụng *Apache Spark*) để tăng tốc độ;

+ **SVM tối ưu hóa:** Nghiên cứu và áp dụng các biến thể SVM được tối ưu hóa cho dữ liệu lớn hoặc các phương pháp giải gần đúng (*Approximate solvers*);

+ **Kết hợp với các công nghệ khác:** Tích hợp SVM với các công nghệ xử lý dữ liệu lớn (Big Data platforms) và các hệ thống giám sát mạng tiên tiến. Một số nghiên cứu đề xuất kết hợp SVM với các thuật toán học máy khác hoặc các mô hình học sâu để tăng cường khả năng phát hiện;

+ **Sử dụng phần cứng chuyên dụng:** Xem xét việc sử dụng các giải pháp tăng tốc phần cứng (ví dụ: GPU, FPGA) cho các tác vụ tính toán nặng của SVM;

+ **Lấy mẫu dữ liệu thông minh (Intelligent Data Sampling):** Thay vì sử dụng toàn bộ dữ liệu lớn để huấn luyện, có thể áp dụng các kỹ thuật lấy mẫu thông minh để chọn ra một tập dữ liệu đại diện, giúp giảm thời gian huấn luyện mà vẫn đảm bảo hiệu suất.

2.4. Phương án triển khai phát hiện xâm nhập bằng SVM

Như đã trình bày ở trên, việc lựa chọn kernel, tinh chỉnh tham số của SVM hay kết hợp với các kỹ thuật khác ảnh hưởng rất nhiều tới hiệu quả của việc phát hiện tấn công, đặc biệt với lượng dữ liệu vô cùng lớn của 5G. Trong đề án này, tác giả thực hiện nghiên cứu trên 2 phương pháp phát hiện xâm nhập mạng:

- Phương pháp 1: Sử dụng học máy truyền thống (SVM) và tối ưu hóa bằng cách loại bỏ các đặc trưng không cần thiết thông qua thuật toán RFE (*Recursive Feature Elimination*); RFE là phương pháp lựa chọn đặc trưng theo kiểu "bao bọc", hoạt động bằng cách huấn luyện mô hình và loại bỏ dần các đặc trưng ít quan trọng nhất cho đến khi còn lại tập đặc trưng tối ưu.

- Phương pháp 2: Sử dụng học máy truyền thống (SVM) kết hợp lựa chọn đặc trưng dựa trên phân tích thống kê (SelectKBest) và xử lý mất cân bằng dữ liệu (SMOTE - *Synthetic Minority Over-sampling Technique*). SelectKBest là phương pháp lựa chọn đặc trưng sử dụng các bài kiểm tra thống kê (như ANOVA F-test) để đánh giá từng đặc trưng độc lập và chọn ra K đặc trưng có điểm số cao nhất. SMOTE là kỹ thuật xử lý mất cân bằng dữ liệu bằng cách tạo ra các mẫu tổng hợp mới cho lớp thiểu số, giúp cân bằng số lượng mẫu giữa các lớp trong tập huấn luyện.

Bảng 2.1: Tiền xử lý dữ liệu

Hạng mục	Phương pháp 1 (SVM + RFE)	Phương pháp 2 (SVM + SelectKBest + SMOTE)
Xử lý giá trị thiếu	Các giá trị số bị thiếu được điền bằng **giá trị trung vị (median)** của cột tương ứng. Các cột dạng chữ được mã hóa bằng *pd.get_dummies*.	Tương tự, cũng điền giá trị số thiếu bằng **median** và mã hóa các cột dạng chữ.
Chuẩn hóa dữ liệu	Sử dụng *StandardScaler* để chuẩn hóa dữ liệu (đưa về trung bình 0, độ lệch chuẩn 1) như một phần của pipeline huấn luyện.	Cũng sử dụng *StandardScaler* trong pipeline.
Bước đặc biệt	Không có	Có một bước làm sạch bổ sung: Loại bỏ các cột có phương sai gần bằng 0. Bước này giúp loại bỏ các đặc trưng không đổi hoặc gần như không đổi, vốn không mang lại thông tin cho mô hình.

Bảng 2.2: Lựa chọn đặc trưng

Hạng mục	Phương pháp 1 (SVM + RFE)	Phương pháp 2 (SVM + SelectKBest + SMOTE)
Thuật toán	Recursive Feature Elimination (RFE).	SelectKBest.
Bản chất	Là một phương pháp "bao bọc" (wrapper method), dựa trên hiệu suất của mô hình. Nó đánh giá tầm quan trọng của đặc trưng trong mối tương quan với các đặc trưng khác.	Là một phương pháp "lọc" (filter method), dựa trên phân tích thống kê. Nó đánh giá từng đặc trưng một cách độc lập.
Quy trình hoạt động	1. Sử dụng một mô hình ước tính (*LogisticRegression*) để huấn luyện trên toàn bộ các đặc trưng. 2. Xếp hạng tầm quan trọng của các đặc trưng. 3. Loại bỏ một số đặc trưng yếu nhất. 4. Lặp lại quá trình cho đến khi chỉ còn lại số lượng đặc trưng mong muốn (**25 đặc trưng**).	1. Sử dụng một bài kiểm tra thống kê (**f_classif - ANOVA F-test**) để tính điểm cho từng đặc trưng một cách độc lập. 2. Chọn ra 20 đặc trưng có điểm số cao nhất.

Ưu điểm	- Có xu hướng chọn ra được một bộ đặc trưng tối ưu hơn vì nó xem xét sự tương tác giữa chúng. - Rất hiệu quả trong việc tìm ra các đặc trưng thực sự quan trọng đối với mô hình cụ thể.	- Rất nhanh về mặt tính toán vì không cần phải huấn luyện lại mô hình nhiều lần. - Đơn giản và dễ triển khai.
Nhược điểm	- Tốn nhiều thời gian tính toán hơn đáng kể.	- Có thể bỏ lỡ các đặc trưng có giá trị khi chúng kết hợp với nhau, vì nó chỉ xem xét chúng một cách riêng lẻ.

Bảng 2.3: Xử lý mất cân bằng dữ liệu

Hạng mục	Phương pháp 1 (SVM + RFE)	Phương pháp 2 (SVM + SelectKBest + SMOTE)
Phương pháp chính	Sử dụng tham số *class_weight='balanced'* trong mô hình SVC.	Sử dụng kỹ thuật **SMOTE**
Cách hoạt động	Điều chỉnh trọng số của hàm mất mát. Mô hình sẽ bị "phạt" nặng hơn nếu dự đoán sai một mẫu thuộc lớp thiểu số. Phương pháp này không làm thay đổi dữ liệu huấn luyện.	Tạo ra dữ liệu tổng hợp. SMOTE sẽ tạo ra các mẫu dữ liệu mới cho các lớp thiểu số trong tập huấn luyện, làm cho số lượng mẫu giữa các lớp trở nên cân bằng hơn một cách vật lý.
Pipeline	Sử dụng Pipeline tiêu chuẩn của scikit-learn.	Bắt buộc phải sử dụng ImbPipeline của thư viện imblearn. Điều này đảm bảo rằng SMOTE chỉ được áp dụng trên dữ liệu huấn luyện trong mỗi bước của quá trình cross-validation, tránh rò rỉ dữ liệu.

Bảng 2.4: Tinh chỉnh tham số thuật toán

Hạng mục	Phương pháp 1 (SVM + RFE)	Phương pháp 2 (SVM + SelectKBest + SMOTE)
Công cụ	GridSearchCV	RandomizedSearchCV
Chiến lược	**Tìm kiếm toàn diện** (*Exhaustive Search*): Thử tất cả các kết hợp có thể có của các siêu tham số được cung cấp (C và gamma) để tìm ra bộ kết hợp tốt nhất tuyệt đối.	**Tìm kiếm ngẫu nhiên** (*Random Search*): Thử một số lượng hữu hạn (n_iter = 10) các bộ tham số được chọn ngẫu nhiên từ không gian tìm kiếm.
Đánh đổi	- Ưu điểm: Đảm bảo tìm ra được giải pháp tối ưu trong lưới tìm kiếm. - Nhược điểm: Rất tốn thời gian, đặc biệt với không gian tham số lớn.	- Ưu điểm: Nhanh hơn đáng kể so với GridSearchCV. - Nhược điểm: Có thể không tìm ra được bộ tham số tốt nhất tuyệt đối, nhưng thường tìm được một giải pháp đủ tốt.

Phương pháp 1 (SVM + RFE) đã chứng tỏ sự vượt trội về mặt hiệu suất. Sự thành công này có thể được quy cho hai yếu tố chính:

- Lựa chọn đặc trưng bằng RFE: Phương pháp này đã chọn ra một bộ 25 đặc trưng chất lượng cao, có khả năng phân loại tốt khi được xem xét cùng nhau;

- Sự đơn giản và hiệu quả: Chỉ cần sử dụng class_weight='balanced' đã đủ để xử lý mất cân bằng khi kết hợp với bộ đặc trưng tốt, mà không cần đến kỹ thuật phức tạp hơn như SMOTE.

Phương pháp 2 (SVM + SelectKBest + SMOTE) là một cách tiếp cận hợp lý và nhanh hơn về mặt tính toán, nhưng bộ đặc trưng được chọn bởi SelectKBest có thể không tối ưu bằng, dẫn đến hiệu suất thấp hơn một chút, đặc biệt là ở khả năng nhận diện một số lớp cụ thể như SYNFlood.

Các công trình được trích dẫn trong đề án cũng tập trung vào việc sử dụng học máy để giải quyết các vấn đề bảo mật 5G, nhưng có thể có những điểm khác biệt về phương pháp tiếp cận chi tiết hoặc phạm vi:

- Hai phương án này tập trung sâu hơn vào việc áp dụng cụ thể thuật toán SVM cho bài toán phát hiện tấn công DDoS và đi sâu vào các kỹ thuật tiền xử lý dữ liệu (RFE, SelectKBest, SMOTE) để tối ưu hóa hiệu suất trên bộ dữ liệu 5G-NIDD. Trong

khi nghiên cứu của Fakhouri [6] cung cấp cái nhìn tổng thể về nhiều ứng dụng và thuật toán ML, đề án này đi vào chi tiết hơn về các phương pháp tối ưu hóa cho một thuật toán cụ thể (SVM) và một loại tấn công cụ thể (DDoS) trong môi trường 5G;

- Việc giới thiệu hai phương pháp (SVM + RFE và SVM + SelectKBest + SMOTE) thể hiện sự nỗ lực trong việc tìm ra cách tiếp cận tối ưu nhất, đặc biệt là trong việc xử lý mất cân bằng dữ liệu và lựa chọn đặc trưng, điều mà công trình của Alshunaifi [12] có thể chưa đi sâu vào chi tiết các kỹ thuật tối ưu hóa này.

2.5. Kết luận chương

Chương 2 đã phân tích chi tiết các hình thức tấn công DDoS trong mạng di động 5G, đặc biệt là qua mạng botnet IoT và bão tín hiệu, vốn gây ảnh hưởng nghiêm trọng đến mặt phẳng điều khiển và tài nguyên mạng. Thuật toán SVM được lựa chọn làm công cụ phát hiện tấn công nhờ khả năng phân loại mạnh mẽ và hiệu quả cao trong môi trường dữ liệu lớn. Việc áp dụng SVM đòi hỏi phải xử lý tốt các vấn đề như lựa chọn kernel, tinh chỉnh tham số và cân bằng dữ liệu để đạt hiệu suất tối ưu. Tác giả đề xuất hai phương pháp triển khai phát hiện tấn công DDoS trong mạng di động 5G là: SVM kết hợp RFE và SVM kết hợp SelectKBest với SMOTE có thể cải thiện độ chính xác và khả năng phát hiện các loại tấn công khác nhau. Chương này khẳng định rằng SVM là một giải pháp khả thi và hiệu quả trong việc phát hiện sớm các cuộc tấn công DDoS trên mạng 5G.

