BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO THỰC TẬP TỐT NGHIỆP

TÌM HIỂU VÀ PHÁT TRIỂN HỆ THỐNG PHÁT HIỆN LỖI TỰ ĐỘNG TRONG HỆ THỐNG TRUYỀN THÔNG LỚN

Giảng viên hướng dẫn: ThS. Nguyễn Chuẩn Nam
Sinh viên thực hiện: Tấn Lai Hoàng
MSSV: 22DH111164
Chuyên ngành: An ninh mạng
Đơn vị thực tập: CÔNG TY TNHH GIÁO DỤC STEM VÀ PHÁT TRIỂN KỸ NĂNG 4C
Khóa: 2022 – 2026

TP.HCM, tháng 12 năm 2025

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO THỰC TẬP TỐT NGHIỆP

TÌM HIỂU VÀ PHÁT TRIỂN HỆ THỐNG PHÁT HIỆN LỖI TỰ ĐỘNG TRONG HỆ THỐNG TRUYỀN THÔNG LỚN

Giảng viên hướng dẫn: ThS. Nguyễn Chuẩn Nam
Sinh viên thực hiện: Tấn Lai Hoàng
MSSV: 22DH111164
Chuyên ngành: An ninh mạng
Đơn vị thực tập: CÔNG TY TNHH GIÁO DỤC STEM VÀ PHÁT TRIỂN KỸ NĂNG 4C
Khóa: 2022 – 2026

TP.HCM, tháng 12 năm 2025

LỜI CẢM ƠN

Lời đầu tiên, em xin được gửi lời cảm ơn chân thành và sâu sắc nhất tới các anh/chị/em/bạn đồng nghiệp trong CÔNG TY TNHH GIÁO DỤC STEM VÀ PHÁT TRIỂN KỸ NĂNG 4C. Trong suốt thời gian thực tập, em đã nhận được sự hướng dẫn, chỉ bảo tận tình cùng những chia sẻ quý báu về chuyên môn và kinh nghiệm thực tế. Nhờ sự giúp đỡ nhiệt tình của mọi người, em đã có cơ hội học hỏi, rèn luyện và trưởng thành hơn rất nhiều, không chỉ trong kỹ năng nghề nghiệp mà còn trong tác phong làm việc chuyên nghiệp.

Tiếp theo, em xin trân trọng gửi lời cảm ơn tới GVHD – ThS. Nguyễn Chuẩn Nam – người đã luôn theo sát, tận tình chỉ dẫn và định hướng cho em trong suốt quá trình thực hiện đề tài nghiên cứu. Những góp ý, nhận xét và sự hỗ trợ quý báu của thầy đã giúp em hoàn thiện hơn về kiến thức chuyên ngành cũng như kỹ năng nghiên cứu khoa học.

Cuối cùng, em xin gửi lời cảm ơn đến các bạn học cùng khóa – những người đã đồng hành, chia sẻ tài liệu, kinh nghiệm và luôn động viên, hỗ trợ em trong suốt quá trình học tập và thực hiện đề tài. Sự giúp đỡ và tinh thần đoàn kết của các bạn là nguồn động lực lớn để em hoàn thành tốt công việc của mình.

LỜI MỞ ĐẦU

Trung tâm Giáo dục STEAMZone là một đơn vị tiên phong trong lĩnh vực giáo dục STEM/STEAM tại Việt Nam, hoạt động với mục tiêu cốt lõi là áp dụng các phương thức giáo dục hiện đại vào quá trình dạy học và đánh giá. Công ty tập trung vào việc giúp học sinh củng cố, hình thành kiến thức mới và ứng dụng kiến thức vào thực tiễn cuộc sống. STEAMZone đặc biệt chú trọng phát triển các kỹ năng thiết yếu của thế kỷ 21, được gọi là 5C, bao gồm: Giao tiếp (Communication), Cộng tác (Collaboration), Tư duy phản biện và giải quyết vấn đề (Critical Thinking & Problem Solving), Sáng tạo (Creativity), và Tư duy máy tính (Computational Thinking), nhằm chuẩn bị cho thế hệ công dân số làm chủ nền kinh tế và xã hội số.

STEAMZone theo đuổi Sứ mệnh cao cả là đảm bảo sự công bằng trong giáo dục STEM/STEAM, cam kết không để bất kỳ học sinh nào bị bỏ lại phía sau thông qua việc cung cấp giáo viên giỏi, nội dung chương trình chất lượng và cơ sở vật chất hiện đại. Tầm nhìn của công ty là góp phần xây dựng lực lượng lao động chất lượng cao cho đất nước. Thế hệ trẻ được đào tạo tại đây sẽ trở thành nguồn nhân lực nòng cốt, thúc đẩy sự tiến bộ toàn cầu trong kỷ nguyên Cách mạng công nghiệp 4.0, Chuyển đổi số và Chuyển đổi xanh. Để đạt được điều này, công ty ứng dụng các phương pháp sư phạm tiên tiến như Học dự án (PBL), Học truy vấn (Inquiry-Based Learning), Học đảo ngược (Flipped Learning), và chú trọng xây dựng môi trường Vui + Học + Sáng tạo.

Trung tâm Giáo dục STEAMZone được trang bị cơ sở vật chất hiện đại, thiết kế theo không gian mở đạt chuẩn quốc tế, nhằm tăng cường tính sáng tạo và linh hoạt trong các hoạt động trải nghiệm. Các thiết bị và công nghệ được ứng dụng gắn liền với Cuộc cách mạng công nghiệp lần thứ 4, bao gồm Trí tuệ nhân tạo (AI), Tự động hóa (Robotics), Kết nối vạn vật (AIoT), Khoa học dữ liệu, In ấn 3D, và Thực tế ảo (VR/AR). Công ty cung cấp đa dạng chương trình đào tạo cho học sinh từ cấp Mầm non đến Trung học Phổ thông, tổ chức các trại hè STEAM trong nước và quốc tế, cùng với các khóa đào tạo chuyên sâu cho giáo viên. Ngoài ra, STEAMZone còn hợp tác với nhiều đối tác quốc tế uy tín như STEM.org (Hoa Kỳ) và các trường đại học lớn để không ngừng nghiên cứu và đổi mới chất lượng giáo dục.

NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

TPHCM, ngày …… tháng …… năm 20……

Giảng viên hướng dẫn

(Ký tên, đóng dấu)

NHẬN XÉT CỦA CƠ QUAN THỰC TẬP

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

TPHCM, ngày …… tháng …… năm 20……

Xác nhận của đơn vị thực tập

(Ký tên, đóng dấu)

MỤC LỤC

DANH MỤC CÁC BẢNG BIỂU

DANH MỤC CÁC BẢNG HÌNH

DANH MỤC CÁC TỪ VIẾT TẮT

Từ viết tắt	Ý nghĩa
AI	Artificial Intelligence (Trí tuệ nhân tạo)
API	Application Programming Interface (Giao diện lập trình ứng dụng)
ARP	Address Resolution Protocol (Giao thức phân giải địa chỉ)
C&C	Command and Control (Máy chủ điều khiển và ra lệnh - thường dùng bởi mã độc)
CLI	Command Line Interface (Giao diện dòng lệnh)
CPU	Central Processing Unit (Bộ xử lý trung tâm)
CSV	Comma-Separated Values (Định dạng tệp văn bản lưu trữ dữ liệu bảng)
DDoS	Distributed Denial of Service (Tấn công từ chối dịch vụ phân tán)
DNS	Domain Name System (Hệ thống phân giải tên miền)
DoS	Denial of Service (Tấn công từ chối dịch vụ)
GUI	Graphical User Interface (Giao diện người dùng đồ họa)
HIDS	Host-based Intrusion Detection System (Hệ thống phát hiện xâm nhập tại máy chủ/máy trạm)
HTTP	Hypertext Transfer Protocol (Giao thức truyền tải siêu văn bản)
ICMP	Internet Control Message Protocol (Giao thức thông điệp điều khiển Internet)
IDE	Integrated Development Environment (Môi trường phát triển tích hợp)
IDS	Intrusion Detection System (Hệ thống phát hiện xâm nhập)
IP	Internet Protocol (Giao thức Internet)
IPS	Intrusion Prevention System (Hệ thống ngăn chặn xâm nhập)
IT	Information Technology (Công nghệ thông tin)
LAN	Local Area Network (Mạng cục bộ)
MAC	Media Access Control (Địa chỉ vật lý của thiết bị mạng)
ML	Machine Learning (Học máy)

MTU	Maximum Transmission Unit (Đơn vị truyền tải tối đa)
NAT	Network Address Translation (Biên dịch địa chỉ mạng)
NIC	Network Interface Card (Card giao tiếp mạng)
NIDS	Network-based Intrusion Detection System (Hệ thống phát hiện xâm nhập mạng)
OSI	Open Systems Interconnection (Mô hình tham chiếu kết nối các hệ thống mở)
PCAP	Packet Capture (Định dạng tệp lưu trữ dữ liệu gói tin mạng)
PDF	Portable Document Format (Định dạng tài liệu di động)
RAM	Random Access Memory (Bộ nhớ truy cập ngẫu nhiên)
RFC	Request For Comments (Tài liệu chứa các đặc tả kỹ thuật và tiêu chuẩn Internet)
TCP	Transmission Control Protocol (Giao thức điều khiển truyền vận)
UDP	User Datagram Protocol (Giao thức gói dữ liệu người dùng)
UI/UX	User Interface / User Experience (Giao diện người dùng / Trải nghiệm người dùng)
VM	Virtual Machine (Máy ảo)
WPA2	Wi-Fi Protected Access 2 (Chuẩn bảo mật Wi-Fi)

TRƯỜNG ĐẠI HỌC

NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN

SỔ NHẬT KÝ

THỰC TẬP

Họ tên HSSV: Tấn Lai Hoàng

Lớp: AN2202

Ngành: An ninh mạng

Cơ quan thực tập: CÔNG TY TNHH GIÁO DỤC STEM VÀ PHÁT TRIỂN KỸ NĂNG 4C

Thời gian thực tập: **Từ 29/08/2025 đến 01/11/2025**

Tháng 12/2025

*** Lưu ý:** Nội dung thực tập được liệt kê và đánh giá theo từng tuần trong đợt thực tập

Tuần	**Thời gian**	**Nội dung thực tập**	**Nhận xét của đơn vị thực tập**	**Chữ ký người HD của đơn vị thực tập**
1	Từ 29/8/2025 đến 30/8/2025	- Gặp gỡ nhân viên và làm quen với môi trường công ty. - Trao đổi và xác nhận vị trí thực tập với ban nhân sự.		
2	Từ 3/9/2025 đến 6/9/2025	- Nghe giới thiệu về thiết bị và nội dung thực tập. - Hỗ trợ công việc ngoài lề.		
3	Từ 8/9/2025 đến 13/9/2025	- Tham gia hỗ trợ giảng dạy IC3 và STEM tại trường Tiểu học Trương Quyền		
4	Từ 15/9/2025 đến 20/9/2025	- Tham gia hỗ trợ giảng dạy IC3 và STEM tại trường tiểu học Trương Quyền - Dạy lập trình bằng HTML và CSS cho học viên tại công ty		
5	Từ 22/9/2025 đến 27/9/2025	- Tham gia hỗ trợ giảng dạy IC3 và STEM tại trường tiểu học Trương Quyền - Tham gia hỗ trợ giảng dạy AI tại trường THCS Colette - Dạy lập trình bằng HTML và CSS cho học viên tại công ty		

