ĐẠI HỌC QUỐC GIA TP. HCM
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

# BÁO CÁO TỔNG KẾT
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ SINH VIÊN NĂM 2023

*Tên đề tài tiếng Việt:*

## XÂY DỰNG HỆ THỐNG PHÁT HIỆN XÂM NHẬP CHỐNG LẠI CÁC CUỘC TẤN CÔNG DO THÁM KIỂM MỚI

*Tên đề tài tiếng Anh:*

## TOWARD AN ROBUST INTRUSION DETECTION SYSTEM AGAINST UNKNOWN SCAN ATTACKS

Khoa/ Bộ môn: Khoa Mạng máy tính và Truyền thông

Thời gian thực hiện: 6 tháng

Cán bộ hướng dẫn: Ts. Lê kim Hùng

Tham gia thực hiện

| TT | Họ và tên, MSSV | Chịu trách nhiệm | Điện thoại | Email |
|---|---|---|---|---|
| 1. | Phạm Ngọc Thiện, 21522627 | Chủ nhiệm | 0929000433 | 21522627@gm.uit.edu.vn |
| 2. | Nguyễn Hoài Phương, 21520408 | Tham gia | | 21520408@gm.uit.edu.vn |
| 3. | Trần Minh Duy, 21522010 | Tham gia | | 21522010@gm.uit.edu.vn |

**ĐẠI HỌC QUỐC GIA TP. HCM**
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

| | |
|---|---|
| **Ngày nhận hồ sơ** | |
| **Mã số đề tài** | |
| ***(Do CQ quản lý ghi)*** | |

# BÁO CÁO TỔNG KẾT

*Tên đề tài tiếng Việt:*

**XÂY DỰNG HỆ THỐNG PHÁT HIỆN XÂM NHẬP CHỐNG LẠI CÁC CUỘC TẤN CÔNG DO THÁM KIỂU MỚI**

*Tên đề tài tiếng Anh:*

**TOWARD AN ROBUST INTRUSION DETECTION SYSTEM AGAINST UNKNOWN SCAN ATTACKS**

*Ngày ... tháng ..... năm ....*
**Cán bộ hướng dẫn**
*(Họ tên và chữ ký)*

*Ngày ... tháng ..... năm ....*
**Sinh viên chủ nhiệm đề tài**
*(Họ tên và chữ ký)*

# THÔNG TIN KẾT QUẢ NGHIÊN CỨU

1. Thông tin chung:

- Tên đề tài:

**XÂY DỰNG HỆ THỐNG PHÁT HIỆN XÂM NHẬP CHỐNG LẠI CÁC CUỘC TẤN CÔNG DO THÁM KIỂU MỚI**

- Chủ nhiệm: Phạm Ngọc Thiện
- Thành viên tham gia: Nguyễn Hoài Phương, Trần Minh Duy
- Cơ quan chủ trì: Trường Đại học Công nghệ Thông tin.
- Thời gian thực hiện: 6 tháng

2. Mục tiêu:

- Tìm hiểu tổng quan về bài toán "Novelty Detection" và các phương pháp tiếp cận đã được đề xuất.
- Xây dựng một hệ thống phát hiện các cuộc tấn công do thám kiểu mới kết hợp học giám sát (supervised) và học không giám sát (unsupervised) sử dụng ba mô hình máy học Random forest (RF), Local Outlier Factor (LOF) và Isolation-based Anomaly Detection Using Nearest-Neighbor Ensembles (iNNE)
- Nhóm nghiên cứu sẽ thực nghiệm và chứng minh tính hiệu quả của hệ thống bằng cách sử dụng bộ dữ liệu CIDDS 001 và CIDDS 002 cho việc huấn luyện và đánh giá mô hình trong cả 2 ngữ cảnh binary classification và multi-class classification.

3. Tính mới và sáng tạo:

- Ứng dụng mô hình máy học để xây dựng một NIDS có khả năng phát hiện và phân loại lưu lượng mạng theo thời gian thực. Mô hình có khả năng phát hiện các loại tấn công kiểu mới chưa được huấn luyện với độ chính xác cao.
- Khả năng mở rộng và nâng cấp dễ dàng của mô hình NIDS. Khi cần mở rộng hoặc cập nhật dữ liệu về các cuộc tấn công, chỉ cần tiến hành huấn luyện lại từng module riêng lẻ, không cần huấn luyện lại toàn bộ mô hình.

4. Tóm tắt kết quả nghiên cứu:

- Hệ thống phát hiện xâm nhập mạng có độ chính xác cao (ACC > 91% với mọi testcase) trên tập dự liệu CIDDS-001 và CIDDS-002.
- Ngoài ra, tỉ lệ phát hiện chính sát các tấn công do thám kiểu mới (novelty attack) cũng đạt mức trung bình 91% cho các testcase.

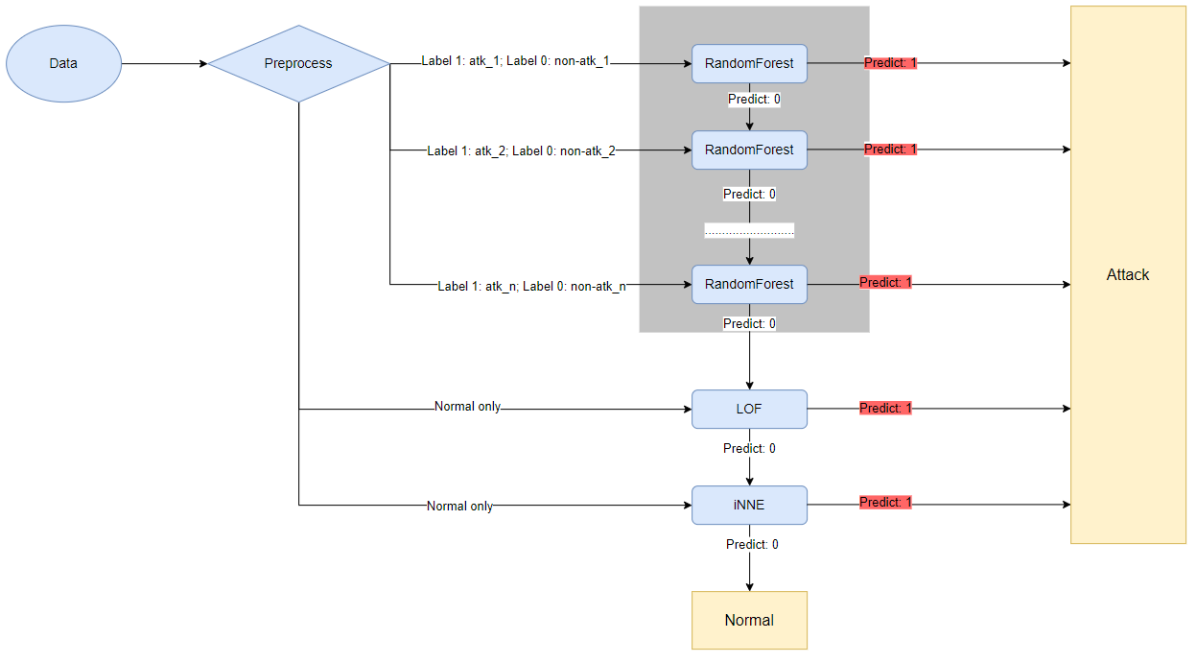
5. Tên sản phẩm:

- Tài liệu báo cáo về tính hiệu quả của hệ thống được đề xuất

6. Hiệu quả, phương thức chuyển giao kết quả nghiên cứu và khả năng áp dụng:

Nghiên cứu hiện tại chỉ mới hoàn tất giai đoạn nghiên cứu ban đầu, tuy vậy tiềm năng và giá trị mang lại của nghiên cứu là rất lớn, hứa hẹn có thể phát triển chiến lược đã đề xuất thành một mô hình kiến trúc hoàn chỉnh và tối ưu hơn. Giải pháp này có thể triển khai tới trong các hệ thống mạng thông tin ở nhiều lĩnh vực kinh tế - xã hội và ở mọi quy mô từ hộ gia đình, tới doanh nghiệp hay tới thành phố thông minh. Nghiên cứu cũng hướng tới ứng dụng rộng rãi cho việc phát hiện và phân loại các mối đe dọa mạng khác chứ không riêng gì các cuộc tấn công do thám.

7. Hình ảnh, sơ đồ minh họa chính




| Cơ quan Chủ trì | Chủ nhiệm đề tài |
|---|---|
| *(ký, họ và tên, đóng dấu)* | *(ký, họ và tên)* |

# TÓM TẮT CÔNG TRÌNH

Trong bối cảnh ngày càng gia tăng của các cuộc tấn công mạng, tấn công do thám (reconnaissance attacks) đóng vai trò quan trọng vì chúng thường là bước đầu tiên trong quá trình xâm nhập hệ thống của kẻ tấn công. Những cuộc tấn công này thu thập thông tin về hệ thống mục tiêu, từ đó tạo điều kiện cho các cuộc tấn công phức tạp hơn. Vì vậy, việc phát hiện và ngăn chặn các cuộc tấn công do thám là cực kỳ quan trọng trong an ninh mạng. Nhằm đáp ứng nhu cầu này, nhóm nghiên cứu đã đề xuất một hệ thống phát hiện tấn công do thám kiểu mới, kết hợp học giám sát và học không giám sát, sử dụng ba mô hình máy học: Random Forest (RF), Local Outlier Factor (LOF) và Isolation-based Anomaly Detection Using Nearest-Neighbor Ensembles (iNNE).

Hệ thống phát hiện xâm nhập mạng (NIDS) được xây dựng dựa trên sự kết hợp của ba mô hình máy học, giúp nâng cao độ chính xác và khả năng phát hiện tấn công do thám. Khi thử nghiệm trên tập dữ liệu CIDDS-001 và CIDDS-002, hệ thống đạt độ chính xác trên 91% cho mọi testcase, chứng minh tính hiệu quả của kiến trúc đề xuất. Một trong những ưu điểm nổi bật của mô hình này là khả năng phát hiện các tấn công kiểu mới với độ chính xác trung bình 91%. Đặc biệt, mô hình có khả năng mở rộng và nâng cấp dễ dàng, chỉ cần huấn luyện lại một số module riêng lẻ khi muốn cập nhật dữ liệu mới, không cần huấn luyện lại toàn bộ hệ thống.

Nghiên cứu này không chỉ mang lại những kết quả thực nghiệm ấn tượng mà còn mở ra nhiều tiềm năng ứng dụng trong tương lai. Mô hình NIDS có thể triển khai trong các hệ thống mạng thông tin ở mọi quy mô, từ hộ gia đình, doanh nghiệp cho đến thành phố thông minh. Việc chuyển giao kết quả nghiên cứu này có thể giúp cải thiện an ninh mạng trong nhiều lĩnh vực kinh tế - xã hội, đồng thời tạo nền tảng cho các nghiên cứu và phát triển tiếp theo. Nghiên cứu cũng hướng tới việc áp dụng rộng rãi cho việc phát hiện và phân loại các mối đe dọa mạng khác, góp phần nâng cao khả năng bảo vệ an ninh mạng một cách toàn diện.

# Mục lục

# NỘI DUNG CÔNG TRÌNH

## 1. ĐẶT VẤN ĐỀ

Trong bối cảnh cách mạng công nghiệp 4.0 đang diễn ra mạnh mẽ, Internet và mạng máy tính ngày càng phát triển mạnh mẽ. Chính vì vậy mà các hoạt động kinh doanh được chuyển sang mô hình số hóa đang được triển khai rộng rãi. Để xây dựng mô hình kinh doanh tốt cần một triển khai một cơ sở hạ tầng thông tin phức tạp. Chính vì sự phức tạp của nó mà dẫn đến những rủi ro và mối đe dọa từ các cuộc tấn công mạng ngày càng tinh vi và đa dạng.

Để nâng cao tính bảo mật của hệ thống mạng trước các cuộc tấn công thì IDS là một giải pháp hàng đầu. Hiện tại, các giải pháp IDS có mặt trên thị trường chỉ phát hiện xâm nhập chủ yếu dựa trên phát hiện các hành vi tấn công đã biết (known-attack) qua việc định nghĩa các signature hoặc profile cơ sở đại diện cho các hành vi bình thường/dự kiến trong mạng,...