CHƯƠNG 3: THỰC NGHIỆM, ĐÁNH GIÁ KẾT QUẢ

3.1. Mô hình triển khai

3.1.1. Các chỉ số đánh giá

Trước khi trình bày các chỉ số đánh giá hệ thống, chúng ta cần thống nhất một số ký hiệu chung:

- **P (*positive*)**: chỉ định dương tính, tức là một phần tử thuộc lớp được phát hiện. Trong trường hợp phát hiện DDoS, nó đề cập đến số lượng các gói tin độc hại;

- **N (*negative*)**: chỉ định âm tính, tức là một phần tử thuộc lớp đối diện. Trong trường hợp này, đây là số lượng các gói tin bình thường;

- Dương tính thực (*TP - True positives*), Âm tính thực (*TN – True negative*): số lượng dự đoán chính xác của các lớp dương tính và âm tính, tương ứng (các giá trị dự đoán và thực tế đều âm tính);

- Dương tính giả (*FP – False positives*), âm tính giả (*FN – False negative*): hai số liệu này lần lượt là số lượng các phần tử được xác định không chính xác là dương tính và âm tính;

- Tỷ lệ dương tính giả (*FPR - False positive rate*): tỷ lệ âm tính được báo cáo là dương tính trên tổng số mẫu $FPR = \frac{FP}{N}$. Giá trị này nằm trong khoảng [0, 1];

- Tỷ lệ âm tính thực (*TNR - True negative rate*): tỷ lệ giữa các phần tử âm tính được xác định chính xác và tổng số phần tử âm tính $TNR = \frac{TN}{N}$. Giá trị này nằm trong khoảng [0, 1];

- Tỷ lệ âm tính giả (*FNR - False negative rate*): tỷ lệ giữa các phần tử dương tính được báo cáo là âm tính và tổng số phần tử dương tính $FNR = \frac{FN}{P}$. Giá trị này nằm trong khoảng [0, 1].

Các chỉ số đánh giá một hệ thống phân loại:

- Độ chính xác (*Accuracy*) là tỷ lệ các mẫu được phân loại chính xác. Nói cách khác, nó là tỷ lệ các mục được xác định chính xác trên tổng kích thước của tập kiểm tra $Acc = \frac{TP+TN}{P+N}$. Giá trị này nằm trong khoảng [0, 1];

- Độ chính xác (*Precision*) còn được gọi là giá trị dự đoán dương tính vì nó trả về tỷ lệ các mẫu có liên quan trong số các mẫu được phát hiện. Nó là tỷ lệ các phần tử dương tính được xác định chính xác là dương tính trên số lượng tất cả các phần tử được xác định là độ chính xác dương tính $Precision = \frac{TP}{TP+FP}$. Giá trị này nằm trong khoảng [0, 1];

- Nhắc lại (Recall) còn được gọi là độ nhạy hoặc *Tỷ lệ dương tính thực*, đó là tỷ lệ các mẫu có liên quan được phát hiện là dương tính. Nó là tỷ lệ các phần tử dương tính được xác định chính xác trên tổng số phần tử dương $Recall = TPR = \frac{TP}{TP+FN}$. Giá trị này nằm trong khoảng [0, 1];

- Điểm F1 đo lường trung bình điều hòa của độ chính xác và nhắc lại $F1 = \frac{2*Precision*Recall}{Precision+Recall}$. Giá trị này nằm trong khoảng [0, 1] và giá trị càng cao thì kết quả càng tốt. Đây là một số liệu đáng tin cậy để đánh giá các bộ dữ liệu không cân bằng;

- Ma trận nhầm lẫn (*Confusion Matrix*) là thước đo việc phân loại trong học máy đầu ra có thể là hai hoặc nhiều lớp. Đối với phân loại nhị phân, nó là một ma trận với 4 tổ hợp khác nhau của các giá trị dự đoán và thực tế. Việc cung cấp các chỉ số cụ thể về hiệu suất và khả năng phản hồi của hệ thống là vô cùng quan trọng.

		Model dự đoán	
		Positive	**Negative**
Thực tế	**Positive**	True Positive (Dự đoán Đúng là Positive)	False Negative (Dự đoán Sai là Negative)
	Negative	False Positive (Dự đoán Sai là Positive)	True Negative (Dự đoán Đúng là Negative)

Hình 3.1: Ví dụ về Confusion matrix

3.1.2. Bộ dữ liệu mẫu

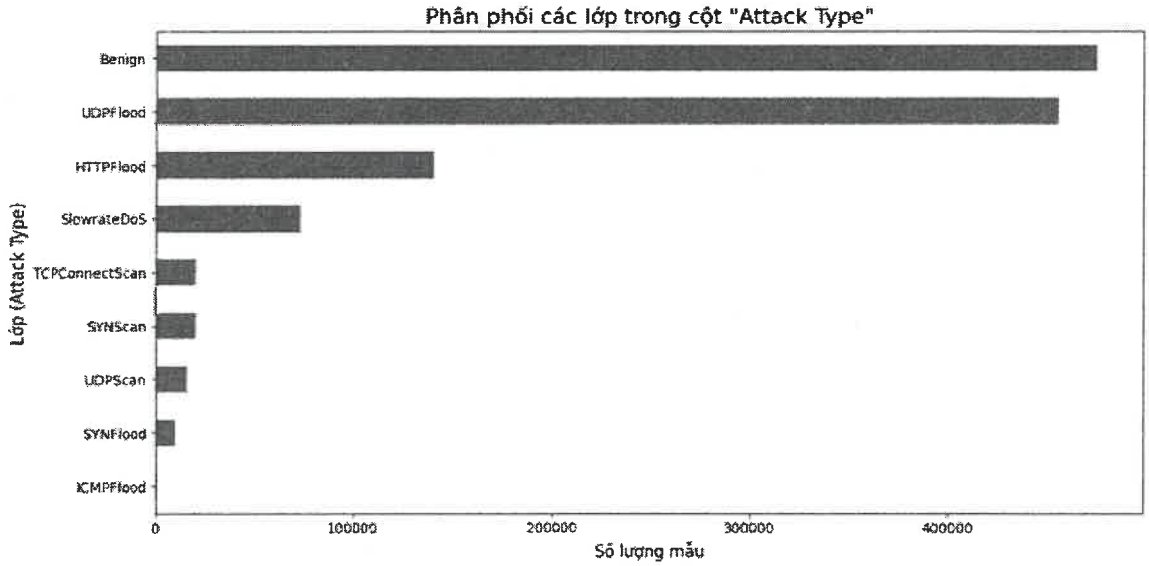
Quá trình thu thập dữ liệu có tầm quan trọng hàng đầu trong thiết kế của một hệ thống phát hiện xâm nhập. Trên thực tế, phân tích dữ liệu tiết lộ hành vi bất thường và cung cấp các chỉ số liên quan có thể giúp thiết kế các hệ thống phát hiện xâm nhập. Ngoài ra, chất lượng của dữ liệu có ảnh hưởng trực tiếp đến hiệu suất của các thuật toán phát hiện, đặc biệt là các thuật toán dựa trên học máy. Dữ liệu này có thể được lấy từ các mạng chuyên dụng trong thế giới thực hoạt động với quyền truy cập hạn chế hoặc các mạng tạm thời và hệ thống thử nghiệm được triển khai đặc biệt để thu thập bộ dữ liệu.

Trong phạm vi nghiên cứu của đề án, tác giả đề xuất sử dụng bộ dữ liệu 5G-NIDD làm bộ dữ liệu mẫu, dùng để thử nghiệm các phương pháp đã trình bày. Bộ dữ liệu 5G-NIDD là bộ dữ liệu xâm nhập mẫu mới nhất, có nguồn gốc từ một mạng thử nghiệm 5G thực [3]. Nó được tạo ra bằng cách sử dụng một mạng 5G đầy đủ chức năng được tạo ra cho các mục đích thử nghiệm 5G, giống với một mạng thực tế. 5G-NIDD bao gồm các tính năng duy nhất, đặc trưng cho các luồng mạng 5G.

Bảng 3.1: Các loại tấn công DoS trong bộ dữ liệu

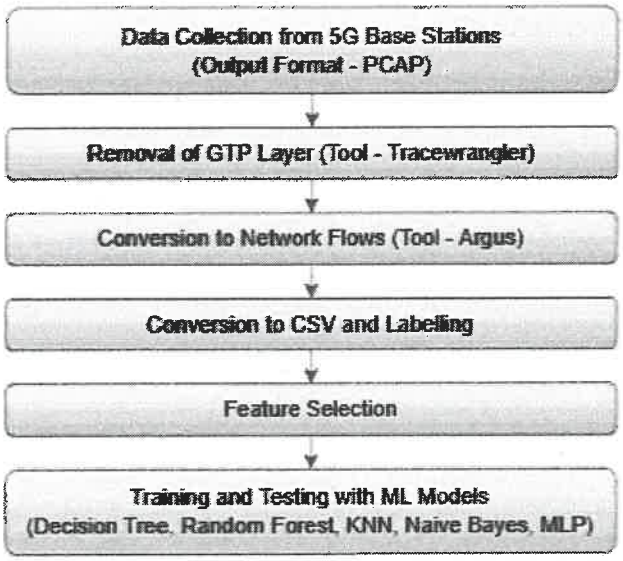
Loại tấn công	Mô tả	Công cụ sử dụng
Goldeneye	Giả mạo hành vi người dùng và gửi các yêu cầu HTTP khiến cho máy chủ web không thể xử lý các yêu cầu hợp pháp.	sudo python3 ./goldeneye.py http://"Target IP address of the web server"
ICMP Flood	ICMP là giao thức dùng để gửi các thông điệp điều khiển và chuẩn đoán mạng, ví dụ ping. Attacker gửi hàng loạt gói tin ICMP, mục tiêu sẽ phản hồi lại từng gói và nguy cơ nghẽn mạng.	sudo hping3 –flood –rand-source -1 -p "Targeted port/Range of ports" "Target IP address"
Slowloris	Gửi các gói tin với tốc độ chậm giả mạo hành vi kết nối kém của người dùng. Attacker duy trì kết nối không hoàn thành đến máy chủ trong một thời gian dài khiến suy giảm tài nguyên hệ thống.	python3 slowloris.py "Target IP address of the web server"

Loại tấn công	Mô tả	Công cụ sử dụng
SSH	Attacker tấn công SYNflood dựa vào cơ chế TCP với SSH	
SYN Flood	Lợi dụng cơ chế handshake của TCP. Attacker gửi hàng loạt gói SYN và không gửi ACK để hoàn tất handshake, buộc mục tiêu phải chờ timeout cho mỗi kết nối.	sudo hping3 -S -p "Targeted port/range of ports" –flood –rand-source "Target IP address"
Torhammer	Attacker lợi dụng mạng Tor để thực hiện tấn công Slowloris	python2 torshammer.py -t "Target IP address of the web server"
UDP Flood	UDP là connection-less và gửi hàng loạt gói tin UDP tới port ngẫu nhiên của mục tiêu. Nếu port không listen thì mục tiêu sẽ phản hồi bằng gói ICMP và làm nghẽn mạng.	sudo hping3 –flood –rand-source –udp -p "Targeted port/range of ports" "Target IP address"



Hình 3.2: Phân phối các lớp trong cột "Attack Type"

Quá trình thu thập và xử lý dữ liệu của 5G-NIDD: Dữ liệu được thu thập ghi lại lưu lượng tấn công và lưu lượng bình thường khi đi qua 2 trạm RAN. Mỗi loại tấn công được ghi lại trong các phiên riêng biệt dưới dạng file .pcap, sau đó các file .pcap được đưa vào trong các quy trình xử lý.