6	Từ 29/9/2025 đến 4/10/2025	- Tham gia hỗ trợ giảng dạy IC3 và STEM tại trường tiểu học Trương Quyền - Tham gia hỗ trợ giảng dạy AI tại trường THCS Colette - Tham gia tiệc Tết Trung Thu tại công ty		
7	Từ 6/10/2025 đến 11/10/2025	- Tham gia hỗ trợ giảng dạy IC3 và STEM tại trường tiểu học Trương Quyền - Tham gia hỗ trợ giảng dạy AI tại trường THCS Colette - Trưng bày sản phẩm của công ty tại Đại hội đại biểu Đảng bộ TPHCM - Học viện Cán bộ TPHCM - Đăng bài giảng tương tác lên website Lumio		
8	Từ 13/10/2025 đến 18/10/2025	- Tham gia hỗ trợ giảng dạy IC3 và STEM tại trường tiểu học Trương Quyền - Đăng bài giảng tương tác lên website Lumio - Đăng bài giảng tương tác lên website Lumio - Hỗ trợ công việc ngoài lề		
9	Từ 20/10/2025 đến 25/10/2025	- Tham gia hỗ trợ giảng dạy IC3 và STEM tại trường tiểu học Trương Quyền - Trưng bày sản phẩm tại công ty để giới thiệu với khách		

10	Từ 27/10/2025 đến 01/11/2025	- Tham gia hỗ trợ giảng dạy AI tại trường THCS Colette - Tham gia hỗ trợ giảng dạy IC3 và STEM tại trường tiểu học Trương Quyền - Lắp đặt trang thiết bị cho cuộc thi lái drone tại trường THCS Đoàn Thị Điểm		

TP.HCM, ngày …… tháng …… năm 20……

Trưởng đơn vị quản lý thực tập

TRƯỜNG ĐẠI HỌC
NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH

Địa chỉ: 828 Sư Vạn Hạnh, Phường Hòa Hưng, Quận 10, TP. Hồ Chí Minh

Điện thoại: (+84 28) 38 632 052 – 38 629 232

Fax: (+84 28) 38 650 991

Email: vpkhoacntt@huflit.edu.vn

PHIẾU NHẬN XÉT SINH VIÊN THỰC TẬP

(Vui lòng đánh dấu [x] vào ô thích hợp)

Họ và tên sinh viên: Tấn Lai Hoàng **Lớp**: AN2202

Cơ quan tiếp nhận: CÔNG TY TNHH GIÁO DỤC STEM VÀ PHÁT TRIỂN KỸ NĂNG 4C

1. Nhận xét của cơ quan về chất lượng công việc được giao

Các công việc được giao:

☐ Hoàn thành xuất sắc ☐ Khá ☐ Yếu

☐ Tốt ☐ Trung bình

Hoàn tất công việc được giao:

☐ Hoàn thành đúng ☐ Thỉnh thoảng đúng ☐ Không đúng thời hạn

Tính hữu hiệu của đợt thực tập đối với cơ quan:

☐ Có giúp ích nhiều ☐ Giúp ích ít

☐ Không giúp ích gì mấy cho hoạt động của cơ quan

2. Nhận xét của cơ quan về bản thân sinh viên

Năng lực chuyên môn sử dụng vào công việc được giao ở mức:

☐ Giỏi ☐ Khá ☐ Trung bình ☐ Yếu

Tinh thần, thái độ đối với công việc được giao:

☐ Tích cực ☐ Bình thường ☐ Thiếu tích cực

Đảm bảo kỷ luật lao động (giờ giấc lao động, nghỉ làm, ...)

☐ Tốt ☐ Trung bình ☐ Kém

Thái độ đối với cán bộ, công nhân viên trong cơ quan:

☐ Chan hòa ☐ Không có gì đáng nói ☐ Rụt rè

3. **Nếu được, xin cho biết một "thành tích nổi bật" của sinh viên** (nếu không có, xin bỏ qua)

.....

.....

4. **Các nhận xét khác** (nếu có)

.....

.....

5. **Đánh giá (theo thang điểm 10)**

a) Điểm chuyên cần, phong cách: ..... b) Điểm chuyên môn: .....

Vui lòng xin cho biết thêm:

- Họ và tên người nhận xét: .....
- Chức vụ trong cơ quan: .....

Trưởng đơn vị	Ngày …… tháng …… năm 2025
(ký tên, đóng dấu và ghi rõ họ tên)	Người nhận xét
	(Ký và ghi rõ họ tên)

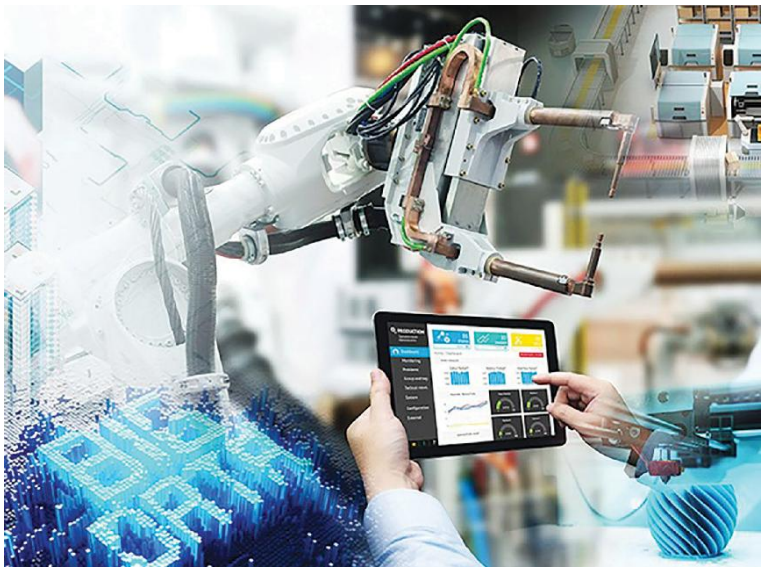
Ghi chú: Một phiếu nhận xét chỉ đánh giá cho một sinh viên

Sinh viên nộp phiếu nhận xét đi kèm báo cáo thực tập

CHƯƠNG I. GIỚI THIỆU

1. Bối cảnh và lý do chọn đề tài

Trong thời đại công nghệ số, hệ thống truyền thông giữ vai trò nền tảng trong việc kết nối và trao đổi thông tin giữa cá nhân, tổ chức và doanh nghiệp. Bên cạnh chức năng liên lạc, các hệ thống này còn đảm bảo hoạt động ổn định cho nhiều dịch vụ trọng yếu như thương mại điện tử, ngân hàng, giáo dục trực tuyến và y tế từ xa. Tuy nhiên, khi quy mô và độ phức tạp ngày càng tăng, việc duy trì tính ổn định, an toàn và chất lượng dịch vụ trở thành thách thức lớn.



Hình 1. Tự động hóa và sản xuất thông minh

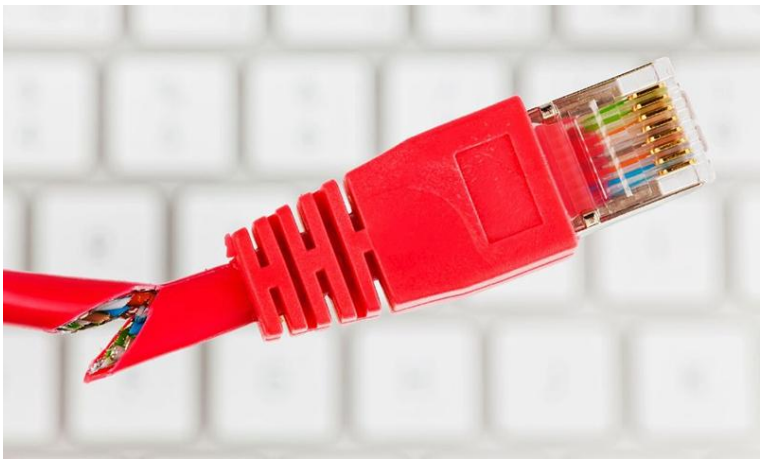
Các lỗi trong hệ thống có thể phát sinh do phần cứng, phần mềm, tấn công mạng, quá tải hoặc sai sót cấu hình, gây gián đoạn dịch vụ và thiệt hại nghiêm trọng nếu không được xử lý kịp thời. Phương pháp giám sát truyền thống khó đáp ứng yêu cầu của môi trường hệ thống lớn và dữ liệu biến động liên tục. Do đó, việc ứng dụng trí tuệ nhân tạo (AI) và học máy (ML) trong phát hiện lỗi tự động là hướng đi tất yếu, giúp phân tích dữ liệu theo thời gian thực, nhận diện bất thường và cảnh báo sớm.

Từ thực tế đó, đề tài **"Tìm hiểu và phát triển Hệ thống phát hiện lỗi tự động trong hệ thống truyền thông lớn"** được lựa chọn nhằm nghiên cứu, xây dựng giải pháp nâng cao khả năng giám sát, phát hiện và xử lý lỗi, góp phần đảm bảo độ tin cậy và an toàn cho hệ thống truyền thông hiện đại.

2. Vấn đề đặt ra trong phát hiện lỗi hệ thống truyền thông

Hệ thống truyền thông lớn có cấu trúc phức tạp, gồm nhiều thành phần phần cứng, phần mềm, giao thức và thiết bị mạng. Trong quá trình vận hành, việc phát hiện lỗi gặp nhiều khó khăn do:

- **Nguyên nhân lỗi đa dạng:** Lỗi có thể xuất phát từ phần cứng, phần mềm, cấu hình mạng hoặc các yếu tố bên ngoài như tấn công mạng, khiến việc xác định và phân loại trở nên phức tạp.
- **Khối lượng dữ liệu khổng lồ:** Hệ thống phải xử lý dữ liệu theo thời gian thực, khiến các phương pháp giám sát thủ công khó theo kịp, dễ bỏ sót hoặc phát hiện chậm.
- **Tính động và phức tạp cao:** Sự mở rộng quy mô, nâng cấp hạ tầng hay biến động lưu lượng làm phát sinh lỗi mới, vượt ngoài khả năng dự đoán của các công cụ truyền thống.
- **Hạn chế của phương pháp cũ:** Các giải pháp thủ công hoặc dựa trên ngưỡng cảnh báo tĩnh chỉ phát hiện lỗi rõ ràng, thiếu khả năng nhận diện bất thường tinh vi và phụ thuộc nhiều vào kinh nghiệm con người.
- **Yêu cầu kịp thời và tự động:** Việc phát hiện chậm trễ có thể gây hậu quả nghiêm trọng, đặc biệt với các hệ thống quan trọng, nên cần một giải pháp tự động và chính xác hơn.



Hình 2. Sự cố vật lý đường truyền Internet

Từ đó, việc ứng dụng **AI và Machine Learning** trong phát hiện lỗi tự động là hướng đi cần thiết, giúp khắc phục hạn chế của phương pháp truyền thống, tăng hiệu quả giám sát và đảm bảo tính ổn định cho hệ thống truyền thông quy mô lớn.

3. Mục tiêu nghiên cứu

Đề tài hướng đến việc nghiên cứu và phát triển **hệ thống phát hiện lỗi tự động trong hệ thống truyền thông lớn** với các mục tiêu cụ thể sau:

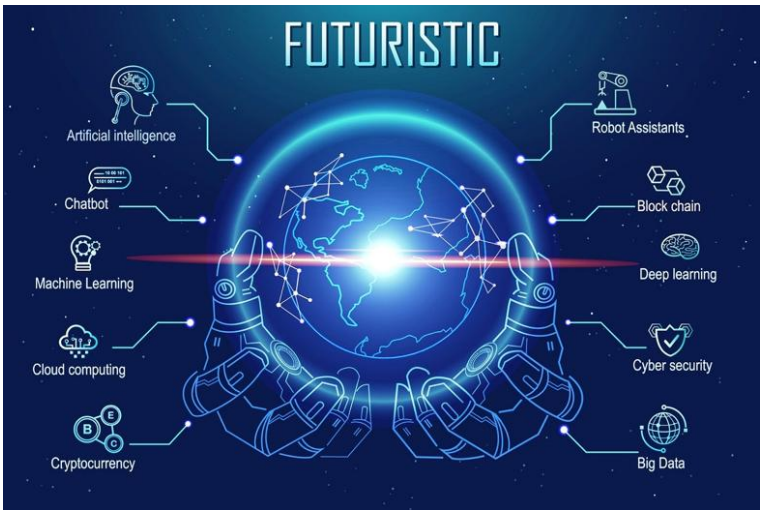
- **Xây dựng cơ sở lý thuyết:** Tìm hiểu tổng quan về hệ thống truyền thông lớn, đặc điểm vận hành, các loại lỗi thường gặp và thách thức trong việc phát hiện, xử lý lỗi.
- **Phân tích phương pháp hiện có:** Đánh giá các phương pháp phát hiện lỗi truyền thống, nêu rõ ưu điểm và hạn chế khi áp dụng trong môi trường quy mô lớn.
- **Ứng dụng công nghệ mới:** Nghiên cứu khả năng sử dụng trí tuệ nhân tạo (AI) và học máy (ML) để phát hiện bất thường, phân loại lỗi và dự đoán sự cố tiềm ẩn.
- **Đề xuất mô hình hệ thống:** Thiết kế quy trình gồm thu thập, tiền xử lý dữ liệu, huấn luyện mô hình, phát hiện lỗi và tích hợp cảnh báo tự động.