Với tình hình như hiện tại, để chống lại các hành vi tấn công mạng tinh vi thì việc nghiên cứu IDS hiệu quả, đáp ứng các yêu cầu bảo mật cao là một nhiệm vụ cấp thiết. Các hệ thống IDS hiện đại không chỉ cần phát hiện các cuộc tấn công đã biết mà còn phải có khả năng nhận diện các mối đe dọa mới, chưa được biết đến. Điều này đòi hỏi sự kết hợp của nhiều công nghệ tiên tiến như học máy, trí tuệ nhân tạo, và phân tích dữ liệu lớn. Đã có rất nhiều các nghiên cứu áp dụng AI vào IDS để nâng cao hiệu quả, song vấn đề trên vẫn là một bài toán nang giải.

Trong các nghiên cứu về chuỗi “cyber kill chain” thì một trong những yếu tố mấu chốt dẫn đến sự thành công của một cuốc tấn công mạng là quá trình thu thập thông tin (reconnaissance) và port scan là một trong những phương pháp thu thập thông tin mạng mà tin tặc thường sử dụng. Như vậy việc ngăn chặn Port Scan giúp cho những kẻ tấn công khó có thể tiếp cận được các thông tin và đặc điểm của mạng nội bộ từ đó giảm thiểu việc xâm nhập mạng một cách nhanh chóng và hiệu quả.

Trong nghiên cứu lần này của nhóm, chúng tôi đem đến một giải pháp Network - based IDS (NIDS) áp dụng Machine Learning nhằm phát hiện các bất thường liên quan đến việc quét mạng (Port Scan). Ngoài việc phát hiện các tấn công quét mạng đã biết thì mô hình này có thể phát hiện được các tấn công chưa biết, các mối đe dọa mới.

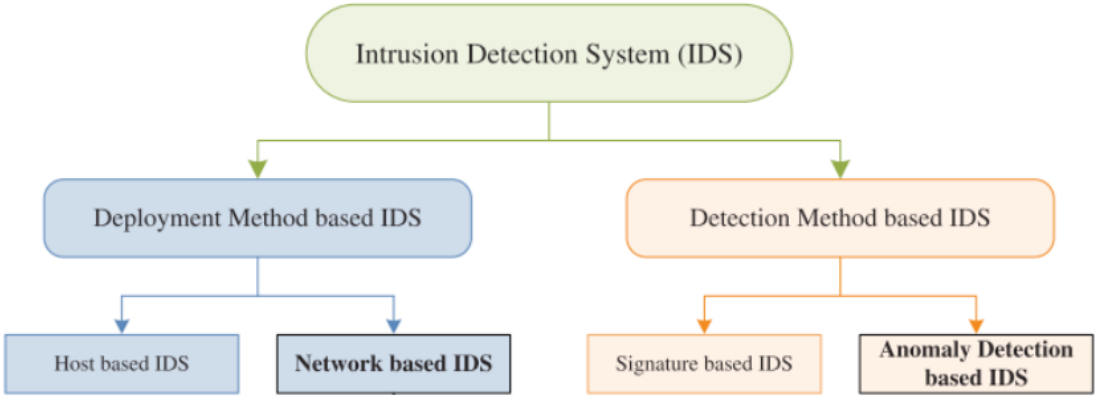
# 2. TỔNG QUAN TÀI LIỆU

## 2.1. Tổng quan kiến thức

### 2.1.1. Tổng quan về Intrusion Detection System – IDS

Theo [8] IDS là sự kết hợp của hai từ "Intrusion" và "Detection System". "Intrusion" đề cập đến sự truy cập trái phép tới thông tin bên trong máy tính hoặc hệ thống mạng nhằm làm tổn hại đến tính toàn vẹn, bảo mật hoặc tính khả dụng của thông tin đó. Trong khi "Detection System"là một cơ chế bảo mật để phát hiện hoạt động bất hợp pháp đó.

Vì vậy, IDS là một công cụ bảo mật liên tục giám sát lưu lượng máy chủ và mạng để phát hiện bất kỳ hành vi đáng ngờ nào vi phạm chính sách bảo mật và làm tổn hại đến tính bảo mật, tính toàn vẹn và tính khả dụng của nó. IDS sẽ tạo cảnh báo về các mối nguy hiểm được phát hiện hành vi đối với máy chủ hoặc quản trị viên mạng.

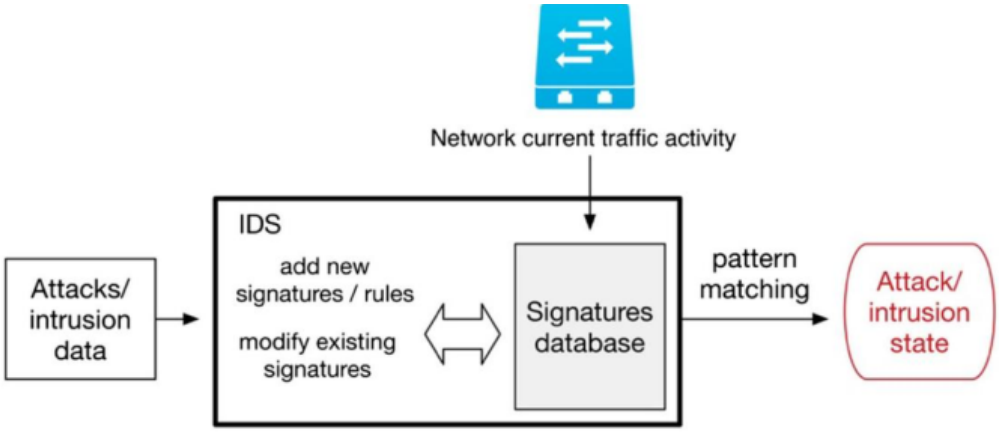



Hình 1. Phân loại IDS

**Các giải pháp IDS truyền thống bao gồm:**

1. **Dựa trên kỹ thuật phát hiện**

- **Signature-based (SIDS)** (hay còn gọi knowledgebased) là một quá trình so sánh các signature với các sự kiện quan sát được để xác định các sự cố có thể có.



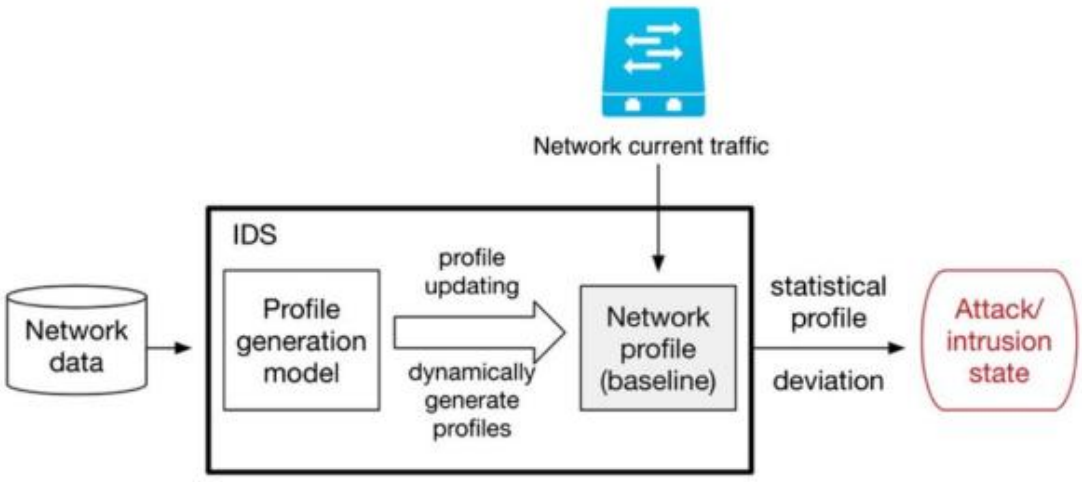

**Hình 2. Mô hình hoạt động Signature-based IDS [10]**

- **Ưu điểm:** Độ chính xác cao khi phát hiện các tấn công đã biết, tỉ lệ cảnh

báo sai thấp.

- **Nhược điểm:** Không thể phát hiện các hành vi bất thường chưa biết trước hoặc các biến đổi nhỏ trong những tấn công đã biết. Yêu cầu phải cập nhật liên tục cơ sở dữ liệu signature**.** Việc triển khai và cập nhật signature khó và tốn thời gian

- **Anomaly-based (AIDS):** (hoặc profile-based) hoạt động dựa trên việc: Tạo ra một profile cơ sở đại diện cho các hành vi bình thường/dự kiến trong mạng. Dựa trên đó, bất kỳ hoạt động mạng đang xem xét nào có sai khác so với profile này đều bị xem là bất thường.



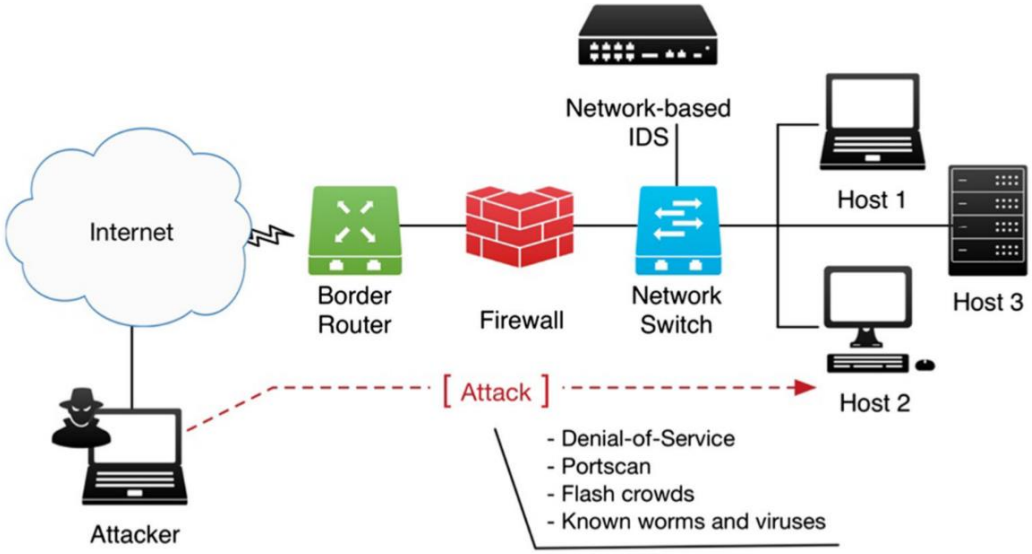

**Hình 3. Mô hình hoạt động Anomaly-based IDS [10]**

- **Ưu điểm:** Phát hiện được cả các hành vi bất thường đã biết và chưa biết, không cần phải có hiểu biết trước. Phát hiện được các tấn công mới (về sau có thể sử dụng trên các signature-based IDS)

- **Nhược điểm**:Tỷ lệ false positives cao (phát hiện nhầm hành vi bình thường là tấn công). Ít hiệu quả quả trong các môi trường mạng động, thay đổi nhiều. Yêu cầu thời gian và tài nguyên để xây dựng được profile đại diện cho mạng.

