HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

PHẠM THANH LONG

NGHIÊN CỨU, THU THẬP PHÂN TÍCH SYSLOG TRONG HỆ THỐNG MẠNG BĂNG RỘNG CỐ ĐỊNH ĐỂ PHỤC VỤ CÁC CÔNG TÁC QUẢN TRỊ CỦA VNPT HÀ NỘI

ĐỀ ÁN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI - 2026

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

PHẠM THANH LONG

NGHIÊN CỨU, THU THẬP PHÂN TÍCH SYSLOG TRONG HỆ THỐNG MẠNG BĂNG RỘNG CỐ ĐỊNH ĐỂ PHỤC VỤ CÁC CÔNG TÁC QUẢN TRỊ CỦA VNPT HÀ NỘI

NGÀNH: HỆ THỐNG THÔNG TIN

MÃ SỐ: 8.48.01.04

ĐỀ ÁN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS.TS. LÊ NHẬT THĂNG

HÀ NỘI - 2026

LỜI CAM ĐOAN

Tôi cam đoan toàn bộ nội dung đề án do tôi tự học tập từ bài giảng các môn học của thầy cô, nghiên cứu trên Internet, sách báo, các tài liệu trong nước và nước ngoài có liên quan cùng với những kinh nghiệm thực tế có được tại đơn vị. Đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kết quả nêu trong đề án là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tác giả đề án



Phạm Thanh Long

LỜI CẢM ƠN

Em xin chân thành cảm ơn PGS-TS. Lê Nhật Thăng đã trực tiếp hướng dẫn, giúp đỡ, dạy bảo, động viên và tạo mọi điều kiện thuận lợi cho em trong quá trình từ xây dựng đề cương sơ bộ đến khi hoàn thiện đề án **"Nghiên cứu, thu thập phân tích Syslog trong hệ thống mạng băng rộng cố định để phục vụ các công tác quản trị của VNPT Hà Nội"**.

Xin gửi lời cảm ơn đến các thầy, cô giáo trong Khoa Đào tạo Sau Đại học, Học viện Công nghệ Bưu chính Viễn thông, những nhà giáo đã truyền dạy cho tôi kiến thức quý báu trong suốt những năm học vừa qua.

Xin được trân trọng cảm ơn các anh, chị đồng nghiệp trong VNPT Hà Nội đã cung cấp thông tin, tài liệu và hỗ trợ giúp đỡ tôi rất nhiều trong quá trình thực hiện đề án.

Do thời gian hoàn thành đề án có hạn cho nên những suy nghĩ cũng như sự thể hiện ý tưởng trong đề án không tránh khỏi những khiếm khuyết. Tôi rất mong được sự động viên và đóng góp ý kiến của các thầy cô giáo.

Xin trân trọng cảm ơn!

MỤC LỤC

LỜI MỞ ĐẦU 1
1. Lý do chọn đề tài. 1
2. Tổng quan về vấn đề nghiên cứu. 2
3. Mục đích nghiên cứu. 3
4. Đối tượng và phạm vi nghiên cứu. 3
5. Phương pháp nghiên cứu. 4
CHƯƠNG 1: CÔNG TÁC QUẢN TRỊ HỆ THỐNG MẠNG BĂNG RỘNG CỐ ĐỊNH VÀ NGHIÊN CỨU DỮ LIỆU SYSLOG HỆ THỐNG MẠNG CỦA VNPT HÀ NỘI 5
1.1. Mô hình hệ thống thiết bị, hệ thống giám sát, báo cáo, công tác vận hành, quản trị mạng băng rộng của VNPT Hà Nội. 5
1.1.1. Cấu trúc và phân lớp kiến trúc hệ thống mạng MAN-E. 5
1.1.2. Hiện trạng hệ thống giám sát, báo cáo đang quản lý và sử dụng. 6
1.1.3. Hạn chế cốt lõi và luận giải khoa học về sự cần thiết của đề án. 10
1.2. Nghiên cứu về dữ liệu cảnh báo trong Syslog của các thiết bị. 11
1.2.1.Tổng quan về Syslog và kiến trúc thu thập dữ liệu log. 11
1.2.2. Bản chất khoa học của dữ liệu nhật ký hệ thống (Syslog Data Cosmology). 15
1.2.3. Phân tích cấu trúc cú pháp dữ liệu cảnh báo đa chủng loại thiết bị (Multi-Vendor Syntax Analysis). 16
1.2.4. Mô hình hóa bài toán chuyển đổi dữ liệu phi cấu trúc và định hướng số hóa báo cáo. 17
1.3. Kết luận chương 1. 19
CHƯƠNG 2: CÁC ỨNG DỤNG CÔNG NGHỆ, HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU ĐỂ SỬ DỤNG PHÂN TÍCH LƯU TRỮ SYSLOG PHỤC VỤ CÔNG TÁC QUẢN TRỊ HỆ THỐNG 20
2.1. Các ứng dụng và các công nghệ áp dụng để phân tích Syslog hệ thống. 20
2.1.1. Cách tiếp cận dựa trên hệ sinh thái mã nguồn mở tập trung (ELK Stack / EFK Stack). 20
2.1.2. Cách tiếp cận dựa trên các giải pháp thương mại quy mô công nghiệp (Enterprise Solutions). 21
2.1.3. Xu hướng áp dụng kiến trúc phi quan hệ và xử lý luồng thời gian thực (Modern Stream Processing Pipeline). 22
2.1.4. Đánh giá và thảo luận so sánh luận chứng giải pháp (Comparative Analysis). 22
2.2. Hệ cơ sở dữ liệu sử dụng để lưu trữ và xử lý Syslog hệ thống. 23