- **Thử nghiệm và đánh giá:** Xây dựng môi trường thử nghiệm, kiểm chứng mô hình trên dữ liệu thực tế hoặc mô phỏng, đánh giá độ chính xác và hiệu quả so với phương pháp truyền thống.
- **Định hướng phát triển:** Đề xuất giải pháp mở rộng và ứng dụng thực tiễn, nhằm nâng cao độ tin cậy, khả năng thích ứng và tự động hóa cho hệ thống truyền thông.

Hình 3. Ứng dụng AI trong quản lý hệ thống truyền thông lớn

4. Đối tượng và phạm vi nghiên cứu

- **Đối tượng nghiên cứu**:
    - Các hệ thống truyền thông quy mô lớn gồm hạ tầng mạng máy tính, máy chủ, thiết bị định tuyến, chuyển mạch và các nền tảng dịch vụ trực tuyến.
    - Các loại lỗi phổ biến như lỗi phần cứng, phần mềm, cấu hình, quá tải và các bất thường liên quan đến an ninh mạng.
    - Các phương pháp và công cụ phát hiện lỗi, đặc biệt là những giải pháp ứng dụng trí tuệ nhân tạo (AI) và học máy (Machine Learning – ML).



Hình 4. AI và ML là một trong những nhân tố phát triển công nghệ tương lai

- **Phạm vi nghiên cứu**:
    - Tập trung vào phát hiện và cảnh báo lỗi tự động trong hệ thống truyền thông lớn, chưa đi sâu vào khắc phục sự cố hay tối ưu hiệu suất mạng.
    - Nghiên cứu các kỹ thuật phát hiện bất thường, phân loại và dự đoán lỗi dựa trên dữ liệu log, dữ liệu giám sát và dữ liệu vận hành.
    - Thử nghiệm mô hình trong môi trường giả lập hoặc với dữ liệu mô phỏng để đánh giá tính khả thi và hiệu quả trước khi ứng dụng thực tế.
    - Giới hạn trong phạm vi nghiên cứu học thuật, chưa mở rộng đến toàn bộ các hệ thống truyền thông chuyên biệt như viễn thông quốc tế hoặc mạng 5G quy mô toàn cầu.

5. Phương pháp nghiên cứu

Đề tài kết hợp nhiều phương pháp để đảm bảo tính khoa học và thực tiễn: nghiên cứu tài liệu, phân tích-so sánh, mô hình hóa, thử nghiệm và đánh giá. Mục tiêu là lựa chọn, xây dựng và kiểm chứng một mô hình phát hiện lỗi tự động phù hợp cho hệ thống truyền thông lớn.

- **Nghiên cứu tài liệu:** Tổng hợp công trình, bài báo và tài liệu chuyên ngành về lỗi hệ thống và ứng dụng AI/ML.

- **Phân tích – so sánh:** Đánh giá ưu – nhược điểm của các phương pháp hiện có để chọn hướng tiếp cận tối ưu.
- **Mô hình hóa:** Thiết kế quy trình hệ thống gồm thu thập dữ liệu, tiền xử lý, huấn luyện mô hình ML, triển khai phát hiện và cảnh báo.
- **Thử nghiệm:** Cài đặt/mô phỏng và chạy mô hình trên dữ liệu thực tế hoặc giả lập để kiểm chứng tính khả thi.
- **Đánh giá:** So sánh theo các tiêu chí (độ chính xác, độ bao phủ, thời gian phát hiện, khả năng thích ứng) so với phương pháp truyền thống.

Các phương pháp nghiên cứu trên được kết hợp linh hoạt, vừa đảm bảo tính khoa học, vừa đảm bảo tính thực tiễn, giúp đề tài đạt được mục tiêu đề ra một cách toàn diện.

Hình 5. Nghiên cứu AI/ML

6. Kết cấu báo cáo

Báo cáo được chia thành **ba chương chính** nhằm đảm bảo tính logic và bao quát toàn diện nội dung nghiên cứu:

- **Chương 1 – Giới thiệu:** Trình bày bối cảnh, lý do chọn đề tài, mục tiêu, đối tượng, phạm vi, phương pháp nghiên cứu và cấu trúc báo cáo. Chương này giúp định hướng và làm rõ ý nghĩa tổng thể của đề tài.

- **Chương 2 – Cơ sở lý thuyết và mô hình hệ thống:** Trình bày kiến thức nền tảng về hệ thống truyền thông lớn, các loại lỗi thường gặp, các phương pháp phát hiện lỗi truyền thống và ứng dụng AI, ML trong phát hiện lỗi. Trên cơ sở đó, chương này đề xuất mô hình hệ thống phát hiện lỗi tự động.
- **Chương 3 – Thử nghiệm, đánh giá và kết luận:** Mô tả môi trường, quy trình thử nghiệm, kết quả và đánh giá hiệu quả của mô hình so với phương pháp truyền thống. Cuối cùng, chương này đưa ra kết luận tổng quát và định hướng phát triển trong tương lai.

Kết cấu trên giúp báo cáo thể hiện rõ ràng tiến trình nghiên cứu, từ cơ sở lý thuyết đến thực nghiệm, qua đó làm nổi bật tính khả thi và giá trị ứng dụng của đề tài.

Hình 6. Trí tuệ nhân tạo và học máy

CHƯƠNG II. CƠ SỞ LÝ THUYẾT VÀ CÔNG NGHỆ NỀN TẢNG

1. Tổng quan về Giám sát và An ninh Mạng

1.1. Các mục tiêu cốt lõi của Giám sát mạng

Việc triển khai giám sát mạng trong các hệ thống lớn nhằm đảm bảo ba trụ cột chính về khả năng vận hành và an ninh. Các mục tiêu kỹ thuật của giám sát mạng bao gồm:

- **Khả năng quan sát (Network Visibility):** Cung cấp cái nhìn toàn cảnh về lưu lượng mạng: ai đang giao tiếp với ai, sử dụng giao thức gì, và băng thông tiêu thụ bao nhiêu. Loại bỏ các "điểm mù" (Blind Spots) trong hạ tầng, nơi mã độc hoặc kẻ tấn công có thể ẩn náu và hoạt động âm thầm.
- **Phát hiện bất thường (Anomaly Detection):** Thiết lập một đường cơ sở (Baseline) cho hoạt động bình thường và nhận diện các sai lệch (Deviation) về lưu lượng hoặc hành vi. Cảnh báo sớm các dấu hiệu tấn công chưa từng biết (Zero-day) hoặc các hành vi quét mạng, làm ngập lụt hệ thống.
- **Hỗ trợ điều tra số (Digital Forensics):** Lưu trữ nhật ký (Log) và dữ liệu gói tin (PCAP) để tái hiện lại diễn biến sự kiện trong quá khứ. Cung cấp bằng chứng số để xác định nguyên nhân gốc rễ (Root Cause Analysis) và quy trách nhiệm sau khi sự cố xảy ra.

Hình 7. Giám sát mạng

1.2. Tầm quan trọng của Giám sát đối với An ninh mạng

Sự hiện diện của hệ thống giám sát thay đổi hoàn toàn thế trận phòng thủ của một hệ thống truyền thông.

Tiêu chí so sánh	Hệ thống KHÔNG CÓ giám sát (Thụ động)	Hệ thống CÓ giám sát (Chủ động)
Cơ chế phát hiện	Chỉ phát hiện sự cố khi dịch vụ đã bị sập hoặc người dùng báo lỗi.	Phát hiện các dấu hiệu rà quét hoặc thăm dò ngay từ khi cuộc tấn công mới bắt đầu.
Thời gian phản ứng	Chậm, thường mất nhiều giờ đến nhiều ngày để xác định nguyên nhân.	Tức thời (Real-time), cho phép ngăn chặn thiệt hại lan rộng.
Khả năng kiểm soát	"Hộp đen" – Quản trị viên không biết chi tiết luồng dữ liệu bên trong.	"Hộp trắng" – Quản trị viên nắm rõ mọi luồng kết nối và trạng thái giao thức.
Rủi ro an ninh	Dễ bị tấn công dai dẳng (APT) mà không hay biết.	Giảm thiểu rủi ro bị chiếm quyền điều khiển hoặc đánh cắp dữ liệu.

Bảng 1. So sánh trạng thái an ninh giữa hệ thống có và không có giám sát

1.3. Các kiến trúc triển khai hệ thống giám sát

Trong kỹ thuật mạng, có hai mô hình kiến trúc chính để thu thập dữ liệu giám sát, mỗi mô hình có đặc điểm kỹ thuật và phạm vi ứng dụng riêng biệt.

Đặc điểm	Giám sát tại Máy trạm (Host-based / HIDS)	Giám sát tại Mạng / Cửa ngõ (Network-based / NIDS)
Vị trí triển khai	Cài đặt phần mềm (Agent) trực tiếp lên từng máy chủ hoặc máy tính người dùng.	Đặt thiết bị giám sát tại các điểm nút giao thông (Chokepoints) như Gateway, Router, Switch Core.
Phạm vi thu thập	Chỉ nhìn thấy các gói tin đi vào và đi ra khỏi chính thiết bị đó.	Nhìn thấy toàn bộ lưu lượng của mạng con (Subnet) đi qua điểm nút, bao gồm cả lưu lượng của các thiết bị không cài Agent.
Cơ chế hoạt động	Phân tích log hệ điều hành, file hệ thống và gói tin cục bộ.	Phân tích tiêu đề (Header) và nội dung (Payload) của gói tin trên đường truyền.
Ưu điểm	Có thể giải mã dữ liệu mã hóa (SSL/TLS) tại đích. Phát hiện được sự thay đổi file hệ thống.	Không ảnh hưởng hiệu năng máy trạm. Không phụ thuộc vào hệ điều hành của máy trạm. Khó bị kẻ tấn công vô hiệu hóa hơn.
Nhược điểm	Tốn tài nguyên (CPU/RAM) của máy chủ. Khó triển khai và quản lý trên quy mô lớn.	Khó phân tích dữ liệu mã hóa. Cần xử lý băng thông lớn tại một điểm tập trung.

Bảng 2. Phân tích các mô hình kiến trúc giám sát

1.4. Chế độ hoạt động của Card mạng trong giám sát (Promiscuous Mode)

Để mô hình giám sát tại mạng (Network-based) hoạt động hiệu quả, thiết bị giám sát cần được cấu hình card mạng ở chế độ đặc biệt.

Chế độ	Nguyên lý hoạt động	Ứng dụng
Chế độ thường (Non-promiscuous)	Card mạng chỉ nhận và xử lý các gói tin có địa chỉ MAC đích trùng với địa chỉ MAC của chính nó (hoặc gói tin Broadcast). Các gói tin khác bị loại bỏ ngay tại tầng phần cứng.	Sử dụng cho các máy trạm, server thông thường để đảm bảo hiệu năng và bảo mật.
Chế độ hỗn tạp (Promiscuous Mode)	Card mạng được cấu hình để **thu nhận toàn bộ** các gói tin đi ngang qua dây cáp hoặc sóng vô tuyến, bất kể địa chỉ đích là gì.	Bắt buộc phải có đối với các thiết bị giám sát mạng (Sniffer), IDS, IPS hoặc các Gateway quản lý để phân tích toàn bộ lưu lượng mạng.