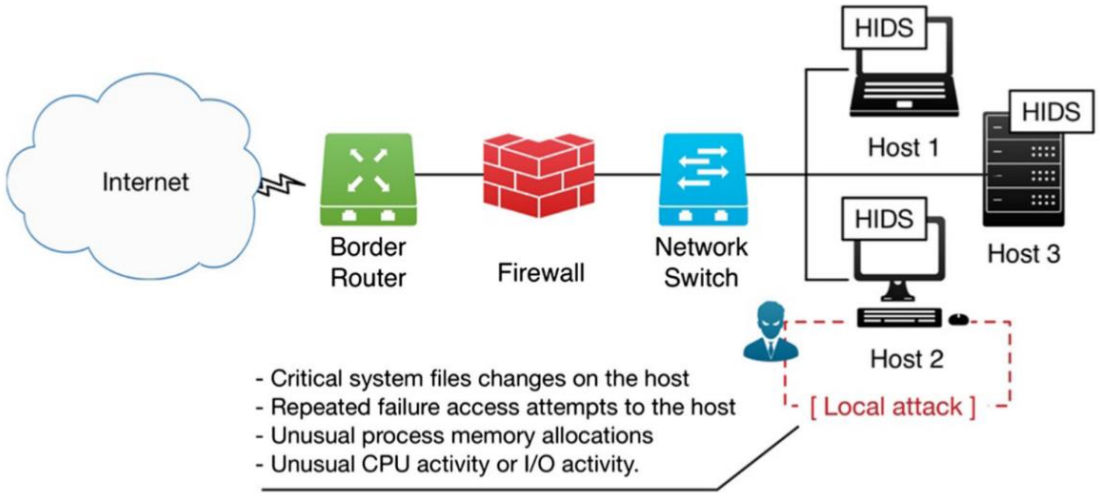
**2. Dựa trên cách triển khai/nguồn dữ liệu**

- **Network-based (NIDS):** theo dõi lưu lượng mạng cho một phần của mạng (network segment) hoặc các thiết bị, phân tích các hoạt động mạng và các giao thức, ứng dụng để xác định các hành vi bất thường. Thường triển khai ở biên mạng, như gần tường lửa hoặc router biên, server VPN, server remote access và mạng không dây




Hình 4. Mô hình hoạt động Network-based IDS cơ bản [10]

- **Host-based (HIDS):** theo dõi các đặc điểm của một host riêng lẻ và các sự kiện xảy ra trong host đó để phát hiện hoạt động bất thường. Được triển khai trên host quan trọng (các server có thể truy cập từ bên ngoài, các server chứa thông tin quan trọng)




Hình 5. Mô hình Host-based IDS cơ bản [10]

**Các thách thức với IDS truyền thống bao gồm:**

1. Kích thước mạng và dữ liệu liên quan ngày càng tăng
2. Thách thức trong việc cải thiện độ chính xác trong phát hiện tấn công, đồng thời giảm tỉ lệ cảnh báo sai.
3. Việc nhiều novel/zero-day attack trở thành trở ngại lớn đối với các IDS truyền thống
4. Các tấn công mạng và các kỹ thuật qua mặt IDS ngày càng phức tạp

**→ Hiểu được những thách thức hiện nại trong lĩnh vực an ninh mạng, các nhà nghiên cứu đã đưa ra một giải pháp tiềm năng dựa trên kỹ thuật Machine Learning và Deep Learning.**

### 2.1.2. Tổng quan Machine Learning – Học máy trong IDS

**Machine learning (ML)** hay máy học là một nhánh của trí tuệ nhân tạo (AI), nó là một lĩnh vực nghiên cứu cho phép máy tính có khả năng cải thiện chính bản thân chúng dựa trên dữ liệu mẫu (training data) hoặc dựa vào kinh nghiệm (những gì đã được học). Machine learning có thể tự dự đoán hoặc đưa ra quyết định mà không cần được lập trình cụ thể.

**Một thuật toán machine learning có:**

- Dữ liệu đầu vào là tập dữ liệu huấn luyện (training dataset)
- Kết quả đầu ra là 1 mô hình (model)
- Model là một thuật toán nhận đầu vào là các dữ liệu mới có cùng định dạng với dữ liệu huấn luyện và đưa ra một dự đoán

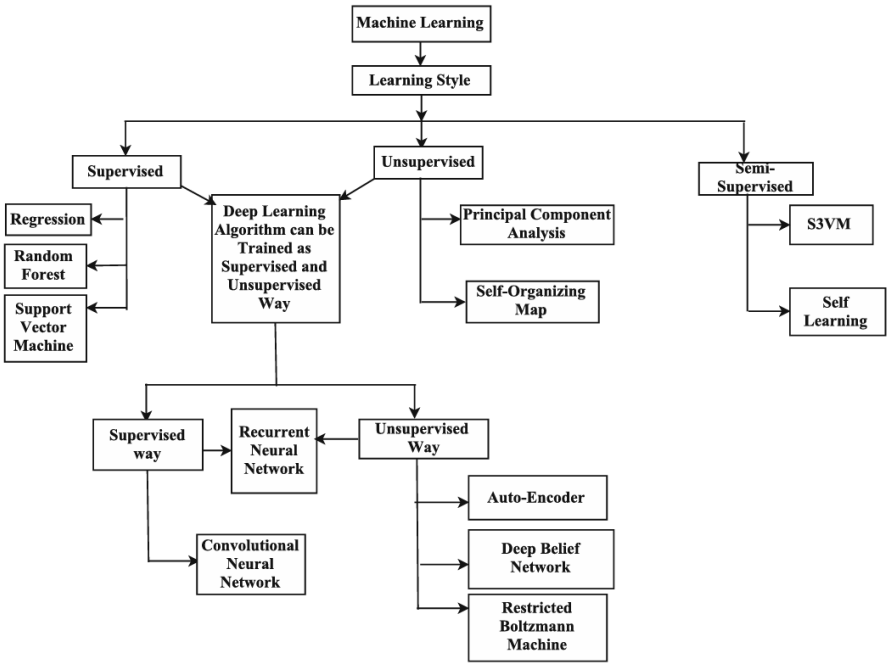
**Một số phương pháp tiếp cận trong IDS được đề cập ở hình. Cụ thể bao gồm:**

- **Supervised learning – Học có giám sát:** là việc cho máy tính học trên dữ liệu đã được gán nhãn (label), hay nói cách khác, với mỗi đầu vào $X_i$, chúng ta sẽ

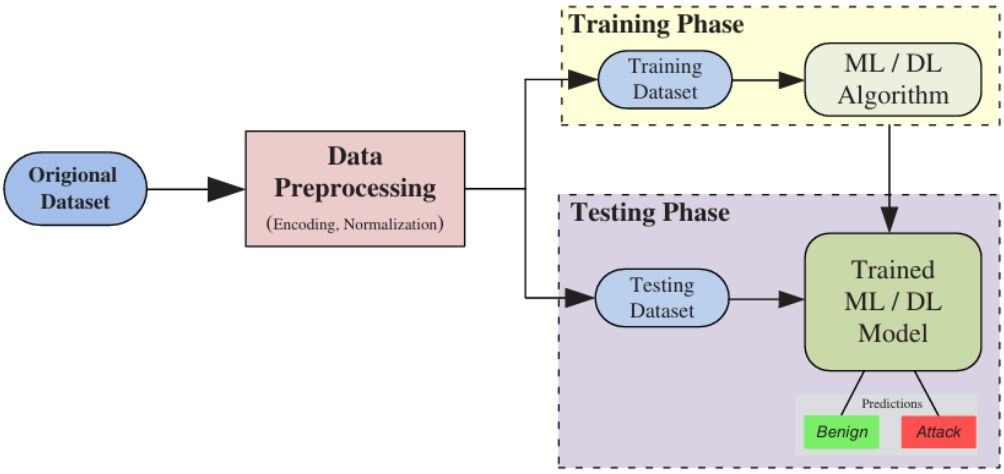
có nhãn $Y_i$ tương ứng. Một số kĩ thuật phổ biến như: Random Forest, Decision Tree, Support Vector Machine.

- **Unsupervised learning – Học không giám sát:** là cho máy tính học trên dữ liệu mà không được gán nhãn, các thuật toán machine learning sẽ tìm ra sự tương quan dữ liệu, mô hình hóa dữ liệu hay chính là làm cho máy tính có kiến thức, hiểu về dữ liệu, từ đó chúng có thể phân loại các dữ liệu về sau thành các nhóm, lớp (clustering) giống nhau mà chúng đã được học hoặc giảm số chiều dữ liệu (dimension reduction). Một số kĩ thuật phổ biến như: K-mean clustering, Principal Component Analysis (PCA).
- **Semi-Supervised Learning – Học bán giám sát:** Dữ liệu đầu vào bao gồm cả dữ liệu đã gán nhãn và chưa gán nhãn.

❖ **Một số cách tiếp cận khác: Reinforcement Learning – Học tăng cường:** đưa ra các dự đoán dựa trên việc thử và sai, dạy cho các máy (agent) thực hiện tốt 1 nhiệm vụ (task) bằng tương tác với môi trường (environment) thông qua hành động (action) và nhận được phần thưởng (reward).



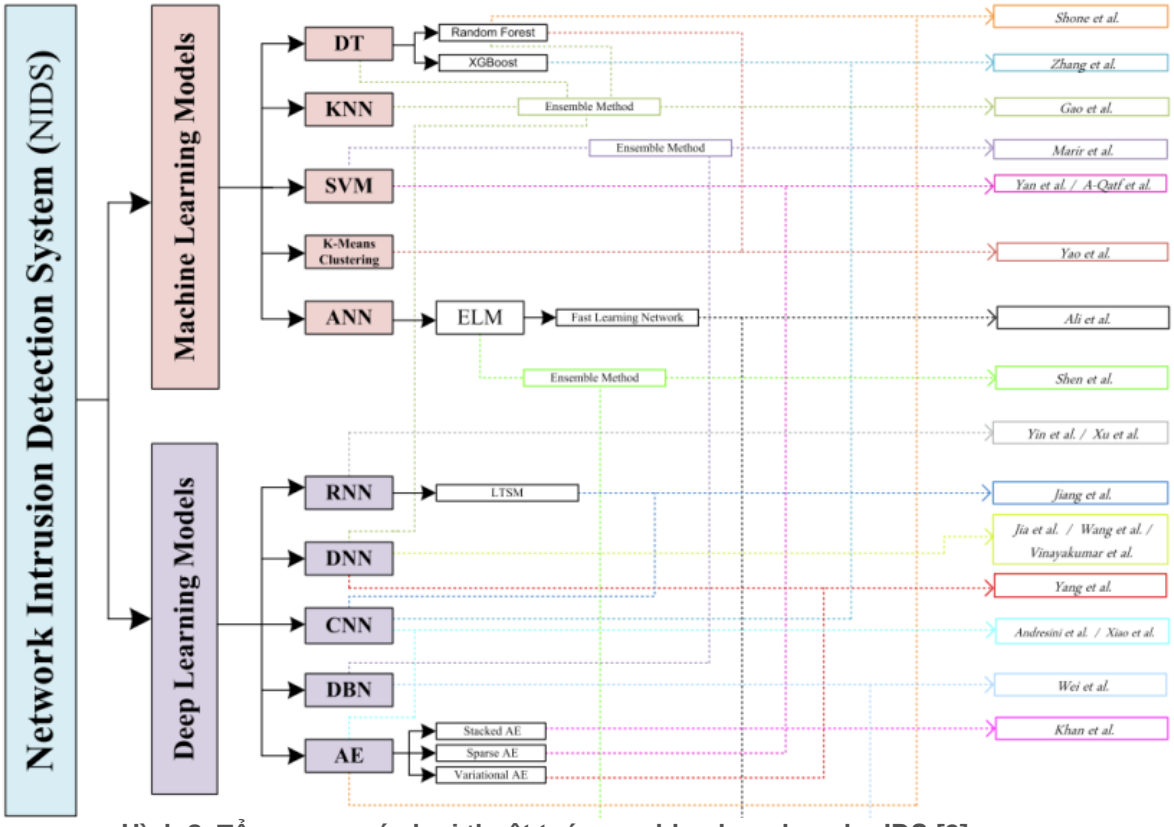

**Hình 6. Tổng quan 1 số hướng tiếp cận Machine Learning [9]**




**Hình 7. Quy trình chung huấn luyện mô hình IDS kết hợp machine learning 3 bước [9]**

**Theo hình trên các bước bao gồm:**

1. ***Bước Tiền xử lý dữ liệu:***
   - Chuyển dữ liệu sang định dạng phù hợp để dùng trong thuật toán
   - Thường bao gồm: Encoding, normalization, data cleaning
   - Chia ngẫu nhiên thành 2 tập dữ liệu training và testing (thường là 80%—20%)
2. ***Bước Training – Huấn luyện:***
   - Thuật toán ML hoặc DL được huấn luyện với tập dữ liệu training
3. ***Bước Testing – Kiểm tra:***
   - Model đã huấn luyện được kiểm tra lại với tập dữ liệu testing và được đánh giá dựa trên các dự đoán mà nó đưa ra




**Hình 8. Tổng quan các loại thuật toán machine learning cho IDS [9]**

### 2.1.3. Tổng quan về Novel Attack

Novel attack là các cuộc tấn công mới hoặc chưa được biết đến trước đây. Các cuộc tấn công mới thường dựa vào các kỹ thuật phức tạp, chiến lược đổi mới hoặc các lỗ hổng zero-day mà nhà cung cấp phần mềm chưa biết hoặc vẫn cần được vá. . Những tấn công này chưa có chữ ký (signature) trong cơ sở dữ liệu của các hệ thống bảo mật truyền thống, khiến chúng trở nên đặc biệt nguy hiểm.