Hình 3.3: Quy trình thu thập, xử lý dữ liệu mẫu 5G-NIDD

1, Loại bỏ lớp GTP: Vì dữ liệu đi qua giao diện vô tuyến nên chứa lớp GTP-U. Loại bỏ lớp GTP là cần thiết để phân tích chính xác. Sử dụng Tracewrangler;

2, Chuyển đổi sang luồng mạng: Giao thức Argus được dùng để chuyển dữ liệu từ dạng gói sang luồng mạng. Mỗi luồng có IP, port, protocol,... Điều này làm giảm kích thước dữ liệu và thuận tiện hơn cho mô hình học máy;

3, Gộp dữ liệu và dán nhãn: Dữ liệu được chuyển sang file CSV và án nhãn. Gộp dữ liệu từ các phiên theo từng trạm RAN sau đó kết hợp lại. Kết quả cuối cùng: ***1.215.890 dòng*** dữ liệu luồng mạng.

4, Mã hóa dữ liệu: Một số đặc trưng để là dữ liệu phân loại (categorical), không thể trực tiếp đưa vào mô hình ML. Vì vậy, cần dùng kỹ thuật One-hot Encoding để chuyển thành số nhị phân.

Bảng 3.2: Các đặc trưng trong bộ dữ liệu mẫu

Tên đặc trưng	Mô tả
Flgs	Flags TCP (SYN, ACK, RST, FIN). Trong ICMP thường là 'e' (echo)
Seq	Sequence Number – Số thứ tự gói tin trong chuỗi hoặc kết nối
Dur	Duration – Thời lượng của kết nối hoặc phiên (tính bằng giây)

Tên đặc trưng	Mô tả
Runtime	
Mean, Sum, Min, Max	Chỉ số thống kê một số đặc trưng (kích thước gói, thời gian trễ, hoặc tốc độ)
Proto	Protocol – Giao thức truyền tải (TCP, UDP, ICMP, v.v)
sTos, dTos	Type of Service (ToS) – Trong IP header, xác định độ ưu tiên của gói tin
sDSb, sDSb	DSCP bits (Differentiated Services Code Point) – Định tuyến ưu tiên trong mạng
sTtl, dTtl	Time to Live – Thời gian sống của gói tin (đếm số lượng hops trước khi bị loại bỏ)
sHops, dHops	Số lượng hops (nút) mà gói tin đi qua từ nguồn đến đích
Cause	Nguyên nhân và sự kiện xảy ra
State	Trạng thái kết nối TCP (ESTABLISHED, FIN_WAIT, ECO = echo)
TotPkts, SrcPkts, DstPkts	Tổng số gói tin, gói từ nguồn, gói từ đích
TotBytes, SrcBytes, DstBytes	Tổng số byte, byte từ nguồn, byte từ đích
Offset	Độ lệch thời gian giữa các gói hoặc phiên
sMeanPktSz, dMeanPktSz	Kích thước trung bình gói tin từ nguồn/đích
Load, SrcLoad, DstLoad	Tải mạng tổng, từ nguồn, từ đích (byte/s, packet/s)
Loss, SrcLoss, DstLoss, pLoss	Gói tin bị mất (tổng, nguồn, đích, phần trăm)
SrcGap, DstGap	Khoảng cách giữa các gói tin liên tiếp từ nguồn/đích (đơn vị thời gian)
Rate, SrcRate, DstRate	Tốc độ truyền tải tổng, từ nguồn và đích (packet hoặc byte/s)
SrcWin, DstWin	TCP Window Size – cửa sổ điều khiển luồng
sVid, dVid	VLAN ID nếu có sử dụng VLAN
SrcTCPBase, DstTCPBase	Base Sequence Number cho TCP (ban đầu)
TcpRtt	Round Trip Time – thời gian phản hồi của TCP
SynAck	Thời gian giữa SYN và ACK
AckDat	Thời gian giữa ACK và dữ liệu thực tế truyền sau đó

3.1.3. Kiến trúc hệ thống

Bộ dữ liệu mẫu 5G-NIDD cung cấp rất nhiều đặc trưng khác nhau trong các dữ liệu mạng (Bảng 3.3). Tuy nhiên, trong phạm vi nghiên cứu thực nghiệm, tác giả chỉ tập trung vào một số đặc trưng cơ bản:

- Thông số gói tin (TotPkts, SrcPkts, DstPkts): Lưu lượng gói tin;
- Thông số byte (TotBytes, SrcBytes, DstBytes): Kích thước dữ liệu;
- Tốc độ truyền (Rate, SrcRate, DstRate);
- Tải mạng (Load, SrcLoad, DstLoad);
- Thời gian (Mean, Dur);
- Thông tin giao thức (Proto, Sport, Dport).

Các đặc trưng được chọn dựa trên tần suất xuất hiện lớn trong bộ dữ liệu. Việc chọn ra các đặc trưng là để tập trung vào những đặc trưng xuất hiện liên tục và liên quan đến lưu lượng, kích thước, thông tin, thời gian, tốc độ, tỷ lệ truyền tải mạng, giúp mô hình học máy phán đoán chính xác hơn và tránh overfitting.

Tiền xử lý dữ liệu:

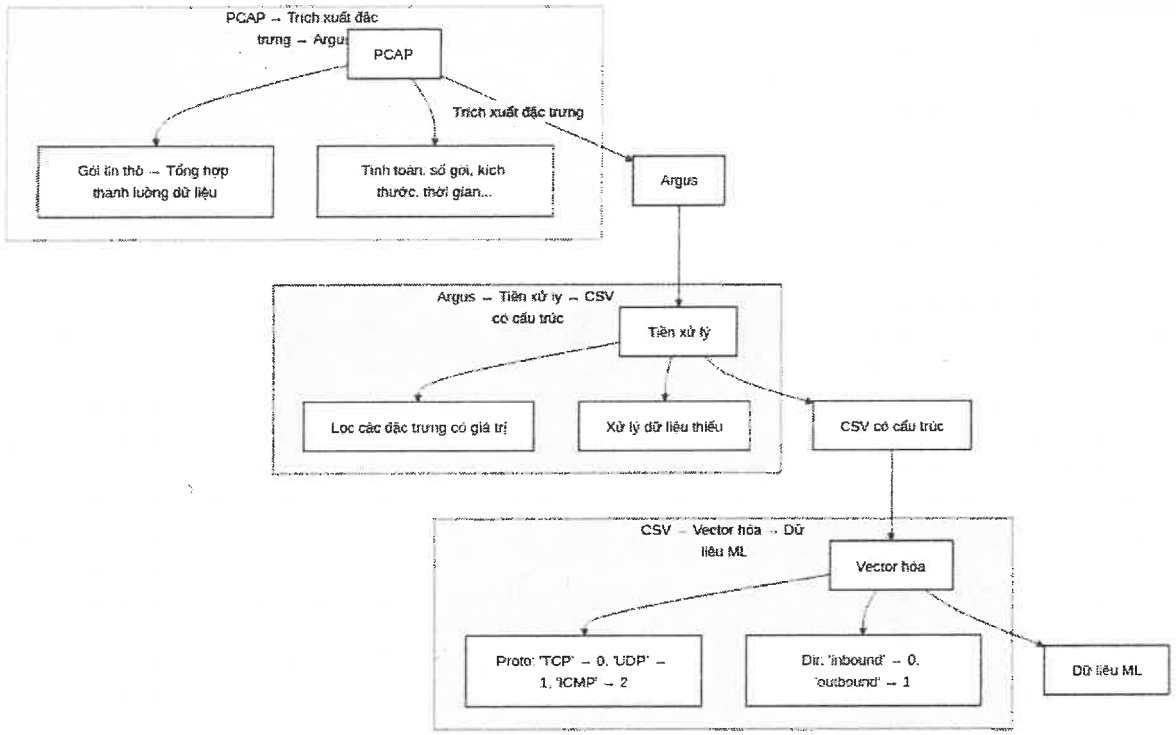
Dữ liệu gốc, thô là dữ liệu log của luồng mạng như PCAP, được thu thập trực tiếp từ mạng 5G. Cần áp dụng các kỹ thuật tiền xử lý dữ liệu như vector hóa, vector reasoning, và chuẩn hóa.

Các bước xử lý dữ liệu thô (PCAP sang CSV):

1. Chuyển PCAP thành định dạng **Argus**: *argus –r file.pcap –w output.argus*

2. Chuyển Argus thành CSV với các trường mong muốn: *ra -r file.argus -s stime dur proto saddr sport daddr dport state pkts bytes rate spkts dpkts sbytes dbytes srate drate load sload dload -c , > raw_flows.csv*

Vector hóa và Chuẩn hóa: Vector hóa các cột phân loại như Proto, State từ dạng chữ thành vectơ số. Mã hóa dữ liệu phân loại bằng LabelEncoder (ví dụ: Proto, State, Dir). Chuẩn hóa dữ liệu số bằng StandardScaler để đưa các đặc trưng về cùng một khoảng giá trị, giúp tăng tốc độ hội tụ và cải thiện hiệu suất mô hình. Xử lý giá trị thiếu và bất thường.



Hình 3.4: Chuẩn hóa dữ liệu đầu vào

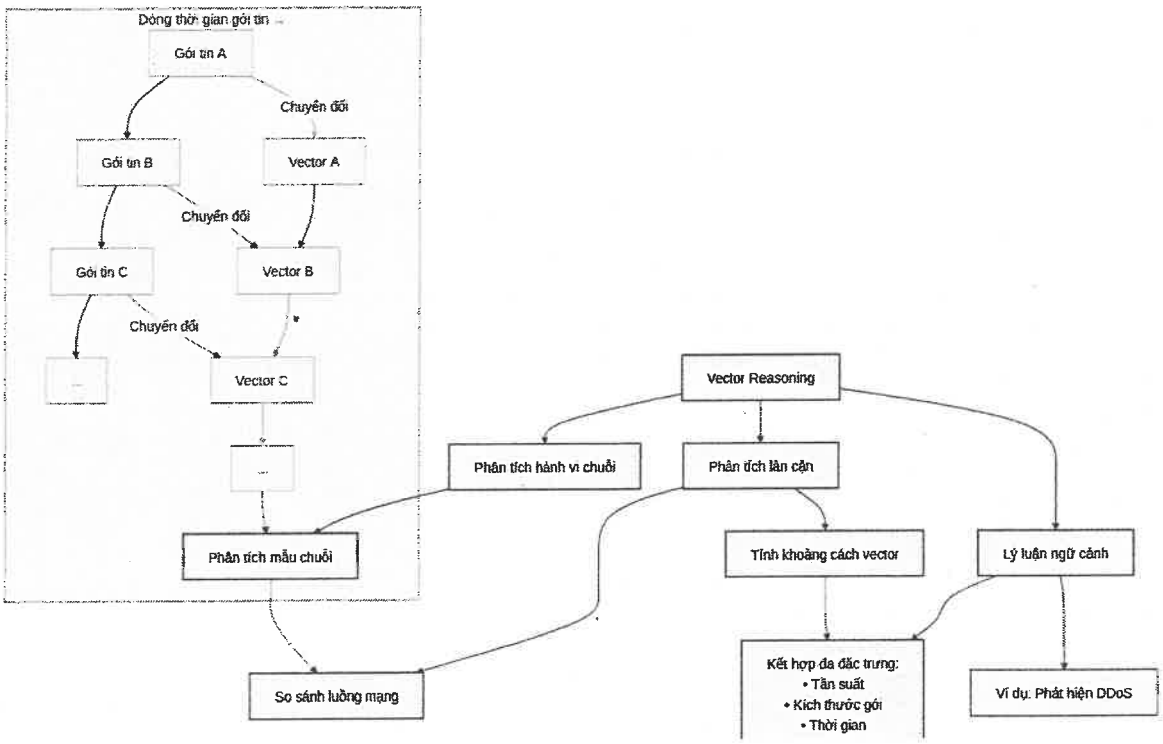
Cân bằng dữ liệu: Sử dụng kỹ thuật SMOTE để xử lý mất cân bằng dữ liệu bằng cách tạo ra các mẫu tổng hợp cho các lớp thiểu số.