Bảng 3. Các chế độ hoạt động của Card giao tiếp mạng (NIC)

2. Các mối đe dọa và sự cố mạng thường gặp

2.1. Phân loại các cuộc tấn công và bất thường mạng

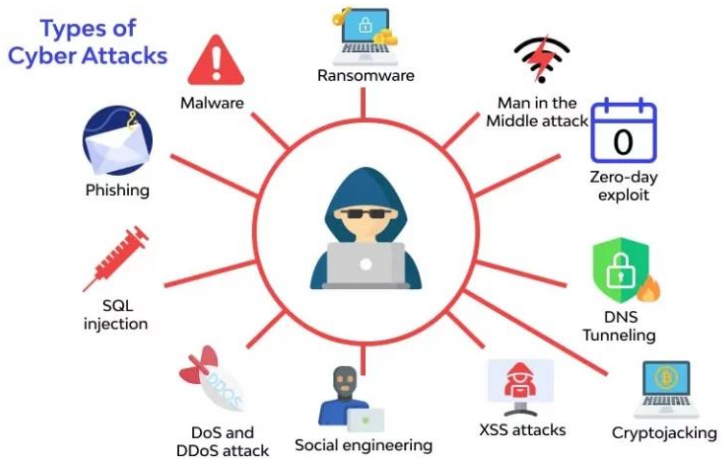
Các mối đe dọa mạng có thể được nhận diện thông qua các đặc trưng kỹ thuật (signatures) và hành vi lưu lượng (traffic behavior). Cơ chế hoạt động và dấu hiệu nhận biết của các loại hình tấn công phổ biến gồm:

- **Trinh sát mạng (Reconnaissance):**
    - **Cơ chế kỹ thuật (Nguyên lý hoạt động):** Đây là giai đoạn thu thập thông tin sơ bộ. Kẻ tấn công sử dụng các kỹ thuật quét (scanning) để xác định cấu trúc mạng, các máy chủ đang hoạt động (Active Hosts) và các cổng dịch vụ đang mở (Open Ports).
        - **Host Discovery:** Gửi gói tin ICMP Echo hoặc ARP Request.
        - **Port Scanning:** Gửi gói tin TCP với cờ SYN (kỹ thuật Half-open) để thăm dò phản hồi.
    - **Dấu hiệu nhận biết trên mạng (Network Indicators):**

        - **Mô hình lưu lượng:** Xuất hiện một địa chỉ IP nguồn duy nhất gửi yêu cầu kết nối đến một dải IP đích rộng (Quét ngang) hoặc đến hàng loạt cổng trên một IP đích (Quét dọc).
        - **Tần suất:** Số lượng gói tin khởi tạo kết nối tăng cao trong thời gian ngắn mà không có dữ liệu trao đổi thực tế.
    - **Hậu quả tiềm tàng:** Lộ lọt thông tin về hạ tầng, phiên bản hệ điều hành và các dịch vụ đang chạy. Đây là tiền đề để kẻ tấn công lựa chọn phương thức khai thác lỗ hổng phù hợp.
- **Tấn công Từ chối Dịch vụ (DoS/DDoS):**
    - **Cơ chế kỹ thuật (Nguyên lý hoạt động):** Tấn công nhằm mục đích làm cạn kiệt tài nguyên hệ thống (CPU, RAM, Băng thông) khiến người dùng hợp lệ không thể truy cập dịch vụ.
        - **Volumetric Attacks:** Làm ngập đường truyền (ví dụ: UDP Flood).
        - **Protocol Attacks:** Tận dụng điểm yếu của giao thức (ví dụ: SYN Flood làm đầy bảng kết nối).
    - **Dấu hiệu nhận biết trên mạng (Network Indicators):**
        - **Lưu lượng:** Băng thông mạng bị chiếm dụng đột ngột, đạt đỉnh (Traffic Spike) vượt xa mức nền (Baseline) thông thường.
        - **Đặc điểm gói tin:** Lượng lớn gói tin có kích thước giống hệt nhau, nội dung vô nghĩa, hoặc giả mạo địa chỉ IP nguồn (IP Spoofing).
    - **Hậu quả tiềm tàng:** Gây tê liệt toàn bộ hệ thống mạng hoặc máy chủ dịch vụ, dẫn đến gián đoạn hoạt động kinh doanh và mất mát doanh thu.
- **Bất thường Giao thức (Protocol Anomalies):**
    - **Cơ chế kỹ thuật (Nguyên lý hoạt động):** Bao gồm các sai lệch so với tiêu chuẩn RFC của giao thức TCP/IP. Những bất thường này thường do lỗi cấu hình, lỗi phần mềm, hoặc do mã độc sử dụng để lẩn tránh sự phát hiện.
        - **Kênh ngầm (Covert Channel):** Giấu dữ liệu trong các trường không sử dụng của gói tin.

- **Cờ TCP không hợp lệ:** Sử dụng tổ hợp cờ phi logic (ví dụ: bật cả SYN và FIN).

- **Dấu hiệu nhận biết trên mạng (Network Indicators):**
  - **Kích thước:** Gói tin có kích thước quá nhỏ (Tiny fragments) hoặc quá lớn (vượt MTU).
  - **Trạng thái:** Xuất hiện nhiều gói tin TCP Reset (RST) ngắt kết nối bất thường; hoặc giao tiếp diễn ra trên các cổng không tiêu chuẩn (Non-standard ports).
- **Hậu quả tiềm tàng:** Có thể là dấu hiệu của việc dữ liệu đang bị đánh cắp (Data Exfiltration), mã độc đang giao tiếp với máy chủ điều khiển (C&C), hoặc hệ thống đang hoạt động kém ổn định.



Hình 8. Các loại tấn công mạng

2.2. Các thách thức trong việc phát hiện

Việc nhận diện chính xác các mối đe dọa trên gặp phải nhiều thách thức về mặt kỹ thuật, đòi hỏi sự kết hợp của nhiều phương pháp phân tích khác nhau.

Thách thức	Mô tả	Hệ quả
Mã hóa dữ liệu (Encryption)	Phần lớn lưu lượng mạng hiện đại (như HTTPS) được mã hóa, làm ẩn nội dung (Payload) bên trong.	Các phương pháp phân tích dựa trên chữ ký (Signature-based) truyền thống trở nên kém hiệu quả vì không đọc được nội dung gói tin. Cần chuyển sang phân tích hành vi (Behavior-based).
Báo động giả (False Positives)	Các hành vi mạng hợp lệ đôi khi có đặc điểm giống tấn công (ví dụ: tải file lớn có thể giống DoS, quét mạng nội bộ hợp pháp giống Reconnaissance).	Gây nhiễu cho hệ thống giám sát, làm lãng phí thời gian của quản trị viên và có thể dẫn đến việc chặn nhầm người dùng hợp lệ.
Tốc độ và Khối lượng dữ liệu	Mạng gigabit tạo ra hàng triệu gói tin mỗi giây. Việc phân tích sâu (Deep Packet Inspection) cho từng gói tin đòi hỏi tài nguyên tính toán khổng lồ.	Hệ thống giám sát cần có cơ chế lọc thông minh hoặc lấy mẫu (Sampling) để đảm bảo hiệu năng mà không bỏ sót các dấu hiệu quan trọng.

Bảng 4. Thách thức kỹ thuật trong giám sát an ninh mạng

3. Các phương pháp phát hiện bất thường (Anomaly Detection)

3.1. Phân loại các phương pháp tiếp cận

Các kỹ thuật phát hiện bất thường hiện đại thường được phân loại dựa trên cách thức học từ dữ liệu và loại thuật toán sử dụng.

Nhóm phương pháp	Nguyên lý hoạt động	Đặc điểm nổi bật	Phạm vi ứng dụng
Thống kê (Statistical)	Xây dựng mô hình phân phối xác suất của dữ liệu bình thường. Các điểm dữ liệu có xác suất xuất hiện thấp dưới ngưỡng quy định được coi là bất thường.	Đơn giản, dễ triển khai. Hiệu quả với dữ liệu tuân theo phân phối chuẩn.	Phát hiện các thay đổi đột ngột về lưu lượng (Traffic Spikes) hoặc lỗi phần cứng đơn giản.
Học máy (Machine Learning)	Sử dụng các thuật toán để tự động học các đặc trưng của dữ liệu mà không cần lập trình quy tắc cụ thể. Bao gồm học có giám sát (Supervised) và không giám sát (Unsupervised).	Khả năng thích ứng cao. Có thể xử lý dữ liệu nhiều chiều (multi-dimensional).	Phân tích log hệ thống, phát hiện xâm nhập mạng, gian lận tài chính.
Học sâu (Deep Learning)	Sử dụng mạng nơ-ron đa lớp (Neural Networks) để mô phỏng khả năng học của não bộ, tự động trích xuất các đặc trưng phức tạp từ dữ liệu thô.	Độ chính xác cực cao với dữ liệu lớn. Tự động trích xuất đặc trưng (Feature extraction).	Xử lý dữ liệu chuỗi thời gian (Time-series), hình ảnh, hoặc tín hiệu phức tạp.

Bảng 5. Tổng quan các nhóm phương pháp phát hiện bất thường

3.2. Các thuật toán Học máy phổ biến trong giám sát mạng

Bảng dưới đây so sánh các thuật toán phổ biến:

Thuật toán	Loại hình	Cơ chế hoạt động	Ưu điểm kỹ thuật
Isolation Forest (Rừng cô lập)	Không giám sát (Unsupervised)	Thay vì mô hình hóa dữ liệu bình thường, thuật toán chủ động "cô lập" các điểm dị biệt bằng cách xây dựng các cây quyết định ngẫu nhiên. Các điểm bất thường sẽ nằm ở các nhánh nông hơn của cây.	**Hiệu suất cao:** Độ phức tạp tuyến tính, xử lý nhanh lượng dữ liệu lớn. **Không cần gán nhãn:** Hoạt động tốt ngay cả khi không có dữ liệu lỗi mẫu.
Local Outlier Factor (LOF)	Không giám sát	Đánh giá mức độ bất thường của một điểm dựa trên mật độ (density) của nó so với các điểm lân cận.	Hiệu quả trong việc phát hiện các bất thường cục bộ (local outliers) trong các cụm dữ liệu có mật độ khác nhau.
One-Class SVM	Không giám sát / Bán giám sát	Tìm một siêu phẳng (hyperplane) bao quanh các điểm dữ liệu bình thường trong không gian đặc trưng. Bất kỳ điểm nào nằm ngoài biên giới này được xem là bất thường.	Hiệu quả với dữ liệu nhiều chiều. Phù hợp khi dữ liệu bất thường rất hiếm hoặc không có sẵn để huấn luyện.

LSTM (Long Short-Term Memory)	Học sâu (Deep Learning)	Một dạng mạng nơ-ron hồi quy (RNN) có khả năng ghi nhớ các phụ thuộc dài hạn trong dữ liệu chuỗi thời gian.	Rất mạnh trong việc phát hiện các bất thường dựa trên ngữ cảnh thời gian (ví dụ: xu hướng tăng chậm của tấn công DoS).

Bảng 6. So sánh các thuật toán Học máy phát hiện bất thường

3.3. Lý thuyết chuyên sâu về Isolation Forest

Trong bối cảnh giám sát lưu lượng mạng thời gian thực, **Isolation Forest** thường được ưu tiên lựa chọn làm giải pháp nền tảng do các đặc tính kỹ thuật vượt trội so với các phương pháp dựa trên khoảng cách hoặc mật độ truyền thống.