Do tính chất chưa được nhận diện và tài liệu hóa, các cuộc tấn công này không thể bị phát hiện bằng các phương pháp bảo mật truyền thống dựa trên chữ ký (signature-based) hoặc các mẫu hành vi (behavioral-based) đã biết. Việc phát hiện các cuộc tấn công mới đòi hỏi các biện pháp bảo mật nâng cao, chẳng hạn như hệ thống phát hiện bất thường và thuật toán học máy, có thể phân tích lưu lượng truy cập mạng và các mẫu hành vi để xác định hoạt động đáng ngờ hoặc độc hại khác với hành vi thông thường.

## 2.2.Các công trình liên quan

Tập trung vào vấn đề nêu trên đã có rất nhiều những nghiên cứu liên quan đến IDS mà chúng cũng có khả năng phát hiện được các cuộc tấn công chưa được huấn luyện. Trong section này chúng tôi cung cấp một bảng khảo sát ngắn gọn liên quan đến các phương pháp xây dựng hệ thống IDSs theo nhiều ngữ cảnh khác nhau.

**Supervised approach:** Trong lĩnh vực IDSs dựa trên học máy có giám sát, chúng tôi tập trung khảo sát về các giải pháp hiện đại từ năm 2021 trở về sau.Ví dụ trong nghiên cứu [4], tác giả thực hiện thực nghiệm và so sánh phát hiện xâm nhập mạng trong hai cách tiếp cận khác nhau là single flow và multi-flow ngữ cảnh NIDS. Theo tác giả, đối với single flow RF là mô hình cho kết quả tốt nhất với f1-score là 85.04%, trong khi đó đối với multi-flow feature mô hình cho kết quả tốt nhất là LSTM với f1-score là 89.82% Hay trong nghiên cứu [2] tác giả đề xuất mô hình hybrid IDS để phân loại network traffic data trong môi trường big data một cách hiệu quả, đồng thời tác giả cũng nhấn mạnh sự ảnh hưởng của các tập dữ liệu bị mất cân bằng trong các mô hình machine learning, do đó tác giả đưa ra giải pháp xữ lý dữ liệu mất cân bằng nhằm tăng hiệu suất của mô hình phân loại bằng cách sử dụng STL SMOTE oversampling và Tomek-Links undersampling. Để kiểm chứng tính hiệu quả tác giả đã so sánh với chín thuật toán machine learning, deep learning khác nhau, dựa trên kết quả thu được thì phương pháp của tác giả cho hiệu suất ấn tượng hơn.

Tương tự trong nghiên cứu [3], tác giả cũng giới thiệu cách tiếp cận khác trong việc xây dựng NIDS hiệu quả, đó là IGAN. IGAN cũng kết hợp giữa hai phương pháp over và under sampling để xử lý dữ liệu bị mất cân bằng sau đó dữ liệu được huấn luyện với mô hình ensemble của Letnet5 và LTSM để phân loại các loại tấn công. Họ tiến hành thực nghiệm trên 2 tập dataset là UNSW-NB15 và CICIDS 2017, kết quả tổng quan cho thấy tính hiệu quả của phương pháp bao gồm 99.06% precision, 98.17% recall, 99.73% F1-Score và 98.97% accuracy.

**Unsupervised approach**: Trái với supervised NIDSs, dữ liệu huấn luyện của các mô hình unsupervised NIDSs không cần phải gán nhãn. Thêm vào đó, các mô hình này có khả năng xác định được các mẫu chưa được phát hiện trước đó. Trong bài báo [5], nhóm tác giả đánh giá bốn mô hình unsupervised (PCA, Isolation Forest, Auto Encoder, One-Class SVM) trên hai tập dữ liệu CIC-IDS-2017, CSE-CIC-IDS-2018 và thu được kết quả tối trên từng trường hợp. Tuy nhiên, khi thực hiện chiến lược đánh giá mới được đề xuất bằng cách phân loại một tập dữ liệu tương tự nhưng chưa được học, giá trị accuracy giảm trung bình 25.63% so với phương pháp đánh giá thông thường. Hay một mô hình unsupervised mới với tên gọi là ENAD được giới thiệu trong bài báo [6], mô hình là sự kết hợp của các mô hình AutoEncoders (AEs) và generative adversarial networks (GANs). Để phát hiện sự bất thường trong lưu lượng mạng một cách hiệu quả, một chiến lược đánh trọng số nhằm định lượng mức độ quan trọng của từng mô hình đối với từng loại tấn công. Sau cùng, mô hình đưa ra dự đoán dựa trên các trọng số đã được đánh của các mô hình riêng lẻ. Kết quả thực nghiệm của nhóm tác giả đã chứng minh tính hiệu quả của mô hình. Cụ thể, khi thực nghiệm trên hai tập dữ liệu UNSW-NB15 và CICIDS2017, các chỉ số đánh giá hiệu suất như precision, recall, F1-score ... tăng vượt trội 14.70% so với nhiều phương pháp khác.

**Multi Novelty Detection (MND) approach:** Trong ngữ cảnh này chúng tôi thực hiện khảo sát những nghiên cứu tập trung vào các hệ thống IDSs có khả năng phát hiện các cuộc tấn công unknown - các cuộc tấn công chưa được huấn luyện. Tiêu biểu là nghiên cứu [1] tác giả đề suất một framework cho deep learning - based IDSSec để phát hiện cuộc tấn công mới (chẳng hạn như zero-day attack). Tác giả giới

thiệu framework DOC++ (một phiên bản mới của DOC) kết hợp với phương pháp tiền xử lý DID (Deep Intrusion Detection) để cải thiện khả năng cho các thuật toán deep learning có thể phát hiện các tấn công content-based. Kết quả được đánh giá trên tập dữ liệu CIC-IDS2017 và CSE-CIC-IDS2018, so với các thuật toán khác bao gồm DOC, OpenMax và AutoSVM thì kết quả của tác giả là cho hiệu suất tốt hơn. Trong bài báo [7], tác giả cải tiến thuật toán CNN để nhận diện các cuộc tấn công đã biết. Ngoài ra nhóm tác giả cũng đề suất phương pháp sử dụng phân tích hành vi logic để nhận biết các cuộc tấn công unkown trên môi trường private cloud. Tập dataset CIDSS 2017 và Custom data dùng để đánh giá hiệu quả của phương pháp đề suất của tác giả. Kết quả thu được trên các tập dữ liệu trên đều out-performance mô hình LSTM và DNN. Đặc biệt unknown detection rate đề suất đạt được accuracy trên 90%, f1-scores đạt từ 50% - 70% và recall đạt từ 70% - 80%.

## 2.3. Tính khoa học tính mới

Tổng kết lại, đã có rất nhiều nghiên cứu xoay quanh vấn đề xây dựng một hệ thống phát hiện xâm nhập. Tuy nhiên các nghiên cứu về phát hiện và ngăn chặn tấn công do thám vẫn còn hạn chế, đặc biệt là việc phát hiện tấn công kiểu mới trong do thám (novel attack). Nhóm nghiên cứu nhận định rằng việc phát hiện và ngăn chặn các cuộc tấn công do thám là vô cùng quan trọng, nó giúp hạn chế , thậm chí là ngăn chặn khả năng thu thập dữ liệu phục vụ cho các mục đích xấu như tấn công hệ thống mạng. Do đó, trong nghiên cứu này, nhóm nghiên cứu đã xây dựng một kiến trúc NIDS có thể xử lí theo thời gian thực, phân loại các cuộc tấn công do thám đã biết và phát hiện được các cuộc tấn công do thám kiểu mới.

Nhóm nghiên cứu ứng dụng các mô hình học máy bao gồm Random Forest, Local Outlier Factor (LOF) và Isolation using Nearest Neighbor Ensemble (iNNE) để xây dựng một NIDS có thể phát hiện và phân loại các cuộc tấn công do thám mạng với độ chính xác cao. Điểm đặc biệt là hệ thống có khả năng phát hiện các cuộc tấn công kiểu mới mà chưa được huấn luyện qua các mô hình máy học.

Một trong những ưu điểm của đề xuất là khả năng mở rộng và nâng cấp dễ dàng của mô hình NIDS. Khi cần mở rộng hoặc cập nhật dữ liệu về các cuộc tấn công, chỉ cần tiến hành huấn luyện lại từng module riêng lẻ hoặc thêm mới các module đã huấn luyện vào mô hình, không cần huấn luyện lại toàn bộ mô hình. Điều này giúp tiết kiệm thời gian và tài nguyên, đồng thời đảm bảo hệ thống luôn được cập nhật và hiệu quả trong việc đối phó với các mối đe dọa an ninh mạng ngày càng phức tạp.

# 3. MỤC TIÊU – PHƯƠNG PHÁP

## 3.1. Mục tiêu công trình

Như đã đề cập ở trên, mục tiêu của nghiên cứu này là phát triển một hệ thống phát hiện xâm nhập mạng (Network Intrusion Detection System - NIDS) sử dụng Machine Learning để phát hiện các bất thường liên quan đến việc quét mạng (Scanning Attack). Hệ thống này không chỉ nhằm phát hiện các cuộc tấn công quét mạng đã biết mà còn có khả năng nhận diện các cuộc tấn công mới và các mối đe dọa tiềm ẩn chưa từng được biết đến trước đây. Chi tiết hơn, các mục tiêu của nghiên cứu bao gồm:

**Phát triển và đánh giá các mô hình Machine Learning:** Tạo ra các mô hình học máy có khả năng phát hiện các hoạt động quét mạng dựa trên dữ liệu mạng, đồng thời đánh giá hiệu quả của các mô hình này trong việc phát hiện các cuộc tấn công đã biết và chưa biết.

**Tăng cường khả năng phát hiện tấn công chưa biết:** Khai thác các kỹ thuật Machine Learning tiên tiến, phổ biến để tăng cường khả năng của hệ thống trong việc nhận diện các kiểu tấn công mới và các mối đe dọa tiềm ẩn.

**Tối ưu hóa hiệu năng và độ chính xác của hệ thống NIDS:** Đảm bảo rằng hệ thống phát hiện xâm nhập không chỉ chính xác mà còn hiệu quả về mặt thời gian xử lý và tài nguyên sử dụng, nhằm đảm bảo khả năng ứng dụng thực tế trong các môi trường mạng khác nhau.

**Đánh giá và so sánh với các phương pháp học máy truyền thống:** So sánh hiệu quả của hệ thống phát hiện xâm nhập dựa trên Machine Learning với các phương pháp phát hiện tấn công truyền thống, để xác định các ưu điểm và nhược điểm của phương pháp mới.

Và đồng thời thông qua các mục tiêu trên, nhóm thực hiện mong muốn nghiên cứu này có thể đóng góp vào việc cải thiện an ninh mạng, đặc biệt là trong việc phát hiện và ngăn chặn các cuộc tấn công quét mạng một cách hiệu quả và kịp thời và có khả năng đáp ứng với sự phát triển của các cuộc tấn công mạng này trong tương lai.