Vector Reasoning: Là kỹ thuật sử dụng các vectơ đặc trưng để thực hiện suy luận và phân tích mẫu hành vi trong dữ liệu mạng. Tập trung vào:

- Suy luận mối quan hệ giữa các vector;

- Phát hiện bất thường trong không gian vector;

- Hiểu ngữ cảnh thông qua mối liên hệ giữa các đặc trưng (ví dụ: kết hợp tần suất, kích thước gói, thời gian để phát hiện DDoS).



Hình 3.5: Phân tích, phát hiện DDoS

3.2. Tiến hành thực nghiệm

3.2.1. Cài đặt môi trường mô phỏng

Để tiến hành các thực nghiệm đánh giá hiệu quả của thuật toán SVM trong việc phát hiện tấn công DDoS trên mạng 5G, một môi trường mô phỏng đã được thiết lập với các thành phần phần cứng và phần mềm cụ thể. Môi trường này đảm bảo tính nhất quán và khả năng tái tạo của các kết quả thực nghiệm.

- Phần cứng (Hardware):

+ **CPU:** Intel Core i7 (ví dụ: thế hệ thứ 10 hoặc tương đương) hoặc AMD Ryzen 7 (ví dụ: series 5000 hoặc tương đương), với ít nhất 4 nhân và 8 luồng xử lý để đảm bảo khả năng xử lý dữ liệu hiệu quả;

+ **RAM:** Bộ nhớ tối thiểu 16GB DDR4 (32GB được khuyến nghị) để xử lý các tập dữ liệu lớn và quá trình huấn luyện mô hình mà không gặp phải tình trạng thiếu bộ nhớ;

+ **Ổ cứng:** SSD với dung lượng tối thiểu 256GB (512GB trở lên được khuyến nghị) để lưu trữ bộ dữ liệu, mã nguồn, các thư viện và các mô hình đã huấn luyện. Việc sử dụng SSD giúp tăng tốc độ đọc/ghi dữ liệu, đặc biệt quan trọng khi làm việc với các file CSV lớn;

+ **GPU:** Mặc dù việc huấn luyện các mô hình SVM truyền thống trong Scikit-learn chủ yếu dựa trên CPU, nhưng nếu có sử dụng các thư viện hoặc phiên bản SVM được tối ưu hóa cho GPU (ví dụ: thông qua cuML của RAPIDS AI), thông tin về GPU (ví dụ: NVIDIA GeForce RTX 3060 hoặc tương đương) cũng cần được ghi nhận. Trong khuôn khổ của nghiên cứu này, việc huấn luyện chủ yếu dựa trên CPU.

- Phần mềm (Software):

+ **Hệ điều hành:** Windows 11 Pro (hoặc Windows 10, Ubuntu 22.04 LTS, macOS Monterey/Ventura);

+ **Ngôn ngữ lập trình:** Python phiên bản 3.9.x (hoặc 3.8.x, 3.10.x).

+ **Môi trường phát triển (IDE/Editor):** *Jupyter Notebook hoặc JupyterLab*: Được sử dụng rộng rãi cho các tác vụ khoa học dữ liệu, cho phép thực thi mã lệnh từng bước, trực quan hóa dữ liệu và ghi chú trực tiếp hoặc *Visual Studio Code* với các extension Python và Jupyter, PyCharm Professional Edition.

+ **Các thư viện Python chính:**

NumPy (phiên bản 1.23.x): Thư viện nền tảng cho tính toán khoa học bằng Python, cung cấp các cấu trúc mảng đa chiều và các hàm toán học hiệu suất cao;

Pandas (phiên bản 1.5.x): Thư viện mạnh mẽ cho việc thao tác và phân tích dữ liệu, đặc biệt là xử lý dữ liệu dạng bảng (ví dụ: đọc file CSV, làm sạch dữ liệu, trích xuất đặc trưng);

Scikit-learn (sklearn) (phiên bản 1.2.x): Thư viện học máy toàn diện, cung cấp triển khai của thuật toán SVM, các công cụ tiền xử lý dữ liệu (StandardScaler, LabelEncoder), các kỹ thuật chia tập dữ liệu (train_test_split), tối ưu hóa siêu tham số (GridSearchCV), và các thước đo đánh giá mô hình;

Imbalanced-learn (imblearn) (phiên bản 0.10.x): Thư viện cung cấp các kỹ thuật xử lý dữ liệu mất cân bằng, bao gồm SMOTE đã được sử dụng trong nghiên cứu này;

Matplotlib (phiên bản 3.6.x) và Seaborn (phiên bản 0.12.x): Các thư viện phổ biến để trực quan hóa dữ liệu, chẳng hạn như vẽ biểu đồ phân phối lớp, ma trận nhầm lẫn, và các biểu đồ đánh giá hiệu suất khác;

Joblib (thường đi kèm Scikit-learn): Được sử dụng để lưu và tải các mô hình đã huấn luyện.

+ **Bộ dữ liệu:** Các thực nghiệm được tiến hành trên bộ dữ liệu **5G-NIDD**. Chi tiết về bộ dữ liệu này, bao gồm cấu trúc, các loại tấn công và quy trình thu thập, đã được trình bày trong mục **3.1.2. Bộ dữ liệu mẫu**. Dữ liệu được sử dụng dưới dạng các file CSV đã được xử lý sơ bộ từ các file .pcap gốc.

3.2.2. Các bước thực hiện

- Bước 1: Tiền xử lý và phân tích dữ liệu thông qua các bước sau:

+ Đọc và tổng hợp dữ liệu từ các file CSV;

+ Xử lý các giá trị thiếu;

+ Loại bỏ các bản ghi trùng lặp;

+ Chuyển đổi các cột số thành định dạng số;

+ Xử lý dữ liệu phân loại:

- Sử dụng Label Encoding cho các đặc trưng phân loại như Protocol, State, và Direction;
- Xử lý các giá trị thiếu trong cột phân loại bằng cách thay thế bằng 'unknown'.

+ Phân tích phân phối lớp:

- Loại bỏ các lớp tấn công hiếm (có ít hơn 10 mẫu);
- Tạo biểu đồ phân phối cho các loại tấn công.

+ Chuẩn hóa dữ liệu:

- Sử dụng StandardScaler để chuẩn hóa các đặc trưng số, đưa chúng về cùng một thang đo, điều này quan trọng cho hiệu suất của SVM.

- Bước 2: Huấn luyện và Tối ưu hóa mô hình:

Đây là giai đoạn cốt lõi, nơi hai phương pháp thể hiện sự khác biệt rõ rệt. Dữ liệu đầu vào được chia thành tập huấn luyện (80%) và tập kiểm thử (20%) một cách có phân tầng (stratified) để đảm bảo phân phối lớp là tương đồng giữa hai tập.

+ Phương pháp 1 (SVM + RFE): Phương pháp này tập trung vào việc lựa chọn đặc trưng một cách kỹ lưỡng để tối ưu hóa hiệu suất.

- Lựa chọn đặc trưng với RFE:

Lý do: Để giảm số chiều dữ liệu, loại bỏ nhiễu và tìm ra tập hợp đặc trưng tinh túy nhất;

Thực hiện: Sử dụng kỹ thuật Recursive Feature Elimination (RFE) từ Scikit-learn để chọn ra 25 đặc trưng quan trọng nhất. RFE hoạt động bằng cách huấn luyện lặp đi lặp lại một mô hình ước tính (LogisticRegression) và loại bỏ dần các đặc trưng yếu nhất;

- Xử lý mất cân bằng:

Sử dụng tham số class_weight='balanced' ngay trong mô hình SVC. Tham số này tự động điều chỉnh trọng số của hàm mất mát, "phạt" nặng hơn khi mô hình dự đoán sai các mẫu thuộc lớp thiểu số mà không cần thay đổi dữ liệu gốc.

- Tối ưu hóa siêu tham số:

Sử dụng GridSearchCV để tìm kiếm toàn diện bộ tham số (C, gamma, kernel) tốt nhất cho SVM. Công cụ này thử tất cả các kết hợp có thể có để đảm bảo tìm ra cấu hình tối ưu nhất. Kết quả tìm được là: {'svm__C': 50, 'svm__gamma': 'scale', 'svm__kernel': 'rbf'}.

+ Phương pháp 2 (SVM + SelectKBest + SMOTE): Phương pháp này tập trung vào tốc độ và xử lý mất cân bằng dữ liệu một cách trực tiếp.

- Lựa chọn đặc trưng với SelectKBest:

Lý do: Một phương pháp nhanh hơn RFE để lựa chọn đặc trưng.

Thực hiện: Sử dụng SelectKBest với hàm tính điểm f_classif để chọn ra 20 đặc trưng có mối quan hệ thống kê mạnh nhất với biến mục tiêu.

- Xử lý mất cân bằng với SMOTE:

Lý do: Để giải quyết triệt để vấn đề mất cân bằng bằng cách tăng số lượng mẫu cho các lớp hiếm.

Thực hiện: Áp dụng kỹ thuật SMOTE từ thư viện imbalanced-learn. SMOTE tạo ra các mẫu dữ liệu tổng hợp mới cho các lớp thiểu số, giúp cân bằng lại phân phối lớp trong tập huấn luyện. Toàn bộ quy trình được đóng gói trong một ImbPipeline để đảm bảo SMOTE chỉ được áp dụng trên tập huấn luyện trong mỗi lần kiểm định chéo, tránh rò rỉ dữ liệu.

- Tối ưu hóa siêu tham số:

Sử dụng RandomizedSearchCV để tìm kiếm ngẫu nhiên một số lượng hữu hạn các tổ hợp tham số. Cách này nhanh hơn GridSearchCV và thường hiệu quả để tìm ra một cấu hình đủ tốt. Kết quả tìm được là: {'svm__kernel': 'rbf', 'svm__gamma': 0.1, 'svm__C': 100}.

- Bước 3: Lưu trữ và triển khai mô hình:

+ Lưu trữ mô hình và metadata:

- Lưu mô hình SVM đã huấn luyện với các tham số tối ưu (bao gồm cả kernel đã chọn);
- Lưu bộ mã hóa nhãn (LabelEncoder), bộ chuẩn hóa (StandardScaler), và các metadata khác như danh sách các đặc trưng đã sử dụng;
- Tạo báo cáo tóm tắt về mô hình.

+ Triển khai mô hình để phát hiện tấn công:

- Tạo hàm dự đoán để sử dụng mô hình đã huấn luyện;
- Xử lý dữ liệu đầu vào mới theo quy trình tiền xử lý giống như khi huấn luyện;
- Dự đoán loại tấn công và mức độ tin cậy.