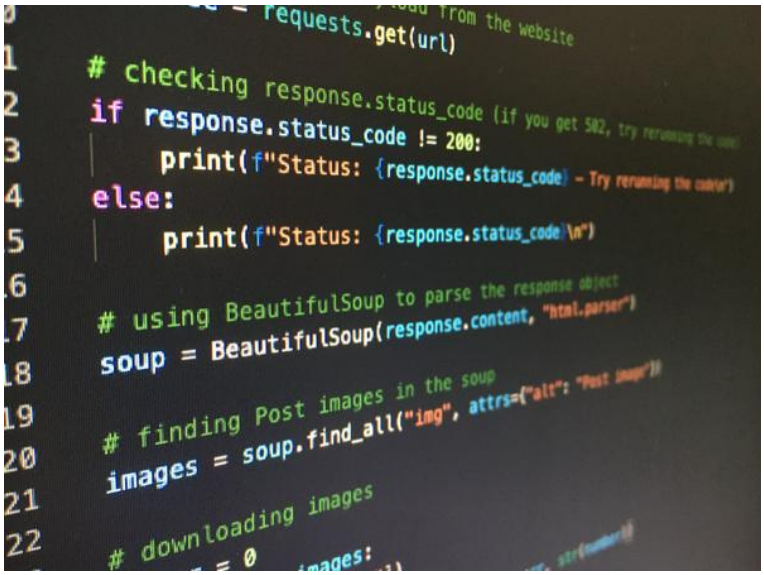
Đặc tính	Giải thích lý thuyết
Giả định nền tảng	Dữ liệu bất thường có hai tính chất: (1) Số lượng ít (Few) và (2) Giá trị khác biệt (Different). Do đó, chúng dễ bị cô lập hơn so với dữ liệu bình thường.
Cơ chế phân vùng	Thuật toán chọn ngẫu nhiên một thuộc tính và một giá trị cắt (split value) để chia dữ liệu. Quá trình này lặp lại đệ quy cho đến khi mọi điểm dữ liệu được cô lập.
Điểm số bất thường (Anomaly Score)	Được tính dựa trên độ dài đường đi trung bình từ gốc cây đến nút lá chứa điểm dữ liệu đó. Đường đi càng ngắn (cô lập càng nhanh) → Điểm số bất thường càng cao.
Ưu thế xử lý mạng	Không cần tính toán khoảng cách Euclide tốn kém (như K-Means hay KNN), giúp giảm tải CPU đáng kể khi xử lý hàng nghìn gói tin/giây.

Bảng 7. Đặc tính kỹ thuật của Isolation Forest

4. Các công nghệ và thư viện hỗ trợ phát triển

4.1. Ngôn ngữ lập trình và môi trường

Trong lĩnh vực an ninh mạng và khoa học dữ liệu, việc lựa chọn ngôn ngữ lập trình đóng vai trò quyết định đến khả năng mở rộng và tích hợp của hệ thống. Python là ngôn ngữ lập trình bậc cao, thông dịch, hướng đối tượng. Nổi bật với cú pháp rõ ràng và hệ sinh thái thư viện khổng lồ. Được xem là ngôn ngữ tiêu chuẩn trong an ninh mạng (Cybersecurity) nhờ khả năng viết mã nhanh (Rapid Prototyping) và tích hợp tốt với các thư viện xử lý gói tin cũng như thuật toán AI.



Hình 9. Ngôn ngữ lập trình Python

4.2. Các thư viện chuyên dụng

Hệ thống giám sát mạng thường được xây dựng dựa trên sự phối hợp của các nhóm thư viện chức năng sau:

Nhóm chức năng	Thư viện tiêu chuẩn	Chức năng và Ứng dụng lý thuyết
Thu thập & Phân tích Gói tin	**Scapy**	Một công cụ thao tác gói tin tương tác mạnh mẽ. Khác với các thư viện chỉ bắt gói tin thụ động, Scapy cho phép: • **Sniffing:** Bắt gói tin từ đường truyền vật lý. • **Dissecting:** Giải mã cấu trúc các giao thức mạng (Ethernet, IP, TCP, UDP, ICMP, ARP...). • **Forging:** Tạo và gửi các gói tin tùy chỉnh để kiểm thử phản ứng của mạng.
Học máy & Phân tích dữ liệu	**Scikit-learn**	Thư viện mã nguồn mở cung cấp các công cụ phân tích dữ liệu và khai phá dữ liệu hiệu quả. Trong bảo mật, nó cung cấp các thuật toán như: • **Isolation Forest:** Để phát hiện bất thường trong luồng dữ liệu. • **Random Forest / SVM:** Để phân loại loại hình tấn công dựa trên dữ liệu đã gán nhãn.
Giao diện người dùng (GUI)	**Tkinter**	Thư viện giao diện đồ họa tiêu chuẩn của Python. Cung cấp các widget (nút bấm, bảng, thanh cuộn) để xây dựng các ứng dụng Desktop, giúp chuyển đổi các dòng lệnh phức tạp thành giao diện trực quan cho người quản trị.
Trực quan hóa dữ liệu	**Matplotlib**	Thư viện vẽ đồ thị 2D/3D. Cho phép chuyển đổi các dữ liệu số liệu thô (tần suất, lưu lượng) thành các biểu đồ (Pie chart, Bar chart, Line plot), giúp nhận diện xu hướng và phân bố của lưu lượng mạng một cách nhanh chóng.

Xuất báo cáo	**FPDF / ReportLab**	Các thư viện hỗ trợ tạo tệp PDF lập trình. Cho phép tự động hóa quy trình tạo báo cáo kỹ thuật, bao gồm việc định dạng văn bản, chèn hình ảnh biểu đồ và tạo bảng số liệu từ kết quả phân tích.

Bảng 8. Các thư viện lõi trong xây dựng hệ thống giám sát

CHƯƠNG III. XÂY DỰNG VÀ ĐÁNH GIÁ HỆ THỐNG GIÁM SÁT MẠNG THÔNG MINH

1. Kiến trúc và Môi trường phát triển

1.1. Môi trường phát triển và Kiểm thử

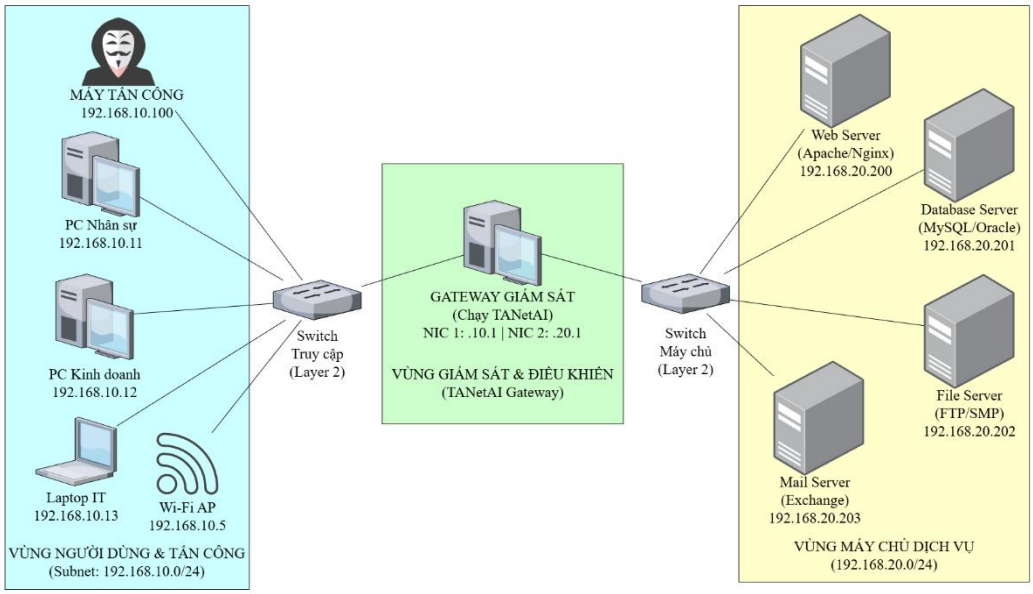
- **Môi trường Phần cứng và Hệ điều hành:** Hệ thống được triển khai trên một mạng nội bộ (LAN) gồm nhiều thiết bị thực tế và giả lập, chạy trên các nền tảng hệ điều hành phổ biến để kiểm chứng tính tương thích đa nền tảng:
    - **Máy Trạm Quản trị & Giám sát (Monitoring Station):**
        - **Hệ điều hành:** Windows (Phiên bản 10/11).
        - **Vai trò:** Đóng vai trò là Gateway trung tâm, nơi cài đặt và vận hành phần mềm TANetAI. Máy này được trang bị 02 giao diện mạng (Network Interfaces) để định tuyến và kiểm soát lưu lượng giữa các phân đoạn mạng khác nhau.
    - **Máy Tấn công và Kiểm thử (Attacker/Tester):**
        - **Hệ điều hành:** Kali Linux (hoặc các bản phân phối Linux chuyên dụng về bảo mật).
        - **Vai trò:** Giả lập các hành vi người dùng và các cuộc tấn công mạng (như Scanning, DoS) để kiểm tra khả năng phát hiện của hệ thống.
    - **Máy Chủ Dịch vụ (Target Servers):**
        - **Hệ điều hành:** Ubuntu Server / Linux.
        - **Vai trò:** Chạy các dịch vụ mạng tiêu chuẩn (Web HTTP, Database) để làm đối tượng bảo vệ.
- **Công cụ và Thư viện Phát triển:** Hệ thống được xây dựng dựa trên các công nghệ mã nguồn mở tiêu chuẩn trong ngành khoa học dữ liệu và an ninh mạng:
    - **Ngôn ngữ lập trình:** Python (Phiên bản 3.x).
    - **Thư viện xử lý mạng:** Scapy (Dùng cho việc bắt gói tin, phân tích giao thức và quét mạng).

- **Thư viện Trí tuệ nhân tạo:** Scikit-learn (Cung cấp thuật toán Isolation Forest cho module phát hiện bất thường).
- **Thư viện Giao diện:** Tkinter (Xây dựng giao diện đồ họa người dùng Desktop).
- **Thư viện Trực quan hóa & Báo cáo:** Matplotlib (Vẽ biểu đồ thống kê) và FPDF (Xuất báo cáo định dạng PDF).

1.2. Sơ đồ Kiến trúc Mạng (Network Topology)

Để kiểm soát toàn bộ lưu lượng trong mạng mà không cần cài đặt Agent lên từng máy con, hệ thống sử dụng mô hình **Gateway Monitoring**. Máy Giám sát (Monitor) được đặt ở vị trí trung gian, chia tách mạng thành hai phân vùng: Vùng Người dùng/Tấn công và Vùng Máy chủ.

Mọi gói tin giao tiếp giữa hai vùng này bắt buộc phải đi qua máy Monitor, cho phép TANetAI thu thập và phân tích toàn bộ dữ liệu một cách minh bạch.



Hình 10. Sơ đồ Hệ thống Mạng Thực tế - Mô hình Gateway Monitor

Chi tiết cấu hình IP định tuyến:

- **Mạng 1 (Vùng Truy cập):** Dải mạng 192.168.10.0/24.

- Các máy trạm (Client) và máy tấn công (Attacker) được gán IP trong dải này (ví dụ .100, .101).
- Default Gateway của các máy này trỏ về địa chỉ 192.168.10.1 (Giao diện 1 của máy Monitor).

- **Mạng 2 (Vùng Máy chủ):** Dải mạng 192.168.20.0/24.
  - Các máy chủ dịch vụ (Web, DB) nằm trong dải này (ví dụ .200, .201).
  - Default Gateway trỏ về 192.168.20.1 (Giao diện 2 của máy Monitor).
- **Máy Monitor (Gateway):**
  - Thực hiện chức năng **IP Forwarding** (Chuyển tiếp gói tin) để kết nối hai mạng.
  - Phần mềm TANetAI lắng nghe trên cả hai giao diện mạng để giám sát song song cả hai chiều lưu lượng.

2. Thiết kế và Hiện thực các Module Lõi (Back-end)

2.1. Module Quản lý mạng (network_manager.py)

Module này đóng vai trò là lớp tương tác phần cứng (Hardware Abstraction Layer), sử dụng thư viện Scapy để thực hiện hai chức năng chính: Khám phá mạng và Thu thập dữ liệu.