## 3.2. Tổng quan giải pháp

Giải pháp được nhóm triển khai dựa trên một mô hình học máy kết hợp sử dụng Random Forest và các phương pháp phát hiện điểm bất thường như LOF và iNNE. Cụ thể:

- Dữ liệu đầu vào được tiền xử lý và chuẩn bị để đưa vào mô hình.
- Sau đó Sử dụng nhiều mô hình Random Forest để phát hiện các loại tấn công cụ thể. Mỗi mô hình được huấn luyện để phân loại một loại tấn công cụ thể**.** Song song với mô hình Random Forest, các phương pháp phát hiện điểm bất thường như LOF và iNNE được áp dụng để phân loại các mẫu dữ liệu bất thường (abnormal) so với phân phối chung.
- Kết hợp các kết quả từ mô hình RandomForest, LOF và iNNE để đưa ra quyết định cuối cùng về việc liệu lưu lượng mạng có phải là tấn công hay không. Nếu bất kỳ mô hình nào phát hiện một tấn công, lưu lượng sẽ được gán nhãn

là tấn công. Ngược lại, nếu không mô hình nào phát hiện bất thường, lưu lượng sẽ được gán nhãn là bình thường. Điều này cung cấp một cách tiếp cận toàn diện để phát hiện và phân loại các hành vi tấn công từ dữ liệu mạng.

## 3.3.Nội dung và phương pháp thực hiện

### 3.3.1. Phương pháp trích xuất đặc trưng dữ liệu mạng

Trong hai bài báo [11], [12] của nhóm tác giả Markus Ring về 2 tập dữ liệu CIDDS 001 và CIDDS 002, họ đã phát triển một phương pháp để tạo ra các bộ dữ liệu nhãn cho hệ thống phát hiện xâm nhập (IDS) dựa trên luồng mạng (NetFlow).

Môi trường ảo hóa dựa trên OpenStack được sử dụng để mô phỏng một môi trường doanh nghiệp nhỏ với nhiều máy chủ và máy khách. Các hoạt động mạng bình thường được mô phỏng bằng các script Python chạy trên các máy khách, thực hiện các hoạt động như duyệt web, gửi email, hoặc in ấn. Các luồng mạng được ghi lại dưới dạng định dạng NetFlow một chiều tại router trong môi trường OpenStack.

Để tạo ra lưu lượng mạng bình thường, các script Python được thiết kế để bắt chước hành vi người dùng thực tế, bao gồm cả việc tuân thủ giờ làm việc và giờ nghỉ trưa. Lưu lượng mạng độc hại được tạo ra bằng cách thực hiện các cuộc tấn công như Tấn công từ chối dịch vụ (DoS), Brute Force, và Quét cổng (Port Scans) trong mạng ảo. Một máy chủ bên ngoài được triển khai để thu thập lưu lượng mạng thực tế từ Internet, bao gồm các cuộc tấn công cập nhật và mới nhất.

Các luồng mạng được ghi lại dưới dạng định dạng NetFlow một chiều, giúp giảm bớt lượng dữ liệu cần phân tích và tránh các vấn đề liên quan đến kết nối mã hóa. Các luồng mạng sau đó được phân tích để trích xuất các đặc trưng liên quan đến cả lưu lượng mạng bình thường và độc hại.

Do nguồn gốc (source), mục tiêu (destination), và thời điểm của các cuộc tấn công đã được biết trước, việc gán nhãn cho dữ liệu NetFlow đã ghi lại trở nên dễ dàng. Dữ liệu được phân loại và gán nhãn thành các luồng mạng bình thường và độc hại, giúp tạo ra một bộ dữ liệu có nhãn đầy đủ và chính xác.

### 3.3.2. Tập dữ liệu

Nhóm tiến hành thực nghiệm và đánh giá mô hình dựa trên tập dataset là CIDSS (Coburg Intrusion Detection Data Sets). Với mỗi tập dataset chúng tôi sẽ lọc và nhóm các Scanning attack theo các giao thức như TCP, UDP, ICMP, ... Tập dữ liệu CIDDS bao gồm 2 phiên bản là CIDDS 001 và CIDDS 002 được ra mắt vào năm 2017 bởi Markus Ring and Sarah Wunderlich với mục đích đánh giá hiệu quả của các mô hình IDS. Dataset được triển khai trên môi trường small business được giả lập trên công cụ ảo hóa OpenStack bao gồm một vài clients và server như Email server hoặc Web Server và được đánh nhãn dựa trên luồng dữ liệu (flow-based). Trong đó CIDSS 001 dataset bao gồm Port Scan, Ping Scan, Dos, BruteForce. Còn đối với CIDDS 002 là một phiên bản nâng cấp hơn và tập trung nhiều về port scan hơn. Trong bộ dataset CIDDS-002, họ thực hiện tổng cộng 43 cuộc tấn công cổng từ bên trong môi trường

mạng nội bộ. Cả hai bộ dataset đều bao gồm 14 features được cung cấp trong file csv bao gồm các feature về network, Protocol type, và thông tin bổ sung thêm về script paramester, chi tiết hơn các Features nằm ở bảng dưới. Tập dữ liệu này được thu thập trong thời gian dài trong khoảng thời gian bốn tuần cho CIDDS-001 và hai tuần đối với CIDDS-002 bằng công cụ NetFlow tạo một nguồn dữ liệu phong phú và đa dạng.

**Bảng 1. Bảng mô tả các đặc trưng trong tập dữ liệu**

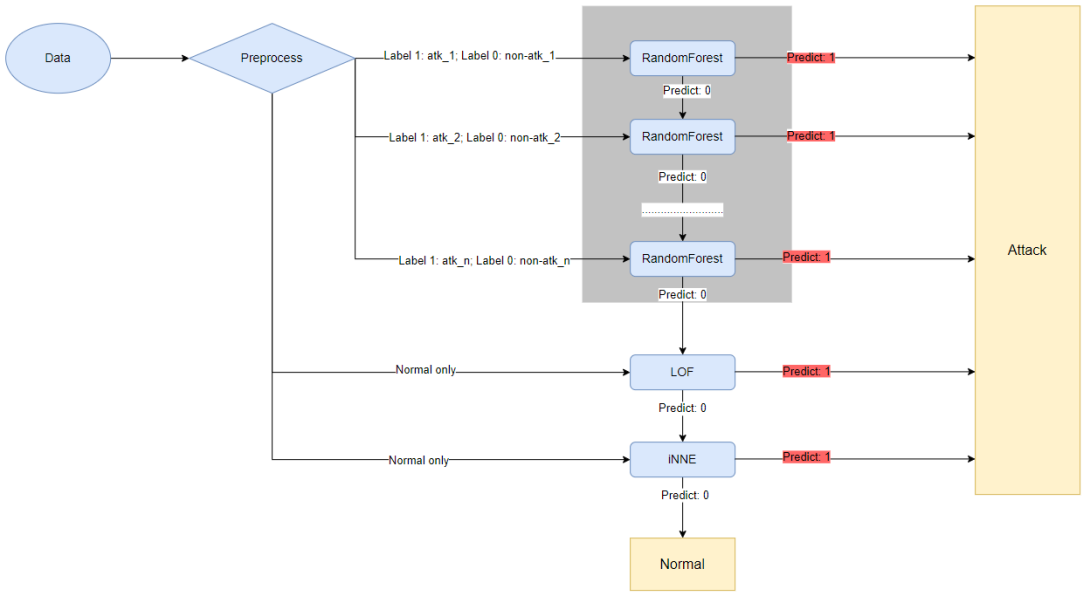
| STT | Tên trường | Mô tả |
|---|---|---|
| 1 | Src IP | Địa chỉ IP nguồn |
| 2 | Src Port | Port nguồn |
| 3 | Dest IP | Địa chỉ IP đích |
| 4 | Dest Port | Port nguồn |
| 5 | Proto | Giao thức trao đổi dữ liệu (e.g. ICMP, TCP, or UDP) |
| 6 | Date first seen | Thời gian bắt đầu lưu lượng được nhìn thấy đầu tiên |
| 7 | Duration | Thời gian tồn tại của lưu lượng |
| 8 | Bytes | Số byte được truyền |
| 9 | Packets | Số gói tin được truyền |
| 10 | Flags | OR concatenation of all TCP Flags |
| 11 | Class | Nhãn các lớp (normal, attacker, victim, suspicious or unknown) |
| 12 | AttackType | Loại tấn công (portScan, dos, bruteForce, —) |
| 13 | AttackID | ID tấn công duy nhất. Tất cả các lưu lượng thuộc về cùng một tấn công sẽ có cùng ID tấn công |
| 14 | AttackDescription | Cung cấp thông tin bổ sung về các thông số tấn công được thiết lập (ví dụ: số lần đoán mật khẩu cho các tấn công SSH-Brute-Force) |

### 3.3.3. Mô hình phát hiện và phân loại tấn công do thám

#### 3.3.3.1. Kiến trúc mô hình phát hiện và phân loại tấn công do thám

Kiến trúc mô hình được chia làm 2 thành phần dựa trên chức năng: (i) phân loại các cuộc tấn công đã biết và (ii) phát hiện các cuộc tấn công do thám kiểu mới. Nhóm nghiên cứu đã sử dụng mô hình Random Forest để giải quyết vấn đề phân loại các

cuộc tấn công do thám đã biết. Sau khi đã phân loại được các cuộc tấn công đã biết với một loại các model Random Forest, phần dữ liệu còn lại sẽ được xử lý bởi mô hình LOF và iNNE để phát hiện các mẫu ngoại lai so với dữ liệu mạng thông thường.




**Hình 9. Tổng quan giải pháp NIDS đề xuất**

Về mô hình Random Forest, đây là mô hình rất nổi tiếng và khá quen thuộc trong các cuộc thi về huấn luyện mô hình học máy cũng như được ứng dụng rất rộng rãi vào các công trình nghiên cứu, các sản phẩm ứng dụng về phân loại, pháp hiện xâm nhập, tấn công mạng. Mô hình Random Forest là sự kết hợp giữa nhiều mô hình Decision Tree được huấn luyện trên các tập dữ liệu khác nhau được rút ra từ tập huấn luyện. Mô hình Random Forest có ưu điểm đó là giảm thiểu được hiện tượng overfitting do có phương sai thấp và ít bị ảnh hưởng bởi nhiễu như mô hình Decision Tree. Khi huấn luyện mô hình, mô hình Random Forest cũng giúp chúng ta đánh giá nhanh tầm quan trọng của các đặc trưng đối với việc phân loại. Điều này cực kì hữu ích đối với những bộ dữ liệu có số chiều lớn. Thuật toán có thể được sử dụng trong cả hai vấn đề phân loại và hồi quy. Qua đó, ta thấy mô hình Random Forest có độ chính xác cao và hiệu quả với bài toán phân loại dựa trên chiến lược học giám sát. Ngoài ra, so với một số mô hình phức tạp khác, Random Forest tương đối đơn giản và dễ hiểu. Việc giải thích các quyết định của mô hình cũng trở nên dễ dàng hơn.

Local Outlier Factor (LOF) là một thuật toán học không giám sát có chức năng phát hiện bất thường dựa trên mật độ, được giới thiệu bởi Markus M. Breunig và các cộng sự năm 2000. Điểm nổi bật của LOF là khả năng đánh giá mức độ bất thường của một điểm dữ liệu dựa trên mật độ của nó so với mật độ của các điểm lân cận. Cụ thể, LOF tính toán một giá trị gọi là "local reachability density" (LRD) cho mỗi điểm dữ liệu, sau đó so sánh LRD của điểm đó với LRD của các điểm lân cận để xác định xem điểm đó có phải là một outlier hay không. Điểm LOF càng cao, điểm dữ liệu đó càng có khả năng là một bất thường. LOF có ưu điểm là không yêu cầu giả định về

phân phối của dữ liệu và có thể xử lý các tập dữ liệu không đồng nhất và có độ nhiễu cao.