- Bước 4: Sử dụng mô hình để phát hiện tấn công: Quy trình phát hiện tấn công bao gồm:

+ Tải mô hình và metadata:

- Tải mô hình SVM đã huấn luyện;
- Tải bộ chuẩn hóa và bộ mã hóa nhãn.

+ Xử lý dữ liệu đầu vào:

- Chuẩn bị dữ liệu đầu vào theo định dạng yêu cầu;
- Áp dụng mã hóa và chuẩn hóa tương tự như trong quá trình huấn luyện.

+ Dự đoán và phân tích kết quả:

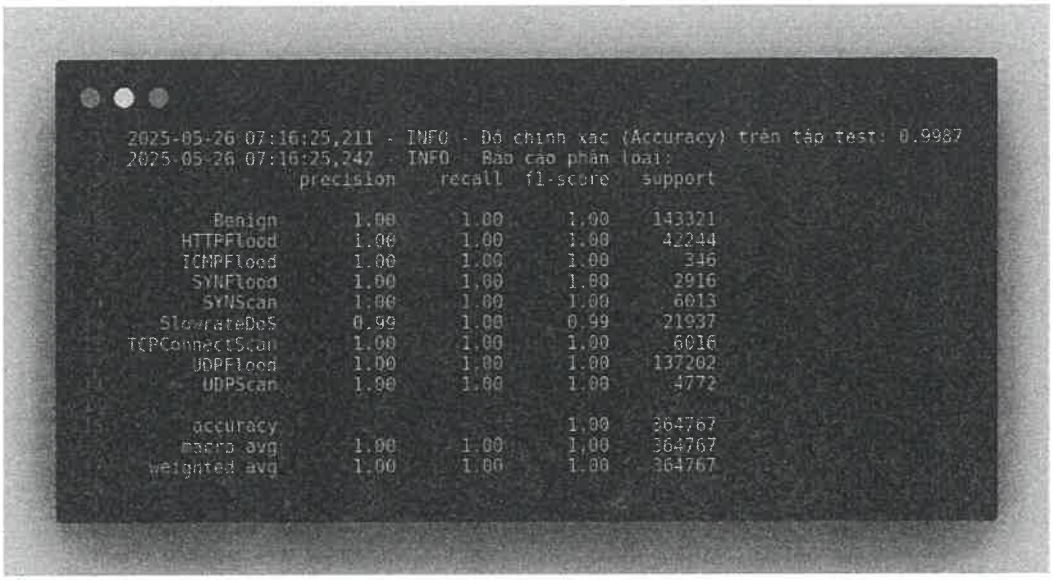
- Sử dụng mô hình để dự đoán loại tấn công;
- Tính toán xác suất cho mỗi loại tấn công (nếu mô hình hỗ trợ predict_proba);
- Cung cấp kết quả chi tiết về loại tấn công được phát hiện, mức độ tin cậy và xác suất cho từng lớp.

3.3. Kết quả thực nghiệm

3.3.1. Kết quả thu được

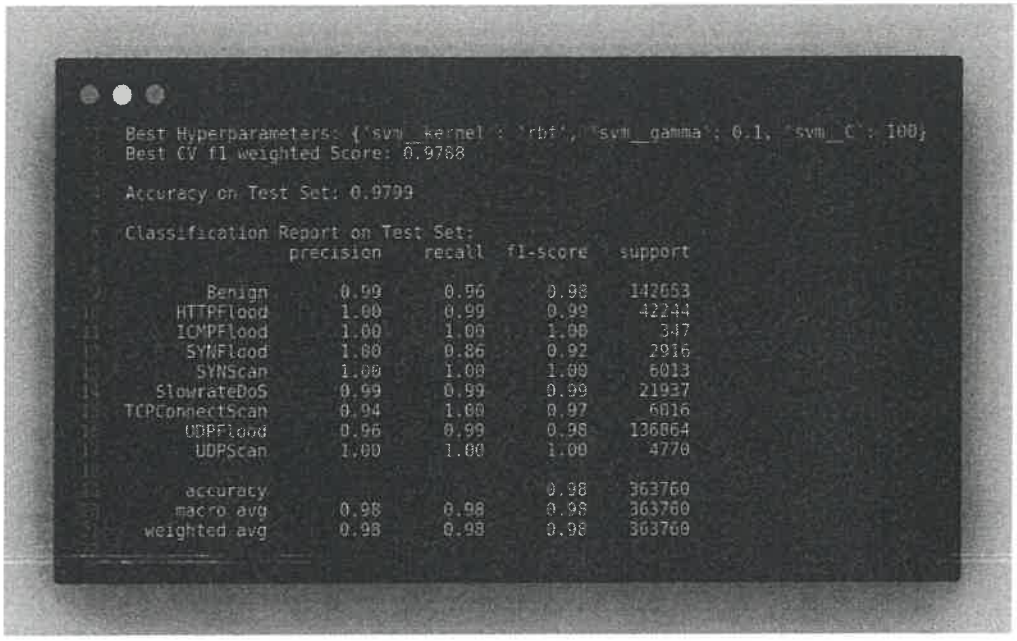
Cả 2 phương pháp thực nghiệm đều cho thấy hiệu suất cao, nhưng có sự khác biệt về kiến trúc, kỹ thuật và kết quả cuối cùng:

- Phương pháp 1: Độ chính xác 99.87%. Đạt điểm F1-score 1.00 trên hầu hết các loại tấn công như Benign, HTTPFlood, ICMPFlood, SYNFlood, SYNScan, TCPConnectScan, UDPFlood, UDPScan.



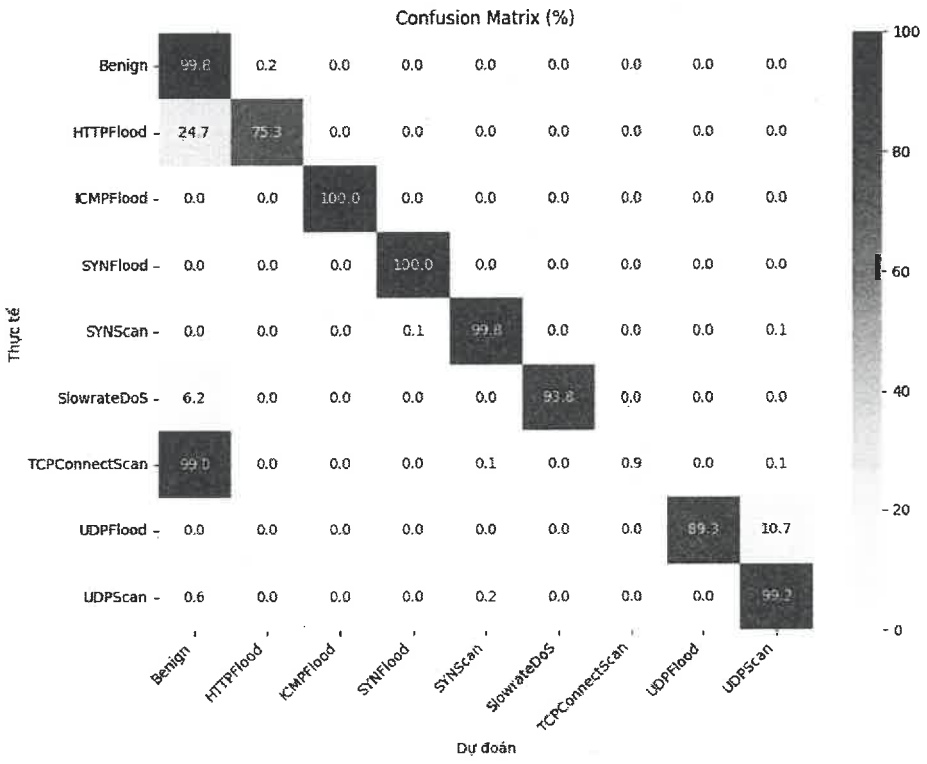
Hình 3.6: Kết quả thực nghiệm phương pháp 1

- Phương pháp 2 (SVM + SelectKBest + SMOTE): Độ chính xác 97.99%. Hiệu suất thấp hơn một chút ở lớp SYNFlood với recall là 0.86. Mặc dù hiệu suất rất tốt, nhưng vẫn thấp hơn phương pháp 1.

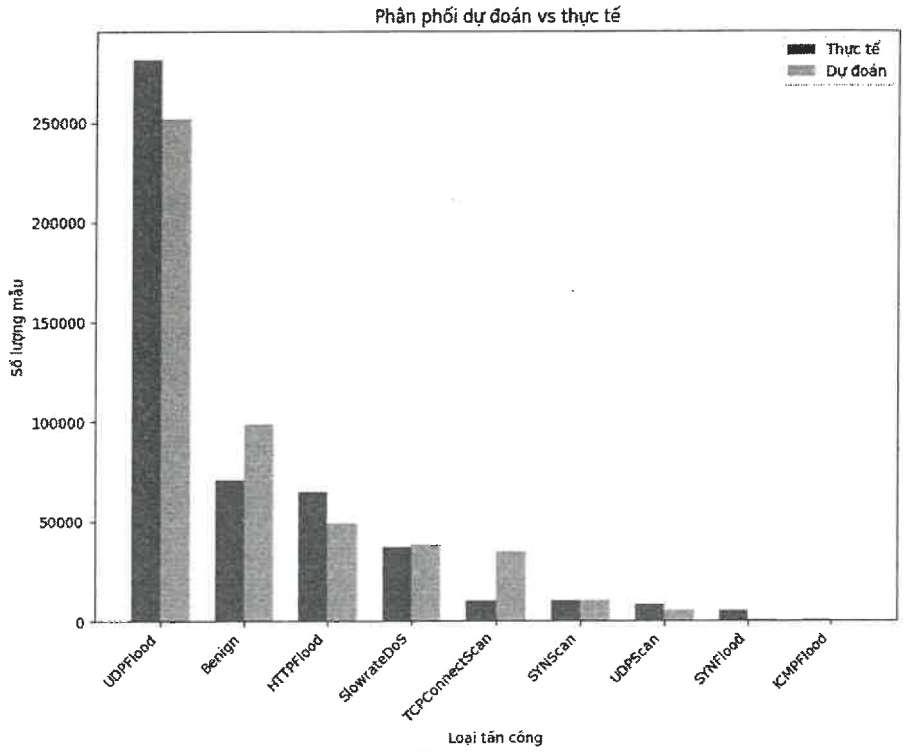


```
Best Hyperparameters: {'svm__kernel': 'rbf', 'svm__gamma': 0.1, 'svm__C': 100}
Best CV f1_weighted Score: 0.9788

Accuracy on Test Set: 0.9799

Classification Report on Test Set:
                precision    recall  f1-score   support

        Benign       0.99      0.96      0.98    142653
    HTTPFlood       1.00      0.99      0.99     42244
    ICMPFlood       1.00      1.00      1.00       347
     SYNFlood       1.00      0.86      0.92      2916
      SYNScan       1.00      1.00      1.00      6013
SlowrateDoS         0.99      0.99      0.99     21937
TCPConnectScan      0.94      1.00      0.97      6016
     UDPFlood       0.96      0.99      0.98    136864
      UDPScan       1.00      1.00      1.00      4770

     accuracy                           0.98    363760
    macro avg       0.98      0.98      0.98    363760
 weighted avg       0.98      0.98      0.98    363760
```

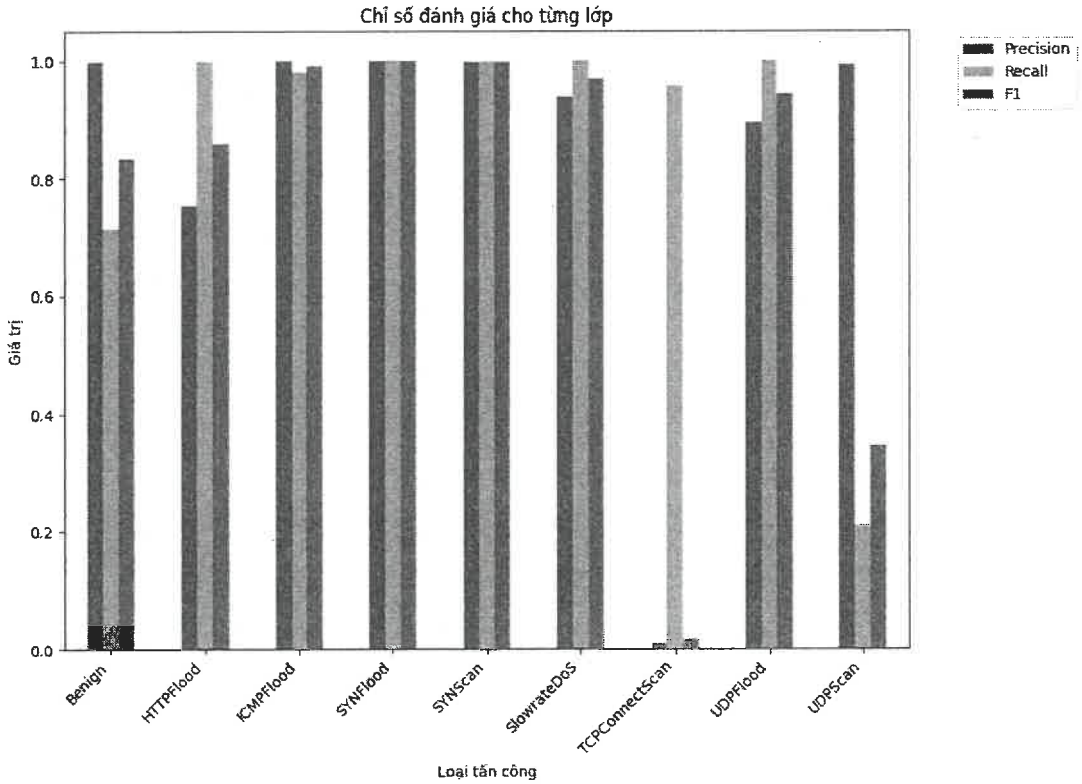
Hình 3.7: Kết quả thực nghiệm phương pháp 2



Hình 3.8: Confusion matrix với các loại tấn công



Hình 3.9: Tương quan giữa dự đoán và thực tế



Hình 3.10: Các chỉ số kết quả Precision - Recall - F1

3.3.2. Đánh giá kết quả

Kết quả thực nghiệm cho thấy cả hai phương pháp tiếp cận đều xây dựng được mô hình SVM hiệu quả để phát hiện tấn công trên mạng 5G. Tuy nhiên, có sự khác biệt rõ rệt về hiệu suất:

- Phương pháp 1 (SVM + RFE) tỏ ra vượt trội, cân bằng tốt giữa hiệu suất cao và sự đơn giản. RFE là một phương pháp lựa chọn đặc trưng mạnh mẽ. Việc sử dụng RFE để lựa chọn một tập hợp gồm 25 đặc trưng chất lượng cao đã giúp mô hình loại bỏ nhiễu và tập trung vào các tín hiệu quan trọng nhất. Điều này dẫn đến độ chính xác cao hơn và khả năng nhận diện cân bằng trên tất cả các loại tấn công. Hơn nữa, phương pháp này chứng minh rằng việc xử lý mất cân bằng bằng cách điều chỉnh trọng số (class_weight='balanced') là đủ hiệu quả khi kết hợp với một bộ đặc trưng tốt;

- Phương pháp 2 (SVM + SelectKBest + SMOTE) là một phương pháp hợp lý nhưng hiệu suất thấp hơn. Xử lý tường minh vấn đề mất cân bằng dữ liệu với SMOTE. SelectKBest là một kỹ thuật lựa chọn đặc trưng nhanh. Mặc dù đã nỗ lực xử lý mất cân bằng dữ liệu một cách tường minh bằng SMOTE, nhưng bộ 20 đặc trưng được chọn bởi SelectKBest có thể không tối ưu bằng. SelectKBest đánh giá các đặc trưng một cách độc lập và có thể đã bỏ lỡ các mối quan hệ tương tác phức tạp, dẫn đến việc mô hình nhận diện một số lớp tấn công (như SYNFlood) kém hơn.

Ưu điểm của các phương pháp thực nghiệm so với phương pháp cũ:

- Hiệu suất cao hơn: Cả hai phương pháp đều cho độ chính xác cao hơn đáng kể so với một cách tiếp cận cơ bản không có lựa chọn đặc trưng hay tối ưu hóa tham số;

- Mô hình gọn nhẹ, dễ kiểm soát nhưng vẫn hiệu quả: Việc giảm số lượng đặc trưng (từ hơn 90 xuống còn 25 hoặc 20) giúp mô hình hoạt động nhanh hơn ở cả giai đoạn huấn luyện và dự đoán;

- Khả năng khái quát hóa tốt: Quy trình tối ưu hóa tham số có hệ thống (GridSearchCV/RandomizedSearchCV) giúp tìm ra mô hình có khả năng hoạt động tốt trên dữ liệu mới, tránh overfitting;

- Xử lý được vấn đề mất cân bằng: Cả hai phương pháp đều có cơ chế riêng để xử lý vấn đề mất cân bằng dữ liệu, giúp cải thiện khả năng phát hiện các cuộc tấn công hiếm gặp.

Các loại tấn công có thể phát hiện: Mô hình có thể phát hiện các loại tấn công DDoS có trong dữ liệu huấn luyện, bao gồm:

- SYN Flood, ICMP Flood, UDP Flood;
- Slowloris;
- Goldeneye;
- Torshammer.

Nhược điểm:

- Phương pháp 1: RFE có thể tốn nhiều thời gian tính toán hơn so với các phương pháp khác;
- Phương pháp 2: Hiệu suất thấp hơn hai phương pháp còn lại. SelectKBest có thể bỏ lỡ sự tương tác giữa các đặc trưng.

Hướng phát triển:

- Thu thập thêm dữ liệu về các loại tấn công mới và biến thể;
- Tối ưu hóa các tham số của thuật toán và cải tiến các bước tiền xử lý;
- Phát triển khả năng học liên tục hoặc học online để mô hình tự cập nhật;
- Tích hợp với các hệ thống bảo mật khác như SIEM, Firewall;
- Cải thiện khả năng xử lý dữ liệu phân tán để tăng tốc độ;
- Phát triển giao diện quản lý và giám sát trực quan hơn;
- Thử nghiệm tích hợp các phương pháp học sâu hoặc mô hình ensemble để cải thiện độ chính xác. Thực hiện các Feature Engineering nâng cao…

3.4. Kết luận chương

Chương 3 đã trình bày quá trình thực nghiệm phát hiện tấn công DDoS trong mạng 5G bằng thuật toán SVM trên bộ dữ liệu 5G-NIDD. Hai phương pháp triển khai được so sánh cho thấy SVM kết hợp RFE đạt độ chính xác cao hơn và khả năng phân loại tốt hơn so với phương pháp dùng SelectKBest và SMOTE. Việc lựa chọn đặc trưng phù hợp và xử lý mất cân bằng dữ liệu đóng vai trò quan trọng trong hiệu suất mô hình. Các chỉ số đánh giá như Precision, Recall và F1-score đều cho thấy mô hình có khả năng phát hiện hiệu quả nhiều loại tấn công khác nhau. Kết quả thực nghiệm khẳng định tính khả thi của việc ứng dụng học máy vào bảo mật mạng 5G.

KẾT LUẬN VÀ ĐỀ XUẤT

Đề án đã tìm hiểu tổng quan kiến trúc mạng 5G và các nguy cơ bảo mật, đặc biệt là tấn công DDoS. SVM được nghiên cứu và áp dụng để phát hiện tấn công DDoS, phân loại lưu lượng mạng thành bình thường hoặc độc hại. Quá trình thực nghiệm trên bộ dữ liệu 5G-NIDD cho thấy SVM có độ chính xác cao trong phát hiện các loại tấn công DDoS phổ biến, phù hợp cho giám sát thời gian thực. Tuy nhiên, việc triển khai thực tế đối mặt với những thách thức về khối lượng dữ liệu lớn, yêu cầu xử lý thời gian thực, tính đa dạng của lưu lượng mạng và tài nguyên tính toán trên hệ thống. Hiệu quả của mô hình phụ thuộc rất nhiều vào chất lượng dữ liệu huấn luyện và việc tinh chỉnh các tham số của SVM.

Để nâng cao hiệu quả phát hiện tấn công DDoS trong mạng 5G, tác giả đề xuất cần tập trung nghiên cứu chuyên sâu vào các hướng sau:

Dữ liệu và đặc trưng: Thu thập thêm dữ liệu về các loại tấn công mới và biến thể. Nghiên cứu sâu hơn về kỹ thuật Feature Engineering nâng cao để trích xuất đặc trưng có ý nghĩa từ dữ liệu mạng 5G.

Tối ưu hóa mô hình: Áp dụng các kỹ thuật tối ưu hóa tham số nâng cao (ví dụ: *tối ưu hóa Bayes*) và phát triển khả năng học liên tục (*Continual learning*) hoặc học trực tuyến (*Online learning*) để mô hình tự động cập nhật và thích nghi với các mẫu tấn công thay đổi. Thử nghiệm kết hợp học sâu (*Deep Learning*) để cải thiện độ chính xác và khả năng phát hiện các cuộc tấn công phức tạp hơn.

Hiệu suất và triển khai: Nghiên cứu kiến trúc phân tán và tính toán song song trên các nền tảng Big Data (như *Apache Spark*) để tăng tốc độ xử lý dữ liệu. Tích hợp hệ thống phát hiện với các giải pháp bảo mật hiện có (*SIEM, Firewall*) để tạo ra giải pháp toàn diện cho mạng 5G. Phát triển giao diện quản lý và giám sát trực quan, đồng thời tự động hóa quy trình cập nhật mô hình.

DANH MỤC TÀI LIỆU THAM KHẢO

[1] Wikipedia, "5G," [Trực tuyến]. Available: https://en.wikipedia.org/wiki/5G.

[2] A. Rachedi, M. H. Rehmani, S. Cherkaoui, and J. J., "IEEE access special section editorial: The plethora of research in Internet of Things (IoT)," *IEEE Access,* vol. 4, p. 9575–9579, 2016.

[3] Sehan Samarakoon, Yushan Siriwardhana, Pawani Porambage, Madhusanka Liyanage, Sang-Yoon Chang, Jinoh Kim, Jonghyun Kim and Mika Ylianttila, "5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network," 2022.

[4] C. Lai, R. Lu, D. Zheng, and X. Shen, "Security and privacy challenges in 5G-enabled vehicular networks," *IEEE Netw.,* vol. 34, no. 2, p. 37–45, 2020.

[5] A. Abouaomar, M. Elmachkour, A. Kobbane, H. Tembin, "Users-Fogs association within a cache context in 5G networks: Coalition game model," *IEEE Symposium on Computers and Communications (ISCC),* pp. 14-19, 2018.

[6] Fakhouri, H.N.; Alawadi, S.; Awaysheh, F.M.; Hani,, "A Comprehensive Study on the Role of Machine Learning in 5G Security: Challenges, Technologies, and Solutions," *Electronics,* vol. 12, p. 4604, 2023.

[7] Jaya Preethi Mohan, Niroop Sugunaraj, Ranganathan , "Cyber Security Threats for 5G Networks," in *IEEE International Conference on Electro Information Technology (eIT)*, 2022.

[8] Abdul Ahad, Zahra Ali, Abdul Mateen, Mohammad Tahi, "A Comprehensive review on 5G-based Smart Healthcare Network Security: Taxonomy, Issues, Solutions and Future research directions," 2023.

[9] I. Ahmad, T. Kumar, M. Liyanage, J. Okwuibe, M. Yl, "Overview of 5G security challenges and solutions," *IEEE Commun. Stand. Mag.,* vol. 2, no. 1, pp. 36-43, 2018.

[10] A. Arfaoui, S. Cherkaoui, A. Kribeche, S. M. Senou, "Context-aware adaptive authentication and authorization in internet of things," *The ICC 2019-2019 IEEE International Conference on Communications (ICC), IEEE,* pp. 1-6, 2019.

[11] H. Moudoud, S. Cherkaoui, and L. Khoukhi, "An IoT blockchain architecture using oracles and smart contracts: the use-case of a food supply chain," *The 2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), IEEE,* pp. 1-6, 2019.

[12] Sulaiman Yousef Alshunaifi, Shailendra Mishra and Mohammed Alshehri, "Cyber-Attack Detection and Mitigation Using SVM for 5G Network," *Intelli gent Autom ation & Soft Comput ing,* vol. 31, 2022.

[13] Cherifa Hamroun, Anne Fladenmuller, Michel Pariente and Guy Pujolle, "Intrusion Detection in 5G and Wi-Fi Networks: A Survey of Current Methods, Challenges, and Perspectives," vol. 13, pp. 40950-40976, 2025.

[14] Dhanya K. A., Sulakshan Vajipayajula, Kartik Srini, "Detection of Network Attacks using Machine Learning and Deep Learning Models," *Procedia Computer Science,* vol. 218, pp. 57-66, 2023.

KiểmTraTàiLiệu

BÁO CÁO KIỂM TRA TRÙNG LẶP

Thông tin tài liệu

Tên tài liệu:	Đề án tốt nghiệp Ngô Văn Nhận
Tác giả:	Nhận Ngô Văn
Điểm trùng lặp:	8
Thời gian tải lên:	01:23 02/08/2025
Thời gian sinh báo cáo:	01:26 02/08/2025
Các trang kiểm tra:	66/66 trang

Kết quả kiểm tra trùng lặp

Có 8% nội dung trùng lặp

Có 92% nội dung không trùng lặp

Có 0% nội dung người dùng loại trừ

Có 0% nội dung hệ thống bỏ qua

Nguồn trùng lặp tiêu biểu

123docz.net arxiv.org tailieu.vn

Người hướng dẫn khoa học
(Ký và ghi rõ họ tên)

TS. Nguyễn Đình Hóa

Học viên
(Ký và ghi rõ họ tên)

Ngô Văn Nhận

BỘ KHOA HỌC VÀ CÔNG NGHỆ
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

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

BÁO CÁO GIẢI TRÌNH
SỬA CHỮA, HOÀN THIỆN ĐỀ ÁN TỐT NGHIỆP

Họ và tên học viên: Ngô Văn Nhận

Chuyên ngành: HTTT

Khóa: 2023 đợt 2

Tên đề tài: Nghiên cứu phương pháp phát hiện tấn công DDoS trong mạng di động 5G

Người hướng dẫn khoa học: TS. Nguyễn Đình Hóa

Ngày bảo vệ: 19/07/2025

Các nội dung học viên đã sửa chữa, bổ sung trong đề án tốt nghiệp theo ý kiến đóng góp của Hội đồng chấm đề án tốt nghiệp:

TT	Ý kiến hội đồng	Sửa chữa của học viên
1	Chỉnh sửa lỗi soạn thảo, lỗi ngữ pháp, chính tả	Học viên đã rà soát, chỉnh sửa các lỗi soạn thảo, các lỗi ngữ pháp
2	Chỉnh sửa mục tiêu nghiên cứu	Chỉnh sửa, nêu rõ mục tiêu cần đạt được sau khi nghiên cứu
3	Nhất quán trong các phần trích dẫn	Học viên đã chỉnh sửa phần trích dẫn tài liệu tham khảo tại trang 19 (mục 1.5.2 trong chương 1)
4	Nêu rõ các công cụ đã sử dụng	Mục 2.4 (chương 2) học viên đã bổ sung mô tả về 2 phương pháp được sử dụng: khái niệm, cách thức hoạt động của thuật toán RFE, SelectKBest, SMOTE
5	Bổ sung kết luận của từng chương	Tiếp thu góp ý của Hội đồng, học viên đã bổ sung kết luận cho các chương, liên kết tới nội dung chương tiếp theo

Hà Nội, ngày 02 tháng 8 năm 2025

Ký xác nhận của

CHỦ TỊCH HỘI ĐỒNG CHẤM ĐỀ ÁN	THƯ KÝ HỘI ĐỒNG	NGƯỜI HƯỚNG DẪN KHOA HỌC	HỌC VIÊN
TS.Nguyễn Duy Phương	TS.Đặng Hoàng Long	TS.Nguyễn Đình Hóa	Ngô Văn Nhận

BỘ KHOA HỌC VÀ CÔNG NGHỆ
HỌC VIỆN CÔNG NGHỆ
BƯU CHÍNH VIỄN THÔNG

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

BIÊN BẢN
HỌP HỘI ĐỒNG CHẤM ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

Căn cứ quyết định số Quyết định số 1098/QĐ-HV ngày 26 tháng 06 năm 2025 của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ. Hội đồng đã họp vào hồi 12 giờ 00 phút, ngày 19 tháng 07 năm 2025 tại Học viện Công nghệ Bưu chính Viễn thông để chấm đề án tốt nghiệp thạc sĩ cho:

Học viên: **Ngô Văn Nhận**

Tên đề án tốt nghiệp: **Nghiên cứu phương pháp phát hiện tấn công DDoS trong mạng di động 5G**

Chuyên ngành: **Hệ thống thông tin** Mã số: **8480104**

Các thành viên của Hội đồng chấm đề án tốt nghiệp có mặt: ...5.../ 05

TT	HỌ VÀ TÊN	TRÁCH NHIỆM TRONG HĐ	GHI CHÚ
1	**TS. Nguyễn Duy Phương**	Chủ tịch	
2	**TS. Đặng Hoàng Long**	Thư ký	
3	**PGS.TS. Đặng Văn Đức**	Phản biện 1	
4	**TS. Vũ Văn Thỏa**	Phản biện 2	
5	**PGS.TS. Trần Thị Oanh**	Uỷ viên	

Các nội dung thực hiện:

1. Chủ tịch Hội đồng điều khiển buổi họp. Công bố quyết định của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ.
2. Người hướng dẫn khoa học hoặc thư ký đọc lý lịch khoa học và các điều kiện bảo vệ đề án tốt nghiệp của học viên. (có bản lý lịch khoa học và kết quả các môn học cao học của học viên kèm theo).
3. Học viên trình bày tóm tắt đề án tốt nghiệp.
4. Phản biện 1 đọc nhận xét (có văn bản kèm theo)
5. Phản biện 2 đọc nhận xét (có văn bản kèm theo)
6. Các câu hỏi của thành viên Hội đồng:

* PGS.TS Đặng Văn Đức
Bổ sung thêm nội dung trong và ngoài nước liên quan đến đề án
Giải thích thêm phần dữ liệu mẫu được đề cập trong đề án
* TS Vũ Văn Thoả
Giải thích dài đến tấn công DDOS trong 5G?

7. Trả lời của học viên:

Học viên đã trả lời về ~~hay~~ câu hỏi về dữ liệu mẫu, trình bày về phương pháp mô tả dữ liệu đề tài luận văn

8. Thư ký đọc nhận xét về quá trình thực hiện đề án tốt nghiệp của học viên *(có văn bản kèm theo)*.

9. Hội đồng họp riêng:

- Bầu Ban kiểm phiếu:
  1. Trưởng Ban kiểm phiếu: TS. Nguyễn Duy Phương
  2. Uỷ viên Ban kiểm phiếu: TS. Đặng Hoàng Long
  3. Uỷ viên Ban kiểm phiếu: TS. Vũ Văn Thoại
- Hội đồng chấm đề án tốt nghiệp bằng bỏ phiếu kín.
- Ban kiểm phiếu làm việc:
- Trưởng Ban kiểm phiếu báo cáo kết quả kiểm phiếu *(có Biên bản họp Ban kiểm phiếu kèm theo)*
- **Điểm trung bình của đề án tốt nghiệp:** 8.0

Kết luận:

1. Các nội dung cần chỉnh sửa, hoàn thiện sau bảo vệ đề án tốt nghiệp:

Cần xem xét bổ sung phương pháp có liên quan đến đề tài

Trình bày rõ các công cụ thực nghiệm, phương pháp thu thập thực nghiệm

Bổ sung tài liệu tham khảo, bố cục luận văn của từng chương

2. Đề nghị Học viện công nhận (hoặc không) và cấp bằng (hoặc không) thạc sĩ cho học viên:
Đề nghị công nhận và cấp bằng thạc sĩ cho học viên

3. Đề án tốt nghiệp có thể phát triển thành đề tài nghiên cứu cho NCS.....Không.....

Buổi làm việc kết thúc vào....11h30....cùng ngày.

Chủ tịch	**Thư ký**
Phương	Long.
TS. Nguyễn Duy Phương	TS. Đặng Hoàng Long

BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

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

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

Tên đề tài đề án:	Nghiên cứu phương pháp phát hiện tấn công DDoS trong mạng di động 5G
Chuyên ngành:	Hệ thống thông tin
Mã số:	8.48.01.04
Tên học viên:	Ngô Văn Nhận
Họ tên người nhận xét:	Vũ Văn Thỏa
Học hàm, học vị:	Tiến sĩ **Chuyên ngành**: Toán học tính toán
Cơ quan công tác:	Học viện Công nghệ Bưu chính Viễn thông

NỘI DUNG NHẬN XÉT

I/ Cơ sở khoa học và thực tiễn, tính cấp thiết của đề tài

Hiện nay, mạng di động 5G đang được triển khai và vận hành rộng rãi tại Việt Nam. Mạng di động 5G đóng vai trò quan trọng trong các hạ tầng nền tảng cho quá trình chuyển đổi số và các dịch vụ số trên Internet. Cùng với đó, vấn đề bảo mật trong mạng 5G cũng đặt ra nhiều thách thức và trở thành mối quan tâm hàng đầu của người dùng và các nhà cung cấp dịch vụ. Một trong các hướng giải quyết vấn đề trên là nghiên cứu và ứng dụng các kỹ thuật phát hiện và chống tấn công mạng trong mạng 5G.

Đề án với tiêu đề "Nghiên cứu phương pháp phát hiện tấn công DDoS trong mạng di động 5G" của học viên Ngô Văn Nhận tập trung nghiên cứu ứng dụng kỹ thuật học máy nhằm xây dựng hệ thống phát hiện tấn công DDoS trong mạng di động 5G. Vì vậy, đề tài đề án có tính cấp thiết, có ý nghĩa khoa học và thực tiễn.

II/ Về nội dung chất lượng của đề án, những kết quả đạt được

Đề án được trình bày trong 56 trang bao gồm phần mở đầu, 3 chương nội dung, phần kết luận và danh mục các tài liệu tham khảo.