- **Chức năng Khám phá thiết bị (Network Discovery):** Hệ thống sử dụng giao thức ARP (Address Resolution Protocol) để lập bản đồ các thiết bị đang hoạt động trong mạng nội bộ, phục vụ cho tính năng giám sát mục tiêu. **Hàm scan_network(ip_range, iface_name):**
  - **Nguyên lý:** Gửi gói tin ARP Request quảng bá (Broadcast) tới toàn bộ dải IP mạng (ví dụ: 192.168.1.0/24).
  - **Xử lý:** Thu thập các gói tin ARP Reply trả về để trích xuất cặp thông tin IP - MAC Address của các thiết bị đang online.
- **Chức năng Bắt gói tin (Packet Sniffing):** Đây là chức năng quan trọng nhất, biến máy tính thành một thiết bị giám sát thụ động.

- **Cấu hình Promiscuous:** Trong hàm khởi tạo __init__, hệ thống kích hoạt chế độ hỗn tạp bằng lệnh conf.sniff_promisc = True. Điều này cho phép card mạng thu nhận cả các gói tin không được gửi đích danh cho nó, hiện thực hóa mô hình giám sát Gateway.
- **Hàm start_sniffing(...):** Sử dụng hàm sniff() của Scapy với tham số store=False để xử lý gói tin theo luồng (stream) thay vì lưu vào RAM, giúp hệ thống hoạt động liên tục mà không bị tràn bộ nhớ.

2.2. Module AI Thống kê (anomaly_detector.py)

Module này chịu trách nhiệm phát hiện các bất thường chưa biết trước (Zero-day anomalies) dựa trên sự sai lệch về mặt thống kê của gói tin.

Đặc trưng (Feature)	Ý nghĩa	Cách trích xuất
Length	Kích thước gói tin	len(packet) - Phát hiện các gói tin quá khổ hoặc quá nhỏ (như trong tấn công phân mảnh).
Protocol	Giao thức tầng mạng	packet[IP].proto - Nhận diện các giao thức lạ hoặc hiếm gặp.
Source Port	Cổng nguồn	packet[TCP].sport - Phát hiện các tiến trình lạ mở cổng ngẫu nhiên.
Dest Port	Cổng đích	packet[TCP].dport - Phát hiện việc truy cập vào các dịch vụ không tiêu chuẩn.

Bảng 9. Cấu trúc dữ liệu đầu vào cho AI

- **Thuật toán:** Sử dụng **Isolation Forest** từ thư viện scikit-learn.
- **Quy trình vận hành:**
  - **Huấn luyện (fit_model):** Hệ thống thu thập n_packets_to_train (mặc định 3000 gói) đầu tiên để xây dựng mô hình hành vi nền.
  - **Dự đoán (process_packet):** Các gói tin tiếp theo được đưa qua mô hình. Nếu kết quả là -1, gói tin bị đánh dấu là "Bất thường".

- **Giải thích (explain_anomaly):** Hệ thống so sánh đặc trưng của gói tin bất thường với giá trị trung bình (Mean) và độ lệch chuẩn (Std) của tập huấn luyện để đưa ra lý do cụ thể (ví dụ: "Kích thước lớn bất thường").

2.3. Module Phân tích Hành vi (behavioral_analyzer.py)

Module này bổ sung cho AI thống kê bằng cách phát hiện các cuộc tấn công có kịch bản rõ ràng (Signature/Behavior-based) dựa trên tần suất và mẫu hình kết nối.

- **Phát hiện Tấn công Lũ lụt (DoS/Flood Detection):**
  - **Logic:** Sử dụng hàm _check_floods để theo dõi số lượng gói tin gửi đến một đích cụ thể trong cửa sổ thời gian trượt flood_window (2 giây).
  - **Ngưỡng:** Nếu số lượng vượt quá flood_count (2000 gói), hệ thống cảnh báo tấn công DoS.
- **Phát hiện Quét mạng (Scanning Detection):**
  - **Logic:** Hàm _check_scans duy trì một bảng băm (hash map) scan_tracker để đếm số lượng kết nối từ một IP nguồn đến các đích khác nhau.
  - **Kịch bản:**
    - **Port Scan:** Một nguồn kết nối đến > 40 cổng trên cùng một máy đích.
    - **Host Scan:** Một nguồn kết nối đến > 40 máy đích khác nhau trên cùng một cổng.

2.4. Các tiện ích dùng chung (app_utils.py & config.json)

Để đảm bảo tính linh hoạt và khả năng đóng gói đa nền tảng, các thành phần cấu hình và tài nguyên được tách biệt.

- **Quản lý Tài nguyên (app_utils.py):**
  - Hàm resource_path: Tự động xác định đường dẫn tuyệt đối của các file tài nguyên (icon, model, config) dù chương trình đang chạy ở dạng mã nguồn hay dạng file thực thi .exe (được giải nén vào thư mục tạm _MEIPASS của PyInstaller).

- VERSION_HISTORY: Lưu trữ toàn bộ nhật ký phát triển phần mềm, giúp quản lý phiên bản tập trung.

- **Cấu hình Động (config.json):** Lưu trữ các tham số nhạy cảm như ngưỡng phát hiện (flood_count, portscan_count) và tham số huấn luyện AI (training_packets). Điều này cho phép quản trị viên tinh chỉnh độ nhạy của hệ thống mà không cần can thiệp vào mã nguồn (Hard-coding).

3. Thiết kế và Hiện thực Giao diện Người dùng (Front-end)

3.1. Cửa sổ chính và Điều phối luồng (main.py)

Module main.py đóng vai trò là bộ khung (Skeleton) của ứng dụng, quản lý vòng đời của cửa sổ chính và điều phối dữ liệu giữa các thành phần.

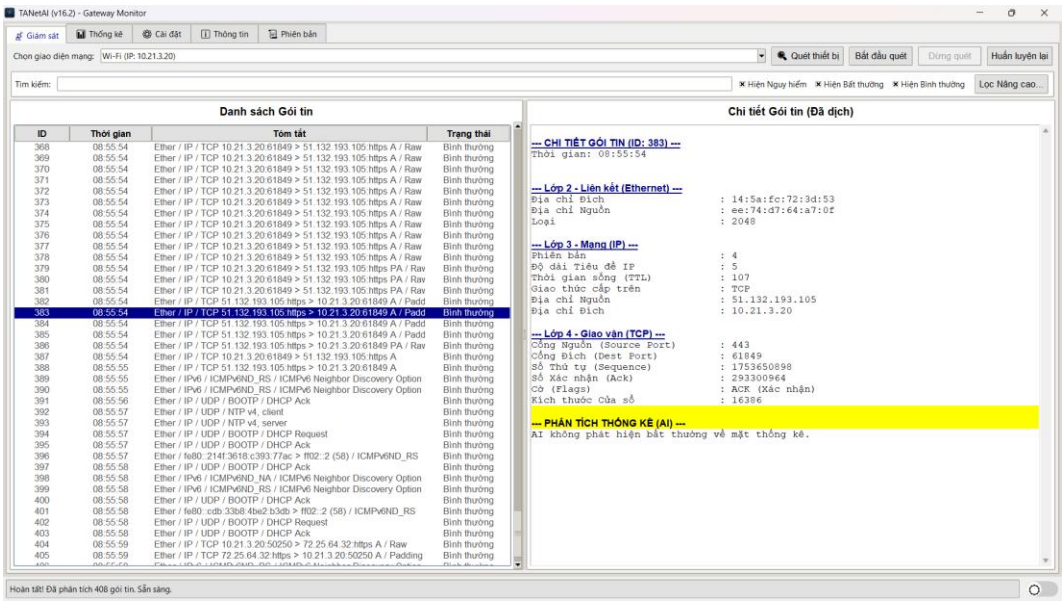
- **Cơ chế Đa luồng (Multithreading Architecture):** Một thách thức lớn trong lập trình mạng là việc bắt gói tin (Sniffing) là một tác vụ chặn (blocking operation). Nếu chạy trên luồng chính, giao diện sẽ bị "đơ" ngay lập tức.
  - **Giải pháp:** Hệ thống tách luồng xử lý thành 2 phần riêng biệt:
    - **Luồng Giao diện (Main Thread):** Chỉ chịu trách nhiệm vẽ UI và nhận sự kiện người dùng.
    - **Luồng Quét (Scanner Thread):** Chạy ngầm hàm run_scanner_thread, thực hiện việc bắt và phân tích gói tin.
  - **Cơ chế giao tiếp:** Hai luồng trao đổi dữ liệu thông qua một hàng đợi an toàn (queue.Queue). Luồng quét đẩy kết quả vào hàng đợi, và luồng giao diện định kỳ (mỗi 100ms) kiểm tra hàng đợi thông qua hàm process_queue để cập nhật hiển thị.
- **Tính năng Tiện ích toàn cục:**
  - **Theme (Giao diện Sáng/Tối):** Lớp ThemeToggle được thiết kế riêng sử dụng Canvas để vẽ nút công tắc trượt hiện đại, thay thế cho nút bấm truyền thống.
  - **Global Copy/Paste:** Hàm setup_global_copy_paste can thiệp vào sự kiện bàn phím và chuột của ứng dụng, cho phép người dùng sao chép nội

dung từ bất kỳ đâu (bảng, ô nhập liệu) bằng phím tắt Ctrl+C hoặc menu chuột phải.

3.2. Tab Giám sát (gui/tab_monitor.py)

Đây là màn hình làm việc chính của quản trị viên, được thiết kế để hiển thị thông tin trực quan và chi tiết.

- **Bảng Danh sách Gói tin (Traffic List):**
  - Sử dụng Widget Treeview để hiển thị dữ liệu dạng bảng với các cột: *ID, Thời gian, Tóm tắt, Trạng thái*.
  - **Mã màu thông minh:** Các dòng được tô màu tự động dựa trên mức độ nghiêm trọng (Normal: Trắng/Xám, Anomaly: Vàng, Danger: Đỏ) giúp nhận diện nhanh mối đe dọa.
- **Khung Chi tiết (Packet Details):**
  - Sử dụng ScrolledText để hiển thị nội dung giải mã của từng lớp giao thức (Ethernet, IP, TCP...).
  - Các thông tin quan trọng được làm nổi bật (Highlight) và dịch sang Tiếng Việt thông qua từ điển TRANSLATION_MAP.
- **Bộ lọc (Filters):** Cung cấp các Checkbox để lọc gói tin theo giao thức (TCP, UDP, ICMP) hoặc theo trạng thái (chỉ hiện Nguy hiểm/Bất thường), giúp người dùng tập trung vào dữ liệu quan trọng.

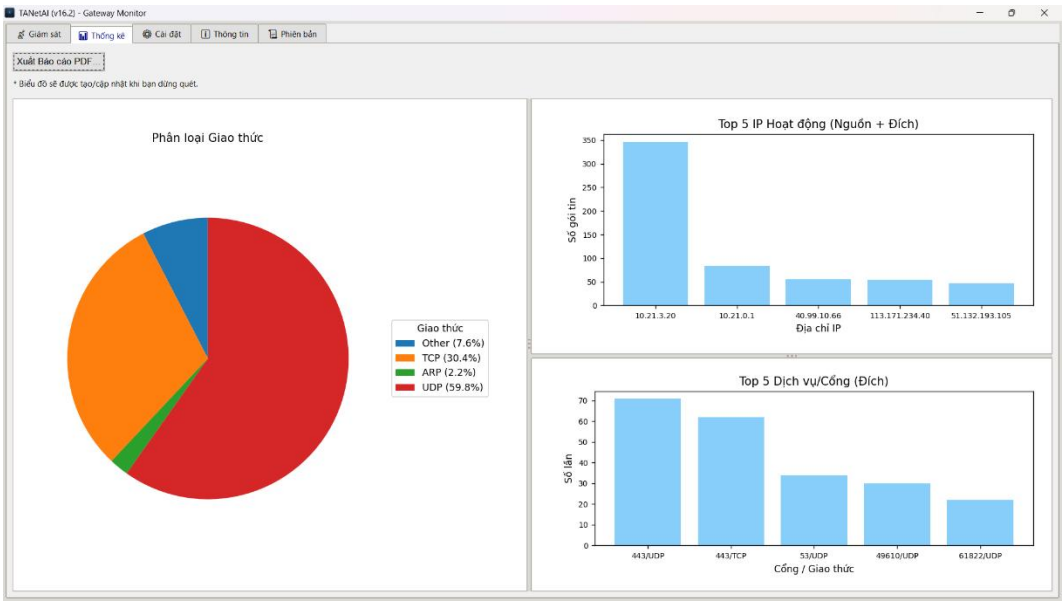


Hình 11. Tab Giám sát

3.3. Tab Thống kê và Trực quan hóa (gui/tab_statistics.py)

Module này chuyển đổi dữ liệu thô thành các biểu đồ quản trị (Dashboard), giúp đánh giá tổng quan tình trạng mạng.