Isolation using Nearest Neighbor Ensemble (iNNE) là một thuật toán học không giám sát được thiết kế dựa trên kỹ thuật nearest-neighbors để giải quyết vấn đề phát hiện outlier. Thuật toán này có cùng ý tưởng với Isolation Forest (iForest) trong việc chia dữ liệu thành các tập con và tính toán điểm cô lập (isolation score) cho mỗi mẫu dựa trên mỗi tập con. Tuy nhiên, thay vì cô lập các mẫu riêng lẻ dựa trên các đặc trưng, iNNE cô lập các mẫu bằng cách chỉ chứa mẫu tương ứng ở trung tâm của một siêu hình cầu (hypersphere). Kích thước của các siêu hình cầu được tính dựa trên khoảng cách của mẫu dữ liệu đến điểm gần nhất của nó để thích nghi với phân phối cục bộ. Nói cách khác, một vùng thưa thớt sẽ có các siêu hình cầu lớn hơn và ngược lại đối với một vùng dày đặc. Đặc điểm này được sử dụng để phát hiện outlier vì các outlier thường phân bố xa khỏi các vùng dày đặc của các inliers. Kích thước của các siêu hình cầu cũng tương ứng với chiều dài đường đi được sử dụng trong iForest, vì cả hai đều được dùng để xác định outlier. Tuy nhiên, ưu điểm của iNNE là thuật toán sử dụng tất cả các đặc trưng để phân vùng các khu vực, do đó tận dụng được mối quan hệ giữa các đặc trưng và tránh được các hạn chế của phương pháp iForest.

#### 3.3.3.2. Tiền xử lý dữ liệu

Sau khi thu thập các file csv CIDDS 001 và CIDDS 002, nhóm tiến hành xử lý và chọn lọc ra các mẫu bình thường (normal) và các mẫu tấn công do thám (scan/portscan/pingscan). Để kết quả mang tính khách quan và tránh bị nhiễu nhóm thực hiện loại bỏ các trường thông tin liên quan đến IP nguồn và IP đích. Tiếp đó, nhóm nghiên cứu tiến hành phân loại và đánh nhãn lại dữ liệu tấn công do thám dựa theo protocol theo như bảng bên dưới. Tiến hành One-hot encode cho các trường dữ liệu là Flags và attackDescription. Các trường attackDescription (bao gồm attackDescription_ICMP, attackDescription_TCP, attackDescription_UDP, attackDescription_normal) sẽ được sử dụng cho việc huấn luyện và đánh giá mô hình theo từng module nhỏ như đã trình bày ở đầu.

**Bảng 2. Bảng mô tả các nhãn**

| Dataset | Origin Label | New Label | Protocol | Final Label |
|---|---|---|---|---|
| **CIDDS** | Normal | 0_normal | Any | 0_normal |
| | Victim | Attack | ICMP | ICMP |
| | Attacker | | TCP | TCP |
| | | | UDP | UDP |

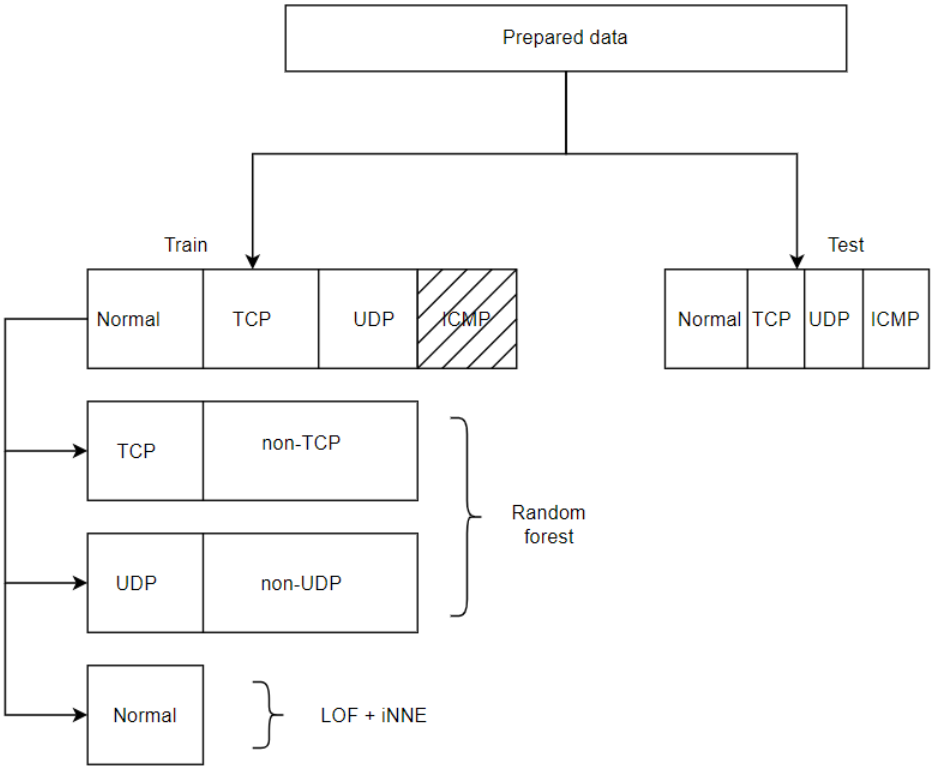
#### 3.3.3.3. Phương pháp huấn luyện mô hình

Từ tập dữ liệu đã chuẩn bị, nhóm nghiên cứu sử dụng 30% số lượng gói tin như là tập kiểm thử để đánh giá mô hình, 70% số lượng gói tin còn lại như là tập huấn luyện.

Tập huấn luyện sẽ được lược bỏ 1 hoặc 2 loại tấn công và xem đó là tấn công do thám kiểu mới mà mô hình NIDS chưa được huấn luyện. Cụ thể sẽ có 6 trường hợp:

**Bảng 3. Bảng các trường hợp.**

| Test Case | Trường giá trị sẽ drop | Giá trị của trường |
|---|---|---|
| 1 | attackDescription_ICMP | 1 |
| 2 | attackDescription_TCP | 1 |
| 3 | attackDescription_UDP | 1 |
| 4 | attackDescription_ICMP hoặc attackDescription_TCP | attackDescription_ICMP = 1 hoặc attackDescription_TCP = 1 |
| 5 | attackDescription_ICMP hoặc attackDescription_UDP. | attackDescription_ICMP = 1 hoặc attackDescription_UCP = 1 |
| 6 | attackDescription_TCP hoặc attackDescription_UDP. | attackDescription_TCP = 1 hoặc attackDescription_UDP = 1 |




**Hình 10. Minh họa huấn luyện mô hình với test case 1 (Drop ICMP)**

Tập huấn luyện sau đó sẽ được sử dụng để huấn luyện cho các mô hình Random Forest, LOF và iNNE. Ví dụ với trường hợp (1), tập huấn luyện sau khi đã loại bỏ các mẫu có giá trị là 1 ở trường attackDescription_ICMP. Sau

đó, tập huấn luyện sẽ được dung để huấn luyện 2 mô hình Random Forest với nhãn (label) khác nhau cho môi mô hình, nhãn lần lượt sẽ là attackDescription_TCP và attackDescription_UDP. Mục đích chính của 2 mô hình Random Forest chính là phát hiện và phân loại tấn công do thám có protocol là TCP và UDP đã biết. Cuối cùng, chọn ra các mẫu lưu lượng mạng thông thường (normal) để huấn luyện cho mô hình LOF và iNNE nhằm mục dích phát hiện các lưu lượng mạng bất thường.

### 3.3.4. Phương pháp thực nghiệm và đánh giá

#### 3.3.4.1. Phương pháp thực nghiệm

Để đánh giá xem NIDS mà tác giả thiết kế liệu có đạt được những mục tiêu chính đã đề ra hay không, nhóm nghiên cứu sẽ huấn luyện mô hình đã đề xuất trong 6 trường hợp như bảng bên dưới, sau đó đánh giá với bộ dữ liệu dùng cho quá trình kiểm tra. Trong quá trình chạy, nhóm nghiên cứu sẽ ghi nhận lại các chỉ số cần thiết để đánh giá độ chính xác của mô hình.

**Bảng 4. Testcase đánh giá mô hình**

| Dataset | Test Alias | Label drop in training set |
|---|---|---|
| **CIDDS 001 002** | A1 | ICMP |
| | A2 | TCP |
| | A3 | UDP |
| | A4 | ICMP + UDP |
| | A5 | UDP + TCP |
| | A6 | TCP + ICMP |

#### 3.3.4.2. Phương pháp đánh giá

Để đánh giá một mô hình với những mô hình khác, cần tới các chỉ số đánh giá chung. Trong phạm vi nghiên cứu lần này nhóm triển khai đánh giá trên phân loại nhị phân và phân loại đa lớp.

Ngoài ra Trong đề tài này, nhóm sẽ tập trung đánh giá các chỉ số quan trọng đối với một NIDS là TPR, FPR và ACC theo công thức được diễn giải sau đây:

- **True Positive Rate (TPR hoặc Recall hoặc Detection Rate):** là tỷ lệ các phát hiện chính xác trên tổng số tất cả các trường hợp tấn công thực tế đã có.

$$TPR = \frac{TP}{TP + FN}$$

- **False positive rate (FPR hoặc Fallout hoặc Fall Alert):** xác định tỉ lệ trường hợp bình thường nhưng bị cảnh báo là tấn công.

$$FPR = \frac{FP}{FP + TN}$$

- **Accuracy (ACC):** là tỉ lệ dự đoán đưa ra đúng với nhãn trên tổng số dự đoán được đưa ra.

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

- **F1-Score:** là trung bình điều hòa (harmonic mean) giữa hai chỉ số Recall và Precison. F1-Score chỉ cao khi cả hai 2 chỉ số Precision và Recall đều cao.

$$F1 = \frac{2 * Recall + Precision}{Recall * Precision}$$

- **MCC:** là độ đo sử dụng trong các bài toán phân loại nhị phân. Nó là một phép đo tổng hợp các kết quả phân loại từ confusion matrix của mô hình. MCC có giá trị từ -1 đến +1, trong đó giá trị +1 đại diện cho sự phân loại hoàn hảo, 0 đại diện cho phân loại ngẫu nhiên và -1 đại diện cho sự phân loại nghịch đảo. MCC cũng thường được sử dụng khi dữ liệu mất cân bằng giữa hai nhãn.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

- **Positive Predictive Value (PPV) hoặc Precision:** Đây là tỷ lệ các trường hợp được dự đoán là tấn công mà thực sự là tấn công trên tổng số các trường hợp được dự đoán là tấn công. PPV cho biết độ chính xác của các dự đoán tấn công, tức là khả năng mô hình không cảnh báo sai.

$$PPV = \frac{TP}{TP + FP}$$

- **Area Under the Curve (AUC):** AUC là một chỉ số đánh giá chất lượng của mô hình phân loại nhị phân. Nó đo lường khả năng mô hình phân biệt được giữa các lớp. AUC có giá trị từ 0 đến 1, với 1 là phân loại hoàn hảo, 0.5 là phân loại ngẫu nhiên.

- **Unknown Detection Rate (UDR):** UDR đo lường khả năng mô hình phát hiện các loại tấn công mà nó chưa từng được huấn luyện, tức là các mẫu unknown. Một NIDS tốt cần có khả năng phát hiện các loại tấn công mới, chưa biết trước.

- **UDR = Số lượng mẫu unknown được dự đoán chính xác / Tổng số lượng mẫu chưa biết (Unknown)**

# 4. KẾT QUẢ - THẢO LUẬN

## 4.1. Môi trường thực hiện

***Cấu hình máy thực hiện: Google Colab***

- CPU: Intel(R) Xeon(R) CPU @ 2.20GHz
- RAM: 12.67 GB
- Ổ cứng: 107.7 GB

***Môi trường phát triển:***

- Trình soạn thảo: Notebook, Visual Studio.
- Ngôn ngữ lập trình: Python.
- Thư viện sử dung: numpy, pandas, mathplotlib, scikit-learn, pyod.