Chương 1 của đề án gồm 17 trang, trình bày tổng quan về cấu trúc mạng 5G, các nguy cơ đe dọa an toàn bảo mật trong mạng 5G. Trên cơ sở nghiên cứu các phương pháp bảo mật hiện có trên mạng 5G, đề án đã khảo sát vai trò và ứng dụng học máy trong bảo mật mạng di động 5G. Tuy nhiên, trong chương này đề án chưa làm rõ cơ chế của các hình thức tấn công trong mạng 5G để từ đó trình bày bài toán đặt ra cần giải quyết trong đề án. Cần bổ sung nội dung kết luận chương 1 để tổng kết các nội dung chính trong chương và các nội dung liên kết đến chương 2.

Chương 2 của đề án bao gồm 14 trang với nội dung chính là khảo sát chi tiết tấn công DDoS trong mạng di động 5G và kỹ thuật học máy SVM (máy vector hỗ trợ) để phát hiện tấn công DDoS trong mạng di động 5G. Tuy nhiên, đề án chưa làm rõ các đặc trưng tấn công DDoS trong mạng di động 5G để làm cơ sở cho phát hiện sau này. Mặt khác, các nội dung về SVM khá sơ sài. Đề án cũng chưa nêu các lý do chọn SVM để phát hiện tấn công DDoS. Cần bổ sung

nội dung kết luận chương 2 để tổng kết các nội dung chính trong chương và các nội dung liên kết đến chương 3.

Chương 3 của đề án gồm 19 trang, trình bày các nội dung triển khai thử nghiệm sử dụng SVM phát hiện tấn công DDoS trong mạng di động 5G trên bộ dữ liệu 5G-NIDD. Đề án đã trình bày các kết quả thử nghiệm và phân tích đánh giá kết quả dựa trên các tiêu chí đánh giá chung cho các kỹ thuật học máy. Tuy nhiên, đề án chưa có phân tích kỹ các kiểu tấn công DoS trong bộ dữ liệu và tiêu chí để phát hiện tấn công DDoS. Vì vậy các kết quả thực nghiệm chưa phù hợp với bài toán đặt ra. Vấn đề lựa chọn các đặc trưng của dữ liệu trình bày trong đề án chưa có cơ sở khoa học và cần có các khảo sát thêm. Nên trình bày các kết quả thực nghiệm dưới dạng bảng thì hợp lý hơn. Cần bổ sung nội dung kết luận chương 3.

Đề án được chia thành các chương mục rõ ràng và cơ bản bám sát các nội dung theo đề cương đã được phê duyệt. Cần bổ sung mục kết luận chương vào cuối mỗi chương để tóm tắt kết quả của chương và liên kết với chương tiếp theo. Nên hiệu chỉnh phần mở đầu của đề án để phù hợp hơn với nội dung và kết quả đạt được của đề án. Phần Kết luận của đề án cần hiệu chỉnh để nêu rõ các kết quả đạt được của đề án và theo đúng quy định của Học viện.

Trong đề án còn có khá nhiều lỗi in ấn phải sửa chữa tại các trang 3, 36, 43, 55, ... Các hình trong đề án cần được soạn thảo rõ hơn và Việt hóa nội dung phù hợp. Phần Danh mục tài liệu tham khảo cần được trình bày theo đúng qui định của Học viện.

III/ Những vấn đề cần giải thích thêm

(1) Giải thích rõ hơn các đặc điểm của tấn công DDoS trong mạng di động 5G. Từ đó, giải thích các kiểu tấn công DoS trình bày tại bảng 3.2 (trang 38) tương ứng như thế nào với các kiểu tấn công mạng trình bày trong mục 1.3 của đề án.

(2) Nêu các lý do đề án chọn kỹ thuật học máy SVM để sử dụng cho bài toán phát hiện tấn công DDoS trong mạng di động 5G?

(3) Giải thích rõ hơn mô hình hệ thống phát hiên tấn công DDoS trong mạng di động 5G trình bày tại hình 3.5 (trang 44). Nêu cách triển khai hệ thống đó trong mạng di động 5G?

IV/ Kết luận

Tôi đồng ý cho học viên Ngô Văn Nhận được bảo vệ đề án với đề tài "Nghiên cứu phương pháp phát hiện tấn công DDoS trong mạng di động 5G" trước Hội đồng chấm đề án tốt nghiệp thạc sĩ của Học viện. Tuy nhiên, để đảm bảo chất lượng đề án tốt nghiệp thạc sĩ chuyên ngành Hệ thống thông tin, học viên cần chỉnh sửa đề án và giải trình rõ trước Hội đồng các vấn đề được góp ý trên đây.

Ngày 14 tháng 07 năm 2025

Người nhận xét

TS Vũ Văn Thỏa

TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VN
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH-VIỄN THÔNG

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

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SỸ

(Dùng cho cán bộ phản biện)

Tên đề tài đề án: **Nghiên cứu phương pháp phát hiện tấn công DDOS trong mạng di động 5G**
Chuyên ngành: Hệ thống thông tin
Mã số: 8.48.01.04
Tên học viên: ***Ngô Văn Nhận***
Họ và tên người nhận xét: PGS.TS. Đặng Văn Đức
Chuyên ngành: Công nghệ thông tin
Cơ quan công tác: Viện Công nghệ thông tin, Viện Hàn lâm KH&CN Việt Nam

NỘI DUNG NHẬN XÉT

I. Cơ sở khoa học và thực tiễn, tính cấp thiết của đề án

Sự phát triển nhanh chóng của mạng di động 5G mang lại nhiều đột phá về băng thông, độ trễ và mật độ kết nối, đồng thời cũng làm gia tăng nguy cơ bị khai thác bởi các hình thức tấn công mạng, đặc biệt là tấn công từ chối dịch vụ phân tán (DDoS). Với cấu trúc phân mảnh, định tuyến mềm dẻo và khả năng ảo hóa tài nguyên, mạng 5G đặt ra nhiều thách thức trong việc giám sát và phát hiện bất thường, khiến các biện pháp an ninh truyền thống trở nên kém hiệu quả. Do đó, việc nghiên cứu và phát triển các phương pháp phát hiện DDoS phù hợp với đặc trưng của mạng 5G là một yêu cầu cấp thiết nhằm đảm bảo tính sẵn sàng và ổn định của hệ thống. Đề tài có ý nghĩa thực tiễn rõ rệt trong việc hỗ trợ xây dựng các hệ thống bảo vệ mạng di động thế hệ mới, đặc biệt trong bối cảnh quá trình triển khai 5G tại Việt Nam đang được đẩy mạnh. Mục tiêu của đề án là nghiên cứu khảo sát mô hình kiến trúc tổng quan của mạng di động 5G, các nguy cơ về mất an toàn bảo mật và phương pháp đảm bảo an toàn bảo mật trong mạng di động 5G, từ đó thực nghiệm giải pháp lựa chọn phát hiện tấn công DDoS để đảm bảo an ninh, an toàn thông tin cho mạng di động 5G. Đề tài đề án tốt nghiệp thạc sỹ của học viên *Ngô Văn Nhận* là có ý nghĩa thực tiễn và có tính cấp thiết.

II. Về nội dung, chất lượng của đề án, các kết quả đã đạt được

Để đạt được mục tiêu đề ra cho đề án tốt nghiệp của mình, học viên đã trình bày các vấn đề chính sau đây:

- Trình bày được khái quát về mạng di động 5G và an toàn bảo mật trong mạng di động, bao gồm sơ lược về kiến trúc mạng 5G, nguy cơ đe doạ an toàn bảo mật trong

mạng 5G, một số hình thức tấn công mạng và phương pháp bảo mật mạng di động 5G.

- Trình bày được phương pháp phát hiện tấn công DDOS trong mạng di động 5G sử dụng học máy SVM.
- Thực nghiệm phát hiện tấn công DDOS bằng SVM trong môi trường mô phỏng. có đánh giá kết quả thu được.

Nội dung nghiên cứu phù hợp đối với cấp độ đề án thạc sỹ kỹ thuật theo định hướng ứng dụng. Bản đề án được trình bày với cấu trúc tương đối phù hợp, ít lỗi in ấn. Tuy nhiên. học viên cần nghiên cứu chỉnh sửa bản đề án của mình theo các góp ý sau đây:

- Cần viết lại mục tiêu đề tài của đề án ở phần Mở đầu. Mục tiêu là cái mà đề án cần đạt được, không phải mô tả những gì đề án làm.
- Cần bổ sung tình hình nghiên cứu, ứng dụng trong và ngoài nước về nội dung tương tự, từ đó khẳng định đề án này là nghiên cứu khảo sát và học hỏi.
- Cần bổ sung sự khác biệt bảo mật và an toàn thông tin trong mạng di động 5G và mạng di động các thế hệ trước.
- Cần nhất quán khi trích dẫn tài liệu tham khảo, ví dụ việc trích dẫn tại trang 19 là không phù hợp.
- Rà soát toàn bộ đề án để chỉnh sửa các lỗi in ấn, sử dụng từ ngữ cho phù hợp (ví dụ "giả vờ" trang 38). Bổ sung đầy đủ các tham chiếu đến tài liệu tham khảo với những kết luận không phải của học viên. Việt hoá các hình vẽ, bảng biểu, nếu sử dụng nguyên gốc hình vẽ từ tài liệu tham khảo thì phải có tham chiếu (ví dụ hình 3.3 trang 39, Bảng 3.1 trang 37), không sử dụng tiếng Anh những nới không cần thiết.
- Rà soát các đoạn, chương mục trong bản đề án sao cho logic, dễ hiểu. Ví dụ, tiểu mục 1.4.1 trang 15 trình bày Trình mô phỏng tấn công SIPDAS trong mục các phương pháp bảo mật hiện có, hơn nữa nội dung tiểu mục này không rõ nghĩa. Đầu các chương cần có tóm tắt nội dung chương, giữa các chương mục cần có kết nối logic, ví dụ mục 2.1.1 Mạng Botnet IoT cần có dẫn dắt tại sao trình bày nội dung này. Cũng như tại sao lại chỉ trình bày hai kỹ thuật tấn công DDOS trong 5G (trang 21),...
- Cả chương 2 và chương 3 trình bày quá gắn với các hàm trong thư viện của ngôn ngữ Python là không phù hợp. Ví dụ mục 2.4 Phương án triển khai phát hiện xâm nhập bằng SVM có câu "... loại bỏ các đặc trưng không cần thiết thông qua RFE..." trong khi RFE chưa được giải thích ở đâu...
- Chương 3 thực nghiệm nên trình bày dưới dạng sơ đồ trực quan, ví dụ UML để dễ theo dõi. Cần bổ sung sơ đồ tổng thể hệ thống thực nghiệm.
- Cần bổ sung kết luận cho các chương.
- Trình bày lại phần kết luận chung cho rõ đề án đã làm gì, các kết quả chính của học viên.

III. Kết luận

Bản đề án tốt nghiệp thạc sỹ của học viên đã trình bày được một số nội dung cơ bản để đạt được mục tiêu đề ra. Tuy nhiên, bản đề án này cần được chỉnh sửa để bảo đảm đầy đủ các tiêu chuẩn cơ bản của đề án thạc sỹ kỹ thuật định hướng ứng dụng theo các góp ý trên đây. Tôi đồng ý để học viên *Ngô Văn Nhận* được bảo vệ trước Hội đồng bảo vệ đề án tốt nghiệp thạc sỹ.

Hà Nội, ngày tháng 07 năm 2025

NGƯỜI NHẬN XÉT

PGS.TS Đặng Văn Đức