- **Công nghệ:** Tích hợp thư viện Matplotlib vào giao diện Tkinter thông qua lớp FigureCanvasTkAgg.
- **Các loại biểu đồ:**
    - **Biểu đồ Tròn (Pie Chart):** Thể hiện tỷ lệ các giao thức mạng. Được tối ưu hóa bằng cách đưa số liệu chi tiết vào bảng chú thích (Legend) để tránh chồng chéo khi có nhiều phần tử nhỏ.
    - **Biểu đồ Cột (Bar Chart):** Thể hiện "Top 5 IP Hoạt động" và "Top 5 Cổng Dịch vụ", giúp phát hiện nhanh các máy trạm đang chiếm dụng băng thông hoặc các cổng đang bị tấn công.
- **Cơ chế cập nhật:** Biểu đồ được vẽ lại tự động mỗi khi người dùng dừng quét hoặc chuyển đổi giao diện (Sáng/Tối) để đảm bảo tính đồng bộ về thẩm mỹ.



Hình 12. Tab Thống kê

3.4. Module Báo cáo (pdf_report.py)

Hệ thống cung cấp khả năng xuất báo cáo vi phạm tự động, phục vụ cho công tác lưu trữ và báo cáo cấp trên.

- **Công nghệ:** Sử dụng thư viện FPDF.
- **Cấu trúc báo cáo:**
    - **Trang bìa:** Tóm tắt thông tin phiên quét (Thời gian, Tổng số gói tin, Số lượng cảnh báo theo phân loại).
    - **Trang Biểu đồ:** Các biểu đồ từ Tab Thống kê được lưu tạm thành ảnh chất lượng cao (High DPI) và nhúng vào PDF theo bố cục dọc (Vertical Stack) để tối ưu hóa khổ giấy A4.
    - **Bảng Chi tiết:** Liệt kê toàn bộ các gói tin bị cảnh báo kèm theo lý do phát hiện (ví dụ: "Sử dụng cổng hiểm", "Phát hiện Scan"), được định dạng bảng (Table) ngay ngắn với font chữ hỗ trợ Tiếng Việt đầy đủ.

4. Quy trình Đóng gói và Triển khai

4.1. Đóng gói cho môi trường Windows

Quy trình đóng gói trên Windows được thực hiện qua hai giai đoạn: Biên dịch mã nguồn thành tệp thực thi (.exe) và Tạo bộ cài đặt (Setup.exe).

- **Biên dịch mã nguồn với PyInstaller** Công cụ PyInstaller được sử dụng để "đóng băng" (freeze) mã nguồn Python và các thư viện phụ thuộc vào một tệp tin duy nhất.
    - **Lệnh thực thi:**

```
&
"C:\Users\tanya\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.13_qbz5n2kfra8p0\LocalCache\local-packages\Python313\Scripts\pyinstaller.exe" --onefile --windowed --uac-admin --icon="my_icon.ico" --name="TANetAI" --collect-data="matplotlib" --collect-data="fpdf2" --add-data="my_icon.ico;." --add-data="config.json;." --add-data="fonts;fonts" main.py
```

    - **Giải thích các tham số kỹ thuật:**
        - --onefile: Gộp toàn bộ tài nguyên (DLL, Python interpreter, code) vào một file .exe duy nhất để dễ quản lý.
        - --windowed: Chạy chương trình ở chế độ cửa sổ (GUI), ẩn cửa sổ dòng lệnh (Console) phía sau để tăng tính thẩm mỹ.
        - --uac-admin: Tự động yêu cầu quyền Quản trị viên (Administrator) khi khởi chạy. Đây là yêu cầu bắt buộc để Scapy có thể truy cập tầng phần cứng mạng.
        - --add-data: Nhúng các tài nguyên tĩnh (file config.json, icon, font chữ) vào bên trong file .exe để chương trình hoạt động độc lập.
- **Tạo bộ cài đặt với Inno Setup:** Sau khi có file TANetAI.exe, phần mềm **Inno Setup** được sử dụng để tạo file cài đặt chuyên nghiệp Setup.exe. Kịch bản cài đặt (.iss) được thiết kế để thực hiện các tác vụ:
    - **Cài đặt Npcap:** Tự động kích hoạt trình cài đặt Npcap (thư viện bắt gói tin nền tảng) nếu máy người dùng chưa có.

- **Triển khai file:** Copy file chương trình và cấu hình vào thư mục Program Files.
- **Tạo lối tắt:** Tạo Shortcut trên Desktop và Start Menu để người dùng dễ dàng truy cập.

4.2. Đóng gói cho môi trường Linux

Đối với môi trường Linux (thường dùng cho các máy chủ hoặc trạm giám sát chuyên dụng), quy trình triển khai tập trung vào việc tự động hóa cài đặt các gói phụ thuộc hệ thống.

- **Biên dịch Binary:** Tương tự như Windows, PyInstaller cũng được dùng trên Linux để tạo ra file nhị phân (Binary) có thể chạy trực tiếp (./TANetAI). File này chứa sẵn trình thông dịch Python, giúp chương trình chạy được ngay cả trên các máy chưa cài Python.
- **Script cài đặt tự động (install.sh):** Thay vì file Setup như Windows, hệ thống cung cấp một kịch bản Shell Script (install.sh) để tự động hóa quy trình triển khai:
  - **Kiểm tra quyền Root:** Đảm bảo người dùng chạy script với quyền cao nhất (sudo) để có thể can thiệp vào cấu hình mạng.
  - **Cài đặt thư viện hệ thống:** Tự động tải và cài đặt các gói thư viện cấp thấp cần thiết từ kho lưu trữ (Repository) như libpcap0.8 (cho chức năng bắt gói tin) và python3-tk (cho giao diện).
  - **Triển khai hệ thống:**
    - Copy file thực thi vào thư mục hệ thống /opt/TANetAI/.
    - Cấp quyền thực thi (chmod +x) cho file chương trình.
  - **Tích hợp Desktop:** Tạo file .desktop trong /usr/share/applications/ để tích hợp TANetAI vào menu ứng dụng của hệ điều hành, cho phép người dùng khởi chạy bằng cách click chuột thay vì gõ lệnh.

5. Kịch bản Kiểm thử và Demo (Testing & Demonstration)

5.1. Mô hình thử nghiệm (Lab Setup)

Môi trường kiểm thử được xây dựng trên nền tảng ảo hóa VMware Workstation với mô hình mạng Gateway, bao gồm 03 máy ảo đóng vai trò khác nhau trong một cuộc tấn công mạng.

Vai trò (Role)	Hệ điều hành	Địa chỉ IP	Chức năng nhiệm vụ
Monitor (Gateway)	Windows 10	LAN 1: 192.168.10.1 LAN 2: 192.168.20.1	Chạy phần mềm **TANetAI**. Đóng vai trò Router trung gian chuyển tiếp gói tin giữa hai mạng con. Thực hiện giám sát và phân tích lưu lượng đi qua.
Attacker (Kẻ tấn công)	Kali Linux	192.168.10.100 (GW: 192.168.10.1)	Sử dụng các công cụ tấn công tiêu chuẩn (Nmap, Hping3) để giả lập các mối đe dọa mạng.
Victim (Nạn nhân)	Ubuntu Server	192.168.20.200 (GW: 192.168.20.1)	Chạy dịch vụ Web (Apache) tại cổng 80. Là mục tiêu của các cuộc tấn công.

Bảng 10. Cấu hình chi tiết các nút mạng trong mô hình thử nghiệm

Cấu hình định tuyến: Máy Monitor được kích hoạt tính năng IP Forwarding, đảm bảo mọi gói tin từ máy Attacker muốn đến máy Victim bắt buộc phải đi qua máy Monitor để hệ thống TANetAI có thể bắt giữ và phân tích.

5.2. Kịch bản 1: Giám sát mạng bình thường & Quét thiết bị

Mục tiêu: Kiểm chứng khả năng khám phá thiết bị (Network Discovery) và giám sát lưu lượng hợp lệ.

- **Thực hiện:**

- Trên TANetAI, kích hoạt chức năng **"Quét thiết bị LAN"** trên dải mạng 192.168.20.0/24.
- Chọn IP mục tiêu là máy Victim (192.168.20.200) để giám sát tập trung.
- Từ máy Attacker, sử dụng trình duyệt truy cập vào Website của máy Victim.

- **Kết quả:**
  - Hệ thống nhận diện chính xác địa chỉ IP và MAC của máy Victim.
  - Trên bảng giám sát, các gói tin HTTP/TCP xuất hiện với trạng thái **"Bình thường"** (Màu trắng). Không có cảnh báo giả.

5.3. Kịch bản 2: Phát hiện Tấn công Trinh sát (Port Scanning)

Mục tiêu: Đánh giá khả năng phát hiện hành vi rà quét cổng của module *Behavioral Analyzer*.

- **Thực hiện:** Từ máy Attacker, sử dụng công cụ Nmap để quét 1000 cổng phổ biến trên máy Victim. Lệnh: nmap -sS -p 1-1000 192.168.20.200
- **Kết quả:**
  - TANetAI phát hiện sự gia tăng đột biến số lượng kết nối SYN từ một nguồn duy nhất đến nhiều đích khác nhau.
  - Giao diện hiển thị cảnh báo **Màu Đỏ** với trạng thái **"NGUY HIỂM"**.
  - Thông báo chi tiết: *"PHÁT HIỆN QUÉT CỔNG (Port Scan): Nguồn 192.168.10.100 đang quét 192.168.20.200"*.

5.4. Kịch bản 3: Phát hiện Tấn công Từ chối Dịch vụ (DoS Flood)

Mục tiêu: Kiểm tra khả năng chịu tải và phát hiện tấn công ngập lụt (Volumetric Attack).

- **Thực hiện:** Từ máy Attacker, sử dụng Hping3 để thực hiện tấn công SYN Flood cường độ cao vào cổng 80 của Victim. Lệnh: hping3 --flood -S -p 80 192.168.20.200
- **Kết quả:**
  - Lưu lượng gói tin tăng vọt lên hàng nghìn gói/giây.

- Giao diện TANetAI vẫn hoạt động ổn định (không bị treo) nhờ cơ chế xử lý đa luồng và cập nhật giao diện theo lô (Batch update).
- Sau khi dừng quét, hệ thống liệt kê hàng loạt cảnh báo **"PHÁT HIỆN TẤN CÔNG LŨ LỤT (DoS/Flood)"** dựa trên việc vượt ngưỡng tần suất (flood_count trong cấu hình).

5.5. Kịch bản 4: Xuất báo cáo điều tra

Mục tiêu: Tổng hợp dữ liệu thành bằng chứng số (Digital Evidence).

- **Thực hiện:** Sử dụng chức năng **"Xuất Báo cáo PDF"** trên giao diện Thống kê.
- **Kết quả:**
  - Tệp PDF được tạo ra chứa đầy đủ thông tin: Thời gian thực hiện, Tổng quan phiên quét.
  - **Biểu đồ:** Thể hiện trực quan tỷ lệ giao thức TCP chiếm đa số (do tấn công Flood) và IP của Attacker đứng đầu danh sách hoạt động.
  - **Nhật ký chi tiết:** Liệt kê danh sách các gói tin vi phạm kèm theo thời gian (Timestamp) chính xác, phục vụ cho công tác truy vết và xử lý sự cố sau này.

6. Đánh giá kết quả và Thảo luận

6.1. Đánh giá Hiệu năng (Performance Evaluation)

Hệ thống được kiểm thử trên cấu hình tiêu chuẩn (2 Core CPU, 4GB RAM) để đo lường mức tiêu thụ tài nguyên trong các trạng thái hoạt động khác nhau.