## 4.2. Kết quả nghiên cứu

Để so sánh độ hiệu quả của mô hình đề xuất nhóm đã tiến hành chạy thực nghiệm với 9 mô hình máy học nổi tiếng bao gồm: XGBoost (XGB), Support Vector Machine (SVM), CNN, Light GBM (LGBM), Naive Bayes (NB), Decision Tree (DT), Random forests (RF), KNN và AdaBoost classifier (Ada). Nhóm sử dụng cùng bộ dữ liệu và phương pháp đánh giá cho quá huấn luyện và thử nghiệm. Sau đó chạy so sánh với hiệu suất của phương pháp mà nhóm đề xuất. Kết quả chi tiết ở hình dưới:

**Bảng 5. Kết quả so sánh giữa 9 mô hình khác nhau với mô hình do nhóm nghiên cứu đề xuất trong Test Case A1**

| TEST CASE A1 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Evaluate / Models | UDR | ACC | TPR | FPR | F1 | AUC | MCC | PPV |
| Decision Tree | 0.5192 | 0.842 | 0.9957 | 0.3167 | 0.8374 | 0.8395 | 0.7175 | 0.8791 |
| Random Forest | 0.4850 | 0.9482 | 1 | 1 | 0.948 | 0.9475 | 0.8994 | 0.9518 |
| LightGBM | 0.5213 | 0.9607 | 0.9934 | 0.073 | 0.9606 | 0.9602 | 0.9233 | 0.9631 |
| XGBoost | 0.5187 | 0.9656 | 0.9946 | 0.0642 | 0.9655 | 0.9651 | 0.9327 | 0.9676 |
| SVM | 0.0013 | 0.1298 | 0.1204 | 0.8605 | 0.1297 | 0.1299 | -0.7403 | 0.1296 |
| Naïve Bayes | 0.6919 | 0.9283 | 0.9209 | 0.0639 | 0.9283 | 0.9285 | 0.8569 | 0.9284 |
| AdaBoost | 0.6518 | 0.9616 | 0.9933 | 0.071 | 0.9615 | 0.9611 | 0.925 | 0.9639 |
| KNN | 0.8168 | 0.9866 | 0.9905 | 0.0175 | 0.9866 | 0.9865 | 0.9732 | 0.9866 |
| CNN | 0.3218 | 0.7833 | 0.8841 | 0.0471 | 0.7713 | 0.7797 | 0.6235 | 0.7022 |
| **Ours (Binary Classification)** | **1** | **0.9489** | **0.9505** | **0.0965** | **0.9489** | **0.9505** | **0.9024** | **0.9518** |
| **Ours (Multi-Classification)** | **1** | **0.9465** | **0.9656** | **0.0965** | **0.8799** | **N/A** | **0.9158** | **0.8531** |

Bảng 6. Kết quả so sánh giữa 9 mô hình khác nhau với mô hình do nhóm nghiên cứu đề xuất trong Test Case A2

**TEST CASE A2**

| Evaluate / Models | UDR | ACC | TPR | FPR | F1 | AUC | MCC | PPV |
|---|---|---|---|---|---|---|---|---|
| **Decision Tree** | 0.0910 | 0.6295 | 0.9999 | 0.7531 | 0.5644 | 0.6234 | 0.3777 | 0.7889 |
| **Random Forest** | 0.0327 | 0.6029 | 1 | 1 | 0.5241 | 0.5964 | 0.3198 | 0.7651 |
| **LightGBM** | 0.0280 | 0.604 | 1 | 0.805 | 0.523 | 0.5975 | 0.331 | 0.781 |
| **XGBoost** | 0.0280 | 0.604 | 1 | 0.8049 | 0.523 | 0.5975 | 0.3311 | 0.781 |
| **SVM** | 0.0184 | 0.0912 | 0.1648 | 0.9848 | 0.0858 | 0.0899 | -0.8276 | 0.0823 |
| **Naïve Bayes** | 0.0297 | 0.5915 | 0.9815 | 0.8113 | 0.5109 | 0.585 | 0.2807 | 0.7316 |
| **AdaBoost** | 0.0280 | 0.6039 | 0.9999 | 0.8051 | 0.5228 | 0.5973 | 0.3307 | 0.7808 |
| **KNN** | 0.3074 | 0.7172 | 0.9998 | 0.5747 | 0.6895 | 0.7125 | 0.5224 | 0.821 |
| **CNN** | 0.5175 | 0.802 | 0.9996 | 0.4017 | 0.802 | 0.7989 | 0.6556 | 0.789 |
| **Ours (Binary Classification)** | **0.9113** | **0.9168** | **0.9172** | **0.0965** | **0.9167** | **0.9172** | **0.8339** | **0.9167** |
| **Ours (Multi-Classification)** | **0.9113** | **0.9144** | **0.9445** | **0.0954** | **0.9464** | **N/A** | **0.8534** | **0.9488** |

Bảng 7. Kết quả so sánh giữa 9 mô hình khác nhau với mô hình do nhóm nghiên cứu đề xuất trong Test Case A3

**TEST CASE A3**

| Evaluate / Models | UDR | ACC | TPR | FPR | F1 | AUC | MCC | PPV |
|---|---|---|---|---|---|---|---|---|
| **Decision Tree** | 0.2721 | 0.937 | 0.9991 | 0.1271 | 0.9366 | 0.936 | 0.8806 | 0.9446 |
| **Random Forest** | 0.9886 | 0.968 | 1 | 1 | 0.968 | 0.9679 | 0.9361 | 0.9682 |
| **LightGBM** | 0.2967 | 0.9377 | 0.9987 | 0.1251 | 0.9374 | 0.9367 | 0.8818 | 0.9451 |
| **XGBoost** | 0.4397 | 0.9502 | 0.9996 | 0.1007 | 0.95 | 0.9494 | 0.9047 | 0.9553 |
| **SVM** | 0.7302 | 0.5697 | 0.5184 | 0.3772 | 0.5689 | 0.5706 | 0.142 | 0.5713 |
| **Naïve Bayes** | 0.9948 | 0.9352 | 0.8885 | 0.0165 | 0.9351 | 0.936 | 0.8748 | 0.9387 |
| **AdaBoost** | 0.9592 | 0.9901 | 0.9959 | 0.0157 | 0.9901 | 0.99 | 0.9804 | 0.9903 |
| **KNN** | 0.8412 | 0.983 | 0.998 | 0.0324 | 0.983 | 0.9828 | 0.9665 | 0.9837 |
| **CNN** | 0.9803 | 0.7833 | 0.9955 | 0.0104 | 0.7713 | 0.7797 | 0.6235 | 0.7022 |
| **Ours (Binary Classification)** | **0.9820** | **0.9478** | **0.9494** | **0.0965** | **0.9478** | **0.9494** | **0.8999** | **0.9506** |
| **Ours (Multi-Classification)** | **0.7624** | **0.9340** | **0.9156** | **0.0965** | **0.8811** | **N/A** | **0.8927** | **0.8624** |

Bảng 8. Kết quả so sánh giữa 9 mô hình khác nhau với mô hình do nhóm nghiên cứu đề xuất trong Test Case A4

**TEST CASE A4**

| Evaluate / Models | UDR | | ACC | TPR | FPR | F1 | AUC | MCC | PPV |
|---|---|---|---|---|---|---|---|---|---|
| | ICMP | UDP | | | | | | | |
| **Decision Tree** | 0.508 | 0.079 | 0.906 | 0.9985 | 0.1895 | 0.9049 | 0.9045 | 0.8257 | 0.9214 |
| **Random Forest** | 0 | 0.729 | 0.942 | 1 | 1 | 0.9413 | 0.9407 | 0.8878 | 0.9471 |
| **LightGBM** | 0.968 | 0.026 | 0.901 | 0.9957 | 0.1972 | 0.8996 | 0.8993 | 0.8159 | 0.9168 |
| **XGBoost** | 0.017 | 0.084 | 0.9 | 0.997 | 0.2 | 0.8988 | 0.8985 | 0.8151 | 0.9168 |
| **SVM** | 0.976 | 0.998 | 0.675 | 0.6435 | 0.2934 | 0.6744 | 0.6751 | 0.3507 | 0.6756 |
| **KNN** | 0.021 | 0.582 | 0.946 | 0.9983 | 0.1081 | 0.9457 | 0.9451 | 0.8967 | 0.9516 |
| **Navie Bayes** | 0.693 | 0.763 | 0.926 | 0.916 | 0.0638 | 0.926 | 0.9261 | 0.8522 | 0.926 |
| **Ada Boost** | 0.004 | 0.008 | 0.897 | 0.9977 | 0.2072 | 0.8955 | 0.8952 | 0.8098 | 0.9148 |
| **CNN** | 0.963 | 0.549 | 0.941 | 0.9946 | 0.1143 | 0.9407 | 0.9401 | 0.8868 | 0.8997 |
| **Ours (Binary Classification)** | **1** | **0.982** | **0.9481** | **0.9497** | **0.0970** | **0.9481** | **0.9497** | **0.9007** | **0.9510** |
| **Ours (Multi-Classification)** | **1** | **0.982** | **0.9480** | **0.9633** | **0.0970** | **0.9142** | **N/A** | **0.9169** | **0.8873** |

Bảng 9. Kết quả so sánh giữa 9 mô hình khác nhau với mô hình do nhóm nghiên cứu đề xuất trong Test Case A5

**TEST CASE A5**

| Evaluate / Models | UDR | | ACC | TPR | FPR | F1 | AUC | MCC | PPV |
|---|---|---|---|---|---|---|---|---|---|
| | ICMP | UDP | | | | | | | |
| **Decision Tree** | 0.240 | 0.031 | 0.545 | 1 | 0.9256 | 0.4145 | 0.5372 | 0.1981 | 0.7637 |
| **Random Forest** | 0.262 | 0.028 | 0.545 | 1 | 1 | 0.4156 | 0.5377 | 0.1994 | 0.7637 |
| **LightGBM** | 0.257 | 0.027 | 0.544 | 1 | 0.9264 | 0.4138 | 0.5368 | 0.1969 | 0.7636 |
| **XGBoost** | 0.258 | 0.029 | 0.545 | 1 | 0.9245 | 0.4157 | 0.5378 | 0.1996 | 0.7638 |
| **SVM** | 0.265 | 0.034 | 0.546 | 0.9963 | 0.9189 | 0.42 | 0.5387 | 0.1934 | 0.7415 |
| **KNN** | 0.255 | 0.024 | 0.543 | 1 | 0.9292 | 0.411 | 0.5354 | 0.1931 | 0.7631 |
| **Navie Bayes** | 0.255 | 0.028 | 0.545 | 1 | 0.9257 | 0.4144 | 0.5371 | 0.1979 | 0.7636 |
| **Ada Boost** | 0.256 | 0.029 | 0.545 | 1 | 0.9248 | 0.4153 | 0.5376 | 0.1992 | 0.7638 |
| **CNN** | 0.215 | 0.024 | 0.539 | 1 | 0.937 | 0.4031 | 0.5315 | 0.1817 | 0.7637 |
| **Ours (Binary Classification)** | **0.982** | **0.9113** | **0.9156** | **0.9160** | **0.0954** | **0.9156** | **0.9160** | **0.8316** | **0.9156** |
| **Ours (Multi-Classification)** | **0.9786** | **0.9113** | **0.9154** | **0.9418** | **0.0954** | **0.9402** | **N/A** | **0.8420** | **0.9390** |

**Bảng 10. Kết quả so sánh giữa 9 mô hình khác nhau với mô hình do nhóm nghiên cứu đề xuất trong Test Case A6**

**TEST CASE A6**

| Evaluate / Models | UDR | | ACC | TPR | FPR | F1 | AUC | MCC | PPV |
|---|---|---|---|---|---|---|---|---|---|
| | ICMP | UDP | | | | | | | |
| **Decision Tree** | 0 | 0.182 | 0.641 | 1 | 0.729 | 0.5828 | 0.6355 | 0.3986 | 0.7931 |
| **Random Forest** | 0 | 0 | 0.567 | 1 | 0.8809 | 0.457 | 0.5596 | 0.2536 | 0.7699 |
| **LightGBM** | 0 | 0.004 | 0.569 | 1 | 0.8757 | 0.4617 | 0.5621 | 0.2593 | 0.7705 |
| **XGBoost** | 0 | 0.004 | 0.569 | 1 | 0.876 | 0.4614 | 0.562 | 0.259 | 0.7706 |
| **SVM** | 0.017 | 0.344 | 0.645 | 0.9586 | 0.6798 | 0.6013 | 0.6394 | 0.3639 | 0.7375 |
| **KNN** | 0.006 | 0.286 | 0.683 | 0.998 | 0.6431 | 0.6435 | 0.6775 | 0.4654 | 0.8051 |
| **Navie Bayes** | 0.648 | 0.022 | 0.576 | 0.9907 | 0.852 | 0.4797 | 0.5693 | 0.2592 | 0.7422 |
| **Ada Boost** | 0.485 | 0.012 | 0.576 | 1 | 0.8627 | 0.4735 | 0.5686 | 0.2734 | 0.7723 |
| **CNN** | 0.002 | 0.436 | 0.742 | 0.9993 | 0.5232 | 0.7213 | 0.738 | 0.5612 | 0.6632 |
| **Ours (Binary Classification)** | **1** | **0.9113** | **0.9168** | **0.9172** | **0.0954** | **0.9167** | **0.9172** | **0.8339** | **0.9167** |
| **Ours (Multi-Classification)** | **0.7696** | **0.9113** | **0.9057** | **0.9219** | **0.0954** | **0.9065** | **N/A** | **0.8306** | **0.8931** |

❖ **Nhận xét kết quả:**

Nhìn chung kết quả của nhóm có độ chính xác cao hơn so với các mô hình ML / DL cơ bản. Kết quả độ chính xác đạt được đều hơn 90%. Các chỉ số khác như MCC, PPV, TPR đều cho kết quả khả thi hơn so với baseline. Đặc biệt là khả năng nhận biết các UDR (Unknow Detection Rate) cũng đạt kết quả đáng mong đợi. Các Test Case cho kết quả cao đối với việc phân loại nhị phần. Trong khi đó, đối với phân loại đa lớp các Test Case cũng cho khá tốt cũng chỉ 76% - 100%. Qua đó, ta có thể thấy được hiệu quả của mô hình mà nhóm đề xuất đối với cùng một tập dữ liệu.

# 5. KẾT LUẬN – ĐỀ NGHỊ

## 5.1. Kết luận

Trong nghiên cứu lần này, nhóm thực hiện đã nghiên cứu và triển khai thành công một mô hình NIDS có khả năng phát hiện các loại xâm nhập mạng liên quan tới các loại scanning attack đã biết cũng như chưa biết với độ chính xác cao và thời gian huấn luyện ngắn bằng việc tận dụng, kết hợp ưu điểm của các mô hình ML. Hơn nữa, mô hình đề xuất của chúng tôi còn thể hiện tính linh hoạt và khả năng mở rộng, nâng cấp dễ dàng, giúp hệ thống luôn được cập nhật và hiệu quả trước các mối đe dọa an ninh mạng ngày càng phức tạp.

## 5.2.Ý nghĩa khoa học

Nghiên cứu này đóng góp quan trọng vào lĩnh vực an ninh mạng, đặc biệt là trong việc phát hiện và phòng chống các cuộc tấn công mạng. Đặt nền móng cho các nghiên cứu tương tự, nhất là khi các cuộc tấn công ngày càng tinh vi đặc biệt là các cuộc tấn công APT, zero-day attack. Việc ứng dụng các mô hình Machine Learning tiên tiến giúp nâng cao hiệu quả phát hiện các xâm nhập mạng với độ chính xác cao, kể cả đối với những mối đe dọa mới chưa từng được biết đến. Đây là một bước tiến quan trọng trong việc bảo vệ hệ thống mạng khỏi các cuộc tấn công phức tạp và không ngừng biến đổi.

Nghiên cứu này còn có thể được áp dụng để phát hiện các loại tấn công mạng khác như DDOS, brute-force,… và nhiều loại tấn công mạng khác. Khả năng ứng dụng rộng rãi của mô hình làm cho nó trở thành một công cụ hữu ích trong việc nâng cao an ninh mạng tổng thể, giúp bảo vệ hệ thống trước các mối đe dọa ngày càng gia tăng và phức tạp.

## 5.3.Hiệu quả về mặt kinh tế - xã hội

Việc phát hiện và ngăn chặn Scanning attack đóng vai trò rất quan trọng trong các hệ thống mạng. Nó đóng vai trò to lớn trong việc thành công hay thất bại của một tấn công. Chính vì vậy, nghiên cứu của nhóm tập trung vào loại tấn công này nhằm bảo vệ tài sản số của mình một cách hiệu quả hơn. Điều này giúp giảm thiểu thiệt hại kinh tế do các cuộc tấn công mạng gây ra, từ đó tăng cường sự ổn định và tin cậy của hệ thống thông tin.

Ngoài ra, việc dễ dàng mở rộng và nâng cấp mô hình NIDS giúp các doanh nghiệp có thể liên tục cập nhật và cải thiện hệ thống an ninh mạng của doanh nghiệp mà không phải đầu tư quá nhiều vào các giải pháp mới. Điều này giúp tối ưu hóa chi phí đầu tư và duy trì, đồng thời đảm bảo rằng hệ thống an ninh mạng luôn ở trạng thái tốt nhất để đối phó với các mối đe dọa ngày càng phức tạp.

Việc xây dựng một hệ thống NIDS cũng đóng vai trò không nhỏ giúp tạo ra một môi trường mạng an toàn hơn, khuyến khích sự phát triển của kinh tế số và các dịch vụ trực tuyến. Nhờ đó, xã hội có thể tận dụng các lợi ích của công nghệ thông tin và

truyền thông một cách an toàn và hiệu quả hơn. Đồng thời còn tạo niềm tin cho khách hàng khi sử dụng các dịch vụ của doanh nghiệp.

## 5.4.Phạm vi áp dụng

Với nghiên cứu này, chúng tôi góp phần tạo thêm tài liệu hữu ích cho các nhà nghiên cứu tìm hiểu và phát triển các hệ thống phát hiện và xâm nhập mạng, các hệ thống này ngoài khả năng nhận biết các cuộc tấn công đã biết mà còn có thể phát hiện ra các loại xâm nhập mới, các loại xâm nhập chưa được biết đến. Giải pháp này có thể được áp dụng và mở rộng quy mô một cách dễ dàng, phù hợp với nhu cầu và điều kiện của từng doanh nghiệp.

Giải pháp này có thể triển khai mạnh mẽ với mọi doanh nghiệp vì khả năng linh hoạt về quy mô của nó. Việc kết hợp các mô hình Random Forest mang lại sự gọn nhẹ, hiệu quả về tài nguyên, đặc biệt phù hợp với các thiết bị IoT có tài nguyên hạn chế.

Điều này mở ra nhiều cơ hội ứng dụng cho các hệ thống phát hiện xâm nhập trên nền tảng IoT, góp phần bảo mật các thiết bị và hệ thống có tài nguyên hạn chế.

Tóm lại, nghiên cứu này thực sự mang lại nhiều tiềm năng ứng dụng, từ việc triển khai linh hoạt, hiệu quả tài nguyên, đến khả năng phát hiện các cuộc tấn công mới. Đây là một công trình nghiên cứu hữu ích, góp phần thúc đẩy sự phát triển của lĩnh vực an ninh mạng.

## 5.5.Hướng phát triển

Kết quả cho thấy mô hình hoạt động tốt và hiệu quả, trong tương lai nhóm nghiên cứu sẽ tiếp tục triển khai đánh giá hiệu của của đề xuất với nhiều tập dữ liệu khác nhau và đồng thời tối ưu hóa, tinh chỉnh các tham số của mô hình để đạt được hiệu quả cao nhất (Sử dụng kỹ thuật như Grid Search hoặc Random Search để tìm các tham số tối ưu).

Bên cạnh đó, nhóm cũng mong muốn có thể triển khai mô hình đề xuất để phát hiện các cuộc tấn công khác ngoài Scanning Attack, điển hình như DDOS Attack,… . Việc mở rộng khả năng phát hiện các loại tấn công khác sẽ giúp hệ thống NIDS trở nên toàn diện và mạnh mẽ hơn trong việc bảo vệ mạng trước các mối đe dọa đa dạng và phức tạp.

# 6. TÀI LIỆU THAM KHẢO VÀ PHỤ LỤC

## 6.1. Tài liệu tham khảo

**[1] Soltani, M., Ousat, B., Siavoshani, M. J., & Jahangir, A. H. (2023). An adaptable deep learning-based intrusion detection system to zero-day attacks. Journal of Information Security and Applications, 76, 103516.**

**[2] Al, S., & Dener, M. (2021). STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment. Computers & Security, 110, 102435.**

**[3] Rao, Y. N., & Suresh Babu, K. (2023). An imbalanced generative adversarial network-based approach for network intrusion detection in an imbalanced dataset. Sensors, 23(1), 550.**

**[4] Oliveira, N., Praça, I., Maia, E., & Sousa, O. (2021). Intelligent cyber attack detection and classification for network-based intrusion detection systems. Applied Sciences, 11(4), 1674.**

**[5] Verkerken, M., D'hooge, L., Wauters, T., Volckaert, B., & De Turck, F. (2022). Towards model generalization for intrusion detection: Unsupervised machine learning techniques. Journal of Network and Systems Management, 30, 1-25.**

**[6] Liao, J., Teo, S. G., Kundu, P. P., & Truong-Huu, T. (2021, July). ENAD: An ensemble framework for unsupervised network anomaly detection. In 2021 IEEE International Conference on Cyber Security and Resilience (CSR) (pp. 81-88). IEEE.**

**[7] Pontes, C. F., De Souza, M. M., Gondim, J. J., Bishop, M., & Marotta, M. A. (2021). A new method for flow-based network intrusion detection using the inverse Potts model. IEEE Transactions on Network and Service Management, 18(2), 1125-1136.**

**[7] Jing, Y., Zhang, Z., Hu, T., Li, Z., & Liu, S. (2022). Sustainable intrusion detection with new attack classification in private clouds. Journal of Networking and Network Applications, 1(4), 150-159.**

**[8] Sarhan, M., Layeghy, S., Gallagher, M., & Portmann, M. (2023). From zero-shot machine learning to zero-day attack detection. International Journal of Information Security, 22(4), 947-959.**

**[9] Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Transactions on Emerging Telecommunications Technologies, 32(1), e4150.**

**[9] Sultana, N., Chilamkurti, N., Peng, W., & Alhadad, R. (2019). Survey on SDN based network intrusion detection system using machine learning approaches. Peer-to-Peer Networking and Applications, 12(2), 493-501.**

**[10]. Fernandes, G., Rodrigues, J. J., Carvalho, L. F., Al-Muhtadi, J. F., & Proença, M. L. (2019). A comprehensive survey on network anomaly detection. *Telecommunication Systems, 70,* 447-489.**

**[11] Ring, M., Wunderlich, S., Grüdl, D., Landes, D., & Hotho, A. (2017, June). Flow-based benchmark data sets for intrusion detection. In *Proceedings of the 16th European conference on cyber warfare and security. ACPI* (pp. 361-369).**

**[12] Ring, M., Wunderlich, S., Grüdl, D., Landes, D., & Hotho, A. (2017). Creation of flow-based data sets for intrusion detection. *Journal of Information Warfare, 16*(4), 41-54.**

## 6.2. Danh mục hình ảnh



## 6.3. Danh mục bảng

## 6.4.Danh mục thuật ngữ viết tắt

| Số thứ tự | Thuật ngữ | Giải thích thuật ngữ |
|---|---|---|
| 1 | IDS | Intrusion Detection Systems |
| 2 | NIDS | Network-based IDS |
| 3 | HIDS | Host-based IDS |
| 4 | ML | Machine Learning |
| 5 | DL | Deep Learning |
| 6 | UDR | Unknow Detection Rate |
| 7 | CIDDS | Coburg Intrusion Detection Data Sets |
| 8 | AIDS | Anomaly-based Intrusion Detection Systems |
| 9 | SIDS | Signature-based Intrusion Detection Systems |
| 10 | APT | Advanced persistent threat |
| 11 | DDOS | Distributed denial-of-Service |
| 12 | IoT | Internet of Thing |
| 13 | LOF | Local Outlier Factor |
| 14 | iNNE | Isolation using Nearest Neighbor Ensemble |
| 15 | MCC | Matthews correlation coefficient |