2.2.1. Phân tích bài toán tổng thể và các ràng buộc đặc thù của hạ tầng mạng MAN-E VNPT Hà Nội 23
2.2.2. Kiến trúc tổng thể và cơ chế luồng dữ liệu của hệ thống đề xuất (System Pipeline Architecture) 24
2.2.3. Nghiên cứu sâu tầng bóc tách cấu trúc và kỹ thuật chuẩn hóa dữ liệu đa chủng loại (Normalization Layer). 26
2.2.4. Phân tích ngữ nghĩa sự cố và công cụ phân loại mức độ nghiêm trọng (Semantic & Fault Classification). 30
2.2.5. Mô hình hóa lớp đệm phân tán Apache Kafka chống bùng nổ dữ liệu (Fault-Tolerant Buffering Layer). 31
2.2.6. Thiết kế mô hình lưu trữ tối ưu trên hệ quản trị cơ sở dữ liệu phi quan hệ MongoDB (Storage Optimization Layer). 33
2.2.7. Nghiên cứu tầng hiển thị thời gian thực và kiến trúc Single Page Application (Visualization & User Experience Layer). 35
2.3. Kết luận chương 2. ..... ***38***
CHƯƠNG 3: TRIỂN KHAI KHAI THÁC DỮ LIỆU SYSLOG CHO CÔNG TÁC QUẢN TRỊ HỆ THỐNG VÀ BÁO CÁO CỦA VNPT HÀ NỘI **40**
3.1. Đánh giá chất lượng các đường truyền dẫn kết nối các thiết bị. ***40***
3.1.1. Kịch bản thu thập dữ liệu và phân tách lỗi truyền dẫn vật lý. 40
3.1.2. Kết quả bóc tách cấu trúc và phân tích ngữ nghĩa dữ liệu lỗi truyền dẫn. 41
3.1.3. Thống kê, tổng hợp và đánh giá chất lượng đường truyền toàn hệ thống. 42
3.1.4. Kết luận mục thực nghiệm 43
3.2. Đánh giá lưu lượng truyền tải giữa các thiết bị. ***43***
3.2.1. Kịch bản thực nghiệm và các ngưỡng cảnh báo lưu lượng. 44
3.2.2. Kết quả bóc tách cấu trúc và phân tích ngữ nghĩa log lưu lượng. 44
3.2.3. Thống kê xu hướng lưu lượng và đánh giá năng lực truyền tải đường trục. 46
3.2.4. Kết luận mục thực nghiệm. 47
3.3. Thống kê các bất thường về vượt ngưỡng và sử dụng. ***47***
3.3.1. Kịch bản thực nghiệm về các hành vi bất thường. 48
3.3.2. Kết quả bóc tách cấu trúc và phân tích ngữ nghĩa log bất thường. 48
3.3.3. Thống kê định lượng các bất thường hệ thống mạng toàn mạng. 49
3.3.4. Kết luận mục thực nghiệm. 51
3.4. Triển khai các báo phục vụ công tác quản trị hệ thống. ***51***
3.4.1. Giao diện bảng điều khiển trung tâm thời gian thực (Real-time Operations Dashboard). 51

3.4.2. Phân hệ chức năng báo cáo chất lượng mạng lưới và KPIs dịch vụ. 53
3.4.3. Đánh giá kiểm thử hiệu năng hiển thị và render giao diện (UI Performance Testing).....55
3.4.4. Kết luận mục thực nghiệm.....56
3.5. Kết luận chương 3. .....***56***
CHƯƠNG 4: KẾT LUẬN **58**
4.1. Kết quả đạt được của đề án. ***58***
4.2. Hướng nghiên cứu và phát triển tiếp theo của đề tài. ***59***
TÀI LIỆU THAM KHẢO **61**

CÁC TỪ VIẾT TẮT

Từ viết tắt	Từ tiếng Anh đầy đủ	Ý nghĩa / Giải nghĩa tiếng Việt
AAA	Authentication, Authorization, and Accounting	Hệ thống Xác thực, Ủy quyền và Ghi nhận tài khoản
ACL	Access Control List	Danh sách kiểm soát truy cập
API	Application Programming Interface	Giao diện lập trình ứng dụng
BRAS	Broadband Remote Access Server	Trạm kiểm soát truy cập biên mạng băng rộng
BSON	Binary JSON	Định dạng dữ liệu JSON nhị phân (dùng trong MongoDB)
CLI	Command Line Interface	Giao diện dòng lệnh
CPU	Central Processing Unit	Bộ vi xử lý trung tâm
CRC	Cyclic Redundancy Check	Kiểm tra chu kỳ dư (Lỗi bit truyền dẫn)
DDM	Digital Diagnostics Monitoring	Giám sát chẩn đoán kỹ thuật số (Module quang)
DOM	Document Object Model	Mô hình đối tượng tài liệu (Cấu trúc giao diện Web)
DoS	Denial of Service	Tấn công từ chối dịch vụ
EFK	Elasticsearch - Fluentd - Kibana	Bộ công cụ quản lý log mã nguồn mở (Biến thể của ELK)
ELK	Elasticsearch - Logstash - Kibana	Bộ công cụ thu thập, lưu trữ và trực quan hóa log
EPS	Events Per Second	Số lượng sự kiện/bản tin phát sinh trong một giây
FTTH	Fiber To The Home	Công nghệ mạng quang băng rộng đến tận nhà thuê bao
HMI	Human - Machine Interface	Giao diện tương tác người - máy
IDS	Intrusion Detection System	Hệ thống phát hiện xâm nhập
IEEE	Institute of Electrical and Electronics Engineers	Viện kỹ sư Điện và Điện tử
I/O	Input / Output	Quá trình nhập / xuất dữ liệu (Đọc ghi ổ đĩa)
IP	Internet Protocol	Giao thức Internet
JSON	JavaScript Object Notation	Định dạng trao đổi dữ liệu văn bản mỏng nhẹ

KPIs	Key Performance Indicators	Các chỉ số đánh giá hiệu năng cốt lõi
LACP	Link Aggregation Control Protocol	Giao thức điều khiển bó kênh / gộp đường truyền
MAN-E	Metropolitan Area Network - Ethernet	Mạng đô thị Ethernet băng rộng
MTTR	Mean Time To Repair	Thời gian khắc phục sự cố trung bình
NMS	Network Management System	Hệ thống quản trị và giám sát mạng
NoSQL	Not Only SQL	Hệ quản trị cơ sở dữ liệu phi quan hệ
NTP	Network Time Protocol	Giao thức đồng bộ thời gian mạng
OMC	Operations and Maintenance Center	Trung tâm Vận hành và Bảo dưỡng mạng lưới
RAM	Random Access Memory	Bộ nhớ truy cập ngẫu nhiên
RDBMS	Relational Database Management System	Hệ quản trị cơ sở dữ liệu quan hệ truyền thống
RCA	Root Cause Analysis	Phân tích nguyên nhân gốc rễ sự cố
RFC	Request for Comments	Bộ các tiêu chuẩn và tài liệu định hướng Internet
SFP	Small Form-factor Pluggable	Module quang thu phát kích thước nhỏ
SLA	Service Level Agreement	Cam kết chất lượng dịch vụ
SNMP	Simple Network Management System	Giao thức quản lý mạng đơn giản
SPA	Single Page Application	Kiến trúc ứng dụng Web đơn trang
TTL	Time To Live	Cơ chế thiết lập thời gian sống / tự động xóa dữ liệu
UDP	User Datagram Protocol	Giao thức truyền tải hướng không kết nối
UPE	User Edge / Underlay Provider Edge	Thiết bị router biên phân cấp truy nhập
VNPT	Vietnam Posts and Telecommunications Group	Tập đoàn Bưu chính Viễn thông Việt Nam
XFP	10 Gigabit Small Form-factor Pluggable	Module quang thu phát tốc độ 10 Gbps

DANH MỤC HÌNH ẢNH

Hình 1. 1 Sơ đồ phân lớp mạng MANE....6
Hình 1. 2 Các chức năng chính của hệ thống CTS....8
Hình 1. 3 Hệ thống giám sát PRTG....8
Hình 1. 4 Kiến trúc giao thức và mô hình truyền tải chuỗi thời gian Syslog....12
Hình 1. 5 Cấu trúc gói tin RFC 3164 và RFC 5424....13
Hình 2. 1 Mô hình kiến trúc ELK (Elasticsearch - Logstash - Kibana)....20
Hình 2. 2 Sơ đồ kiến trúc phân lớp và luồng dữ liệu hệ thống quản lý Syslog tập trung....25
Hình 2. 3 Cơ chế điều phối xử lý dữ liệu song song và chịu lỗi sử dụng Kafka Consumer Group....33
Hình 2. 4 Cơ chế tối ưu hóa hiệu năng render giao diện thời gian thực trên Angular....36
Hình 3. 1 Bảng giám sát trung tâm....52
Hình 3. 2 Tổng hợp các log bất thường của các hệ thống và mức độ....53
Hình 3. 3 Thống kê các vệ tinh có cảnh báo log sự cố nhiều nhất trong ngày....54
Hình 3. 4 Thống kê các lỗi sự số diễn ra nhiều nhất trong ngày....54

DANH MỤC CÁC BẢNG

Bảng 1. 1 So sánh đặc tính kỹ thuật các công cụ quản trị hiện hữu tại VNPT Hà Nội....10
Bảng 1. 2 Phân cấp toán học và ý nghĩa hệ thống của không gian thuộc tính Severity....15
Bảng 1. 3 Ma trận phân tích ngữ nghĩa dữ liệu Syslog và định hướng thuộc tính số hóa....17
Bảng 2. 1 Ma trận so sánh các đặc tính kỹ thuật của các nền tảng phân tích Syslog22
Bảng 2. 2 Các chỉ số chỉ tiêu kỹ thuật thiết kế của hệ thống quản lý Syslog đề xuất....37
Bảng 3. 1 Thống kê tần suất xuất hiện lỗi truyền dẫn vật lý theo dòng thiết bị....42
Bảng 3. 2 Thống kê thời gian và mức độ nghẽn mạch trên các tuyến đường trục mạng MAN-E....46
Bảng 3. 3 Thống kê tần suất xuất hiện các bất thường vượt ngưỡng phần cứng và sử dụng....50
Bảng 3. 4 Kết quả thực nghiệm đo lường hiệu năng hiển thị của giao diện Angular....55

LỜI MỞ ĐẦU

1. Lý do chọn đề tài.

Tập đoàn Bưu chính Viễn thông Việt Nam (VNPT) với vai trò là một Tập đoàn Viễn thông, Công nghệ thông tin hàng đầu Việt Nam trong cuộc cách mạng công nghệ 4.0 và công cuộc chuyển đổi số. Ngay từ đầu năm 2017, VNPT đã sớm xây dựng chiến lược VNPT4.0, trong đó định vị VNPT phải trở thành nhà cung cấp dịch vụ số hàng đầu tại Việt Nam vào năm 2025 và trở thành Trung tâm dịch vụ Số (Digital Hub) của Châu Á vào năm 2030, đồng thời xác định phải giữ vai trò chủ đạo trong cách mạng số tại Việt Nam.

VNPT Hà Nội cũng đã sớm có nhiều định hướng và phát triển nhiều chương trình số hóa, hòa chung vào cuộc cách mạng chuyển đổi số cùng VNPT.

Hệ thống mạng băng rộng cố định của VNPT Hà Nội dùng để cung cấp các dịch vụ Internet, truyền hình MyTV, điện thoại, truyền tải cho mạng di động... do có nhiều chủng loại và hãng thiết bị khác nhau nên kèm với đó là rất nhiều các phần mềm, chương trình giám sát, cảnh báo, theo dõi, báo cáo, quản trị khác nhau. Do có nhiều phần mềm chương trình, nên người khai thác cần phải thực hiện tác vụ và theo dõi ở nhiều nơi, cùng với đó công tác báo cáo cần tổng hợp xử lý ở nhiều chương trình khác nhau.

Một số hiện trạng và nhu cầu:

- Hiện trạng hệ thống giám sát mạng băng rộng hiện tại của VNPT Hà Nội gặp khó khăn trong việc khai thác, phân tích và tổng hợp số liệu do hạn chế về tính định kỳ và tùy biến theo tiêu chí giám sát.

- Nhu cầu thực tế cần có sự giám sát và cảnh báo các sự cố bất thường trên thiết bị theo thời gian thực và tổng hợp số liệu để phân tích lưu lượng, san tải khi có bất thường về lưu lượng trên mạng lưới.

- Việc nghiên cứu và thu thập các số liệu của hệ thống giúp hiểu rõ hơn về hiệu suất hoạt động của mạng, các yếu tố ảnh hưởng đến chất lượng dịch vụ, cũng như những vấn đề cần được giải quyết và cải thiện. Thông qua việc tổng hợp số hóa

dữ liệu, chúng ta có thể phân tích và đánh giá các thông số kỹ thuật, tốc độ truy cập, độ ổn định của mạng, và mối quan hệ giữa các yếu tố này với trải nghiệm người dùng.

Những điểm này cho thấy sự cần thiết phải phát triển hệ thống giám sát, phân tích và thống kê báo cáo để đáp ứng nhu cầu quản lý và giám sát hiệu quả hơn mà còn hướng tới việc cải thiện dịch vụ, tăng cường sự hài lòng của khách hàng, và đóng góp vào việc phát triển ngành viễn thông của Việt Nam. Đồng thời, nó cũng tạo cơ hội cho các nhà nghiên cứu và các đơn vị quản lý mạng để đưa ra những đề xuất và chiến lược phát triển bền vững cho mạng băng rộng trong tương lai.

2. Tổng quan về vấn đề nghiên cứu.

Xuất phát từ như cầu thực tế vận hành và quản trị mạng băng rộng của VNPT Hà Nội. Đề tài này tập trung vào xây dựng thu thập số liệu và phân tích phát hiện các cảnh báo có trong các bản tin Syslog từ các thiết bị trong hệ thống và tích hợp tập trung trên một website quản lý chung. Từ đó phát triển website số hóa xây dựng báo cáo tổng hợp từ các số liệu thu thập được để phục vụ công tác theo dõi, cảnh báo, đánh giá, phục vụ xây dựng các kế hoạch quản trị, tối ưu, phát triển mạng lưới, nâng cao chất lượng dịch vụ.

Theo đó, việc nghiên cứu tìm hiểu hiện trạng mô hình hệ thống và các chương trình phần mềm theo dõi quản lý giám sát các hệ thống thiết bị, các chương trình đang quản lý quản trị đối hệ thống, từ đó xây dựng bảng danh mục lỗi cần quan tâm, thực hiện kiểm tra xử lý của từng loại thiết bị. Thu thập và phân tích các cảnh báo dựa trên bản tin Syslog của từng loại thiết bị trong hê thống để phát hiện các các cảnh báo lỗi bất thường, các cảnh báo cần xử lý ngay, đây là những vấn đề nghiên cứu chính mà đề tài cần thực hiện nhằm phục vụ hiệu quả cho công tác quản trị mạng băng rộng của VNPT Hà Nội.

Các vấn đề lưu ý khi thực hiện:

- Khối lượng dữ liệu lớn: với hệ thống mạng lớn nhiều thiết bị, sẽ tạo ra một lượng lớn dữ liệu Syslog, đòi hỏi các giải pháp lưu trữ và xử lý hiệu quả.

- Đa dạng định dạng: nhiều chủng loại thiết bị khác nhau có thể sử dụng các định dạng Syslog khác nhau, gây khó khăn cho việc chuẩn hóa dữ liệu.

- Độ phức tạp của dữ liệu: Syslog thường có cấu trúc không đồng nhất và chứa nhiều thông tin nhiễu, đòi hỏi các kỹ thuật phân tích phức tạp.

Tóm lại phân tích dữ liệu Syslog để phát hiện các cảnh báo bất thường là một lĩnh vực nghiên cứu và ứng dụng quan trọng trong lĩnh vực phòng ngừa và phát hiện sự cố nhanh chóng. Bằng cách khai thác thông tin có trong các thông báo Syslog, có thể cải thiện đáng kể khả năng giám sát bảo vệ hệ thống, nâng cao hiệu năng và giảm thiểu rủi ro.

3. Mục đích nghiên cứu.

Phân tích và đưa ra bảng danh mục các lỗi có thể nhận diện từ Syslog cần phải xử lý và các hành động kịch bản xử lý liên quan. Từ đó xây dựng mô hình tham chiếu, tìm kiếm so sánh các cảnh báo thu được từ Syslog với bảng danh mục lỗi, và tự động triển khai các hành động xử lý liên quan như cảnh báo (qua E-mail, SMS...). lưu trữ để phục vụ công tác theo dõi báo cáo. Một số mục tiêu cụ thể:

Thu thập, lưu trữ dữ liệu:

- Tiền xử lý dữ liệu: tiến hành các hoạt động tiền xử lý như lọc, chuẩn hóa, và trích xuất các thông tin cần thiết từ dữ liệu Syslog thô.

- Phân tích dữ liệu: xây dựng các mô hình tìm kiếm, so sánh kiểm tra để phát hiện các lỗi cần có hành động xử lý.

Thực nghiệm và đánh giá kết quả.

4. Đối tượng và phạm vi nghiên cứu.

Đối tượng nghiên cứu: hệ thống thiết bị của mạng băng rộng, hệ thống giám sát, và hệ thống báo cáo, các bản tin Syslog của các hệ thống.

Phạm vi nghiên cứu: hệ thống thiết bị, giám sát, báo cáo của mạng băng rộng trong VNPT Hà Nội.

- Các công nghệ, hệ quản trị cơ sở dữ liệu sử dụng để xây dựng các module kết nối, thu thập dữ liệu và xây dựng chương trình quản lý, bao gồm Angular, JWT, Restful API, SNMP, SYSLOG, SignalR, và MongoDB.

5. Phương pháp nghiên cứu.

Phương pháp nghiên cứu được sử dụng là kết hợp giữa các kiến thức lý thuyết và việc nghiên cứu, khảo sát phân tích của các hệ thống giám sát, báo cáo trong mạng băng rộng của VNPT Hà Nội.

Lý thuyết: nghiên cứu lý thuyết hệ thống thông tin và cơ sở dữ liệu. Các công nghệ, hệ quản trị cơ sở dữ liệu sử dụng. Nghiên cứu các chủng loại, cấu trúc Mib, cấu trúc Syslog của các thiết bị trong mạng băng rộng của VNPT Hà Nội.

Thực nghiệm: tổ chức cơ sở dữ liệu về thông tin số tại đơn vị, nghiên cứu thực tế về dữ liệu giám sát, cảnh báo của các hệ thống, các trang báo cáo đang theo dõi vận hành khai thác. Số hóa dữ liệu cảnh báo, thiết bị, phân tích dữ liệu để áp dụng vào sản xuất tại đơn vị.

CHƯƠNG 1: CÔNG TÁC QUẢN TRỊ HỆ THỐNG MẠNG BĂNG RỘNG CỐ ĐỊNH VÀ NGHIÊN CỨU DỮ LIỆU SYSLOG HỆ THỐNG MẠNG CỦA VNPT HÀ NỘI

1.1. Mô hình hệ thống thiết bị, hệ thống giám sát, báo cáo, công tác vận hành, quản trị mạng băng rộng của VNPT Hà Nội.

1.1.1. Cấu trúc và phân lớp kiến trúc hệ thống mạng MAN-E.

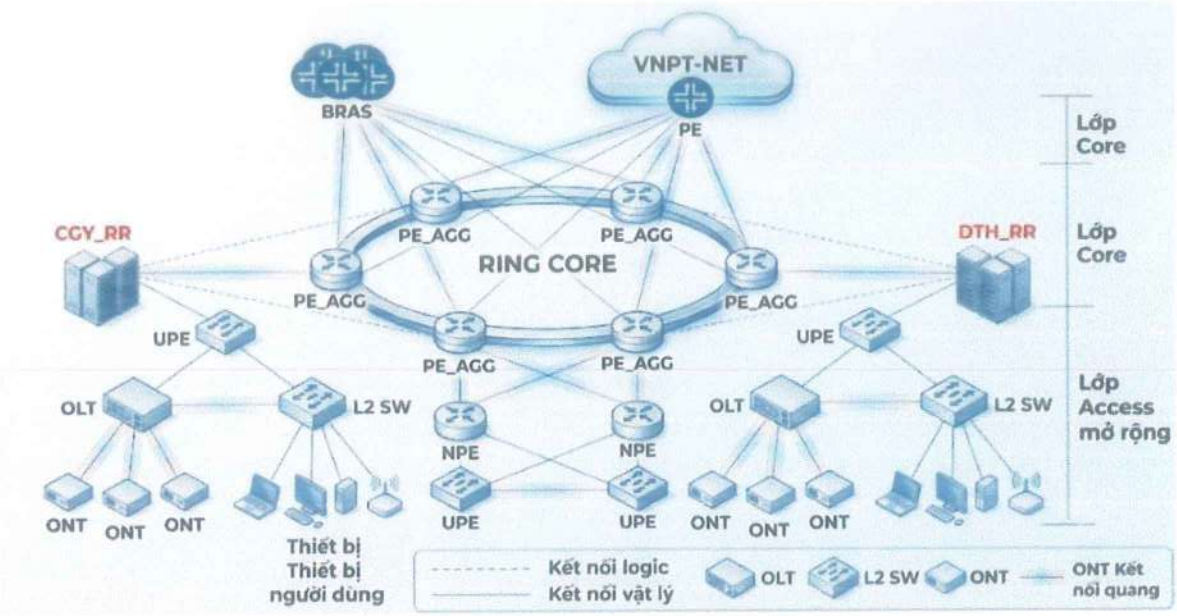
Cấu trúc hệ thống:

Hệ thống mạng MAN-E tổ chức gồm các thành phần mạng lõi và mạng truy nhập, bao gồm các router kết nối với nhau theo vòng ring, hình cây, đảm bảo khả năng dự phòng tuyến khi đứt kết nối.

Mạng lõi (ring core): Lớp này bao gồm các thiết bị biên dịch vụ cao cấp dòng PE (Provider Edge) và PE-AGG (Provider Edge Aggregation) lắp đặt tại các trung tâm lớn. Để tối ưu hóa hiệu năng routing và hạn chế độ trễ truyền dẫn do số lượng bước nhảy (hop), số lượng thiết bị trong một vòng Ring Core được thiết kế chỉ từ 5 đến 7 điểm trong một Ring. Vị trí lắp đặt các Core tại điểm thu gom truyền dẫn và dung lượng trung chuyển qua đó cao. Các thiết bị này được kết nối ring với nhau bằng một đôi hay nhiều đôi cáp quang trực tiếp, sử dụng giao diện kết nối Ethernet cổng 100Gbps.

Mạng truy nhập (Ring Access): Bao gồm các UPE lắp đặt tại các trạm Viễn thông, kết nối với nhau và kết nối tới mạng lõi bằng một hoặc nhiều đôi sợi quang trực tiếp. Tùy theo điều kiện, mạng truy nhập có thể sử dụng kết nối dạng hình sao, ring và trong một ring tối đa từ 4 - 6 thiết bị UPE, hoặc đấu nối tiếp nhau và đấu nối tiếp tối đa từ 4 - 6 thiết bị UPE, vị trí lắp đặt các UPE truy nhập thường đặt tại các điểm thuận tiện cho việc thu gom truyền dẫn kết nối đến các thiết bị truy nhập như Switch L2, OLT GPON.

Sơ đồ mô hình hệ thống:



Hình 1. 1 Sơ đồ phân lớp mạng MANE

Mạng MAN-E của VNPT Hà Nội sử dụng để tập trung lưu lượng (IP/MPLS Aggregation over Ethernet): đảm bảo tập trung lưu lượng từ các thiết bị truy nhập như OLT GPON, SW L2… tới BRAS, PE/NET.

Hạ tầng mạng lưới được đồng bộ hóa dựa trên tổ hợp thiết bị của nhiều hãng công nghệ lớn trên thế giới, cụ thể:

Hạ tầng mạng lõi và hội tụ MAN-E: được cấu thành từ các hệ thống định tuyến dung lượng lớn như Cisco ASR-9010, ASR-9922 và Juniper MX960. Các nút mạng này thiết lập cấu trúc liên kết mạch vòng (Ring topology) tốc độ 100Gbps để đảm bảo khả năng ứng cứu, tự động chuyển tuyến bảo vệ (Protection Switching) khi xảy ra sự cố đứt cáp quang trục.

Hệ thống cổng kiểm soát đa dịch vụ (BRAS - Broadband Remote Access Server): sử dụng dòng tổng đài trung tâm Juniper MX để thực hiện chức năng xác thực tài khoản thuê bao (qua giao thức PPPoE), cấp phát dải địa chỉ IP (IP Pool Management) và thực thi các chính sách băng thông gói cước cho từng khách hàng.

Hạ tầng mạng truy nhập PON: phân bổ rộng khắp các xã/phường bằng các thiết bị tổng đài quang OLT của nhiều nhà sản xuất như Huawei (dòng MA5608T/MA5800), ZTE (dòng C300/C600) và Nokia/Alcatel-Lucent. Sự đa dạng

chủng loại thiết bị (Multi-Vendor Environment) tạo nên tính linh hoạt đa dạng thiết bị nhưng lại là thách thức lớn cho công tác quản trị tập trung [1], [2].

1.1.2. Hiện trạng hệ thống giám sát, báo cáo đang quản lý và sử dụng.

1.1.2.1. Đánh giá chi tiết năng lực và hạn chế của các hệ thống giám sát hiện hữu.

Hiện nay, công tác vận hành và giám sát kỹ thuật mạng lưới tại VNPT Hà Nội đang phụ thuộc vào hai hệ thống chính [2]:

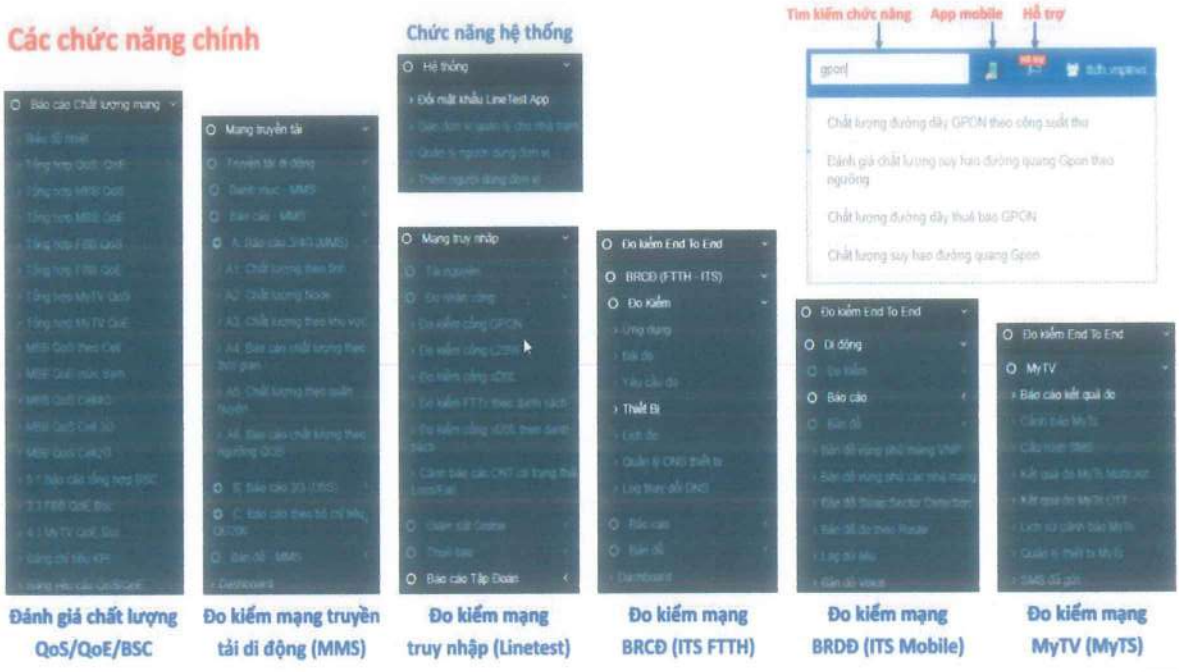
a) Hệ thống đo kiểm tập trung CTS.

Hệ thống CTS được phát triển nhằm mục đích theo dõi và đo kiểm chất lượng dịch vụ (QoS/QoE) đầu cuối của các dịch vụ Viễn thông. CTS sở hữu các đặc điểm vận hành như sau:

Chức năng đạt được: Hệ thống thực hiện thu thập dữ liệu đo kiểm định kỳ, kết xuất các biểu đồ trực quan, tích hợp bản đồ số Google Maps để hiển thị trạng thái chất lượng mạng theo từng khu vực địa lý hoặc theo từng thiết bị mạng cụ thể. Giao diện Web và ứng dụng di động cho phép điều hành trực quan hóa các tham số đường truyền.

Hạn chế cốt lõi: CTS tập trung vào tham số dịch vụ đầu cuối (end-to-end), hoàn toàn không có khả năng đi sâu vào phân tích các sự cố phần cứng nội tại của thiết bị mạng hoặc các cảnh báo logic tức thời từ hệ điều hành router phát ra.