Trạng thái hệ thống	Mức tiêu thụ CPU	Mức tiêu thụ RAM	Đánh giá độ trễ (Latency)
Chờ (Idle)	< 1%	~45 MB	Không có độ trễ.
Giám sát bình thường	5% - 10%	~65 MB	Phản hồi tức thì.
Chịu tải tấn công (Flood)	30% - 50%	~150 MB	Giao diện phản hồi chậm nhẹ (do cập nhật danh sách liên tục), nhưng luồng thu thập vẫn hoạt động tốt.

Bảng 11. Kết quả đo lường hiệu năng vận hành

6.2. Phân tích Ưu điểm và Hạn chế

Dựa trên kết quả thực nghiệm, bảng dưới đây tổng hợp các điểm mạnh và những mặt còn tồn tại của hệ thống TANetAI.

Khía cạnh	Ưu điểm (Strengths)	Hạn chế (Limitations)
Tính Ổn định	**Hoạt động bền bỉ:** Hệ thống duy trì kết nối liên tục trong thời gian dài, không xảy ra lỗi thoát đột ngột (crash). Cơ chế đa luồng giúp tách biệt việc xử lý dữ liệu và hiển thị.	**Yêu cầu phần cứng:** Do đặc thù của ngôn ngữ Python thông dịch, hệ thống đòi hỏi cấu hình phần cứng trung bình. Trên các máy tính cấu hình yếu (CPU đơn nhân, RAM thấp), hiện tượng quá tải cục bộ có thể xảy ra.
Khả năng Xử lý	**Xử lý song song:** Có thể vừa giám sát thời gian thực, vừa phân tích AI và ghi log cùng lúc mà không bị gián đoạn.	**Nguy cơ nghẽn cổ chai:** Trong tình huống bị tấn công Flood cường độ cực lớn (>50.000 gói/giây), hàng đợi xử lý (Queue) có thể bị đầy nhanh chóng, dẫn đến việc một số gói tin cảnh báo bị trễ hoặc lược bỏ.
Trải nghiệm	**Trực quan & Dễ dùng:** Giao diện được thiết kế thân thiện, hỗ trợ chế độ Sáng/Tối và các thao tác tiện ích (Copy/Paste, Lọc nhanh).	**Hướng dẫn sử dụng:** Chưa tích hợp sẵn các chỉ dẫn tương tác (Interactive Guide) ngay trên giao diện để hỗ trợ người dùng mới làm quen nhanh.

Bảng 12. Tổng hợp Ưu điểm và Hạn chế của hệ thống

7. Kết luận và Hướng phát triển

7.1. Kết luận

Đề tài đã hoàn thành việc xây dựng hệ thống **TANetAI** – một giải pháp giám sát an ninh mạng tích hợp trí tuệ nhân tạo. Hệ thống đã giải quyết thành công bài toán phát hiện sớm các dấu hiệu trinh sát (Scanning) và tấn công từ chối dịch vụ (DoS) trong môi trường mạng nội bộ. Với khả năng hoạt động độc lập hoặc triển khai như một Gateway giám sát, TANetAI cung cấp cho người quản trị một công cụ đắc lực để quan sát và bảo vệ hạ tầng mạng.

7.2. Hướng phát triển

Để hoàn thiện sản phẩm và mở rộng khả năng ứng dụng trong thực tế, các hướng nghiên cứu và phát triển tiếp theo được đề xuất bao gồm:

- **Đa nền tảng hóa (Cross-platform Support):**
    - Hiện tại hệ thống hoạt động tốt trên Windows và Linux.
    - **Mục tiêu:** Tối ưu hóa mã nguồn và quy trình đóng gói để phát hành phiên bản dành riêng cho hệ điều hành **macOS**, mở rộng phạm vi người dùng.
- **Nâng cao khả năng phòng thủ (Advanced Prevention):**
    - Bổ sung các module phát hiện và ngăn chặn các loại tấn công mạng phổ biến khác như **ARP Spoofing** (giả mạo địa chỉ MAC), **DNS Spoofing** (giả mạo tên miền) và phát hiện mã độc kết nối ra ngoài (C&C Beaconing).
    - Nghiên cứu tích hợp cơ chế tương tác với Firewall hệ thống để tự động chặn (Block) các địa chỉ IP tấn công.
- **Tối ưu trải nghiệm người dùng (User Experience):** Xây dựng tính năng **"Hướng dẫn tương tác" (Interactive Tutorial)** ngay bên trong phần mềm. Tính năng này sẽ hướng dẫn người dùng từng bước (step-by-step) từ cách chọn giao diện mạng, thiết lập cấu hình AI đến cách đọc hiểu các cảnh báo, giúp giảm thiểu thời gian đào tạo vận hành.

PHỤ LỤC

1. Công việc thực tập

Hình 13. Giảng dạy AI tại Trường THCS Colette

2. Video demo

https://www.youtube.com/playlist?list=PLgg0rsNkk2Y2fLyG1EhXLtrA9zwG1rS96

3. Link chương trình

https://github.com/TanyanArchitect/Thuc-tap-cong-nghiep

TÀI LIỆU THAM KHẢO

1. Các loại lỗi thường gặp

- https://www.numberanalytics.com/blog/mastering-fault-management-telecom-networks
- https://umboss.com/blog/network-fault-management/
- https://itecspec.com/spec/3gpp-32-111-1-4-fault-management-concept-and-requirements/
- https://www.innovile.com/resources/insights/mastering-telecom-fault-management-ensuring-seamless-network-operations/
- https://datacalculus.com/en/blog/telecommunications-carriers/telecommunications-technical-support-engineer/fault-detection-and-management-for-telecom-support-engineers
- https://automationforum.co/installation-troubleshooting-of-communication-systems-part-1/
- https://www.innovile.com/products/mobile-telecom-fault-management-system/
- https://www.geeksforgeeks.org/computer-networks/types-of-errors-in-computer-network/
- https://www.generalnote.com/computer-network/error-detection-and-correction/types-of-errors
- https://link.springer.com/chapter/10.1007/978-3-031-19297-5_4
- https://en.wikipedia.org/wiki/Bit_slip
- https://en.wikipedia.org/wiki/Interference_%28communication%29
- https://arxiv.org/abs/2102.11245
- https://en.wikipedia.org/wiki/Byzantine_fault
- https://en.wikipedia.org/wiki/Falsing

2. Công cụ giám sát tự động cơ bản

- https://www.manageengine.com/network-monitoring/network-fault-management.html
- https://ijcrt.org/papers/IJCRT1812394.pdf
- https://en.wikipedia.org/wiki/Paessler_PRTG
- https://en.wikipedia.org/wiki/Zabbix
- https://en.wikipedia.org/wiki/Prometheus_%28software%29
- https://en.wikipedia.org/wiki/Naemon
- https://www.techradar.com/pro/datadog-network-monitoring-review
- https://www.techradar.com/pro/paessler-prtg-network-monitor-23-4-review
- https://www.techradar.com/pro/opennms-review
- https://www.techradar.com/pro/auvik-review

3. Dự đoán lỗi

- https://link.springer.com/article/10.1007/s10489-020-02026-2
- https://en.wikipedia.org/wiki/Predictive_maintenance
- https://www.itu.int/dms_pub/itu-s/opb/jnl/S-JNL-VOL4.ISSUE3-2023-A31-PDF-E.pdf
- https://link.springer.com/chapter/10.1007/978-981-15-8411-4_91
- https://www.philippe-fournier-viger.com/2021_SURVEY_NETWORK_FAULT_MANAGEMENT.pdf
- https://etd.aau.edu.et/bitstreams/5dfaeeae-16a2-4cfd-a7f9-fe0e6f03b689/download
- https://d197for5662m48.cloudfront.net/documents/publicationstatus/213474/preprint_pdf/49a770516c11594c8504d7f61f416878.pdf

4. Đặc điểm của hệ thống truyền thông lớn

- https://en.wikipedia.org/wiki/Ultra-large-scale_systems
- https://en.wikipedia.org/wiki/Scalability

5. Hệ thống cảnh báo và khắc phục tự động

- https://statetechmagazine.com/article/2025/05/self-healing-networks-how-are-they-used-perfcon
- https://powertechjournal.com/index.php/journal/article/view/206
- https://solveforce.com/self-healing-networks-the-future-of-autonomous-networking/
- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5228591&
- https://ijsred.com/volume6/issue6/IJSRED-V6I6P123.pdf
- https://www.researchgate.net/publication/378517886_Self-Healing_Networks_AI-Based_Approaches_for_Fault_Detection_and_Recovery
- https://www.mdpi.com/2073-431X/14/6/233
- https://selinc.com/solutions/p/flisr/
- https://vitria.com/wp-content/uploads/2024/10/New-Frontier-The-Self-Healing-Network.pdf
- https://link.springer.com/chapter/10.1007/978-3-031-75608-5_25

6. Huấn luyện và triển khai mô hình

- https://neptune.ai/blog/mlops-best-practices
- https://en.wikipedia.org/wiki/MLOps
- https://www.geeksforgeeks.org/machine-learning/machine-learning-deployment/
- https://www.ibm.com/think/topics/model-deployment
- https://arxiv.org/abs/2205.02302
- https://arxiv.org/abs/2202.03541
- https://neptune.ai/blog/retraining-model-during-deployment-continuous-training-continuous-testing
- https://en.wikipedia.org/wiki/Feature_store

- https://en.wikipedia.org/wiki/ModelOps

7. Phát hiện bất thường (Anomaly detection)

- https://datacalculus.com/en/blog/telecommunications-carriers/telecommunications-network-support-engineer/network-anomaly-detection-for-telecom-carriers
- https://www.anodot.com/blog/anomaly-detection-in-telecommunications/
- https://en.wikipedia.org/wiki/Isolation_forest
- https://www.mdpi.com/1424-8220/23/11/5340
- https://en.wikipedia.org/wiki/Local_outlier_factor
- https://en.wikipedia.org/wiki/Autoencoder
- https://arxiv.org/abs/1811.05269
- https://www.siberoloji.com/machine-learning-for-anomaly-detection-in-network-traffic/
- https://www.mdpi.com/2078-2489/12/5/215
- https://www.reddit.com/r/Nokia_stock/comments/oo2bib/nokia_and_vodafone_harness_machine_learning_on/

8. Phát hiện lỗi thủ công

- https://www.ericsson.com/en/blog/2019/3/automating-fault-detection-for-management-systems-using-ml

9. Phân loại lỗi (Fault Classification)

- https://www.mdpi.com/2076-3417/15/13/7580
- https://cuestionesdefisioterapia.com/index.php/es/article/view/3030
- https://www.mdpi.com/1996-1073/16/22/7680
- https://digital-library.theiet.org/doi/full/10.1049/tje2.12324
- https://www.mdpi.com/2076-3417/15/11/6263
- https://dergipark.org.tr/en/pub/dumf/issue/70546/1096691

- https://www.mdpi.com/2079-9292/11/10/1609
- https://www.reddit.com/r/MachineLearning/comments/1d9jxt0/d_is_multilabel_classification_the_best_approach/

10. Thu thập và xử lý dữ liệu

- https://www.mdpi.com/1999-4893/15/11/432
- https://link.springer.com/article/10.1007/s10489-020-02026-2
- https://www.mdpi.com/1424-8220/24/2/401
- https://arxiv.org/pdf/2311.14469
- https://en.wikipedia.org/wiki/Data_preprocessing
- https://www.academia.edu/122664728/Anomaly_Detection_In_Telecommunication_Networks_Leveraging_Novel_Big_Data_And_Machine_Learning_Techniques_For_Proactive_Fault_Management
- https://arxiv.org/abs/1609.08650
- https://www.sciencedirect.com/science/article/pii/S2351978918303858
- https://en.wikipedia.org/wiki/Morlet_wavelet
- https://arxiv.org/abs/2202.04212
- https://www.sciencedirect.com/science/article/pii/S2405896324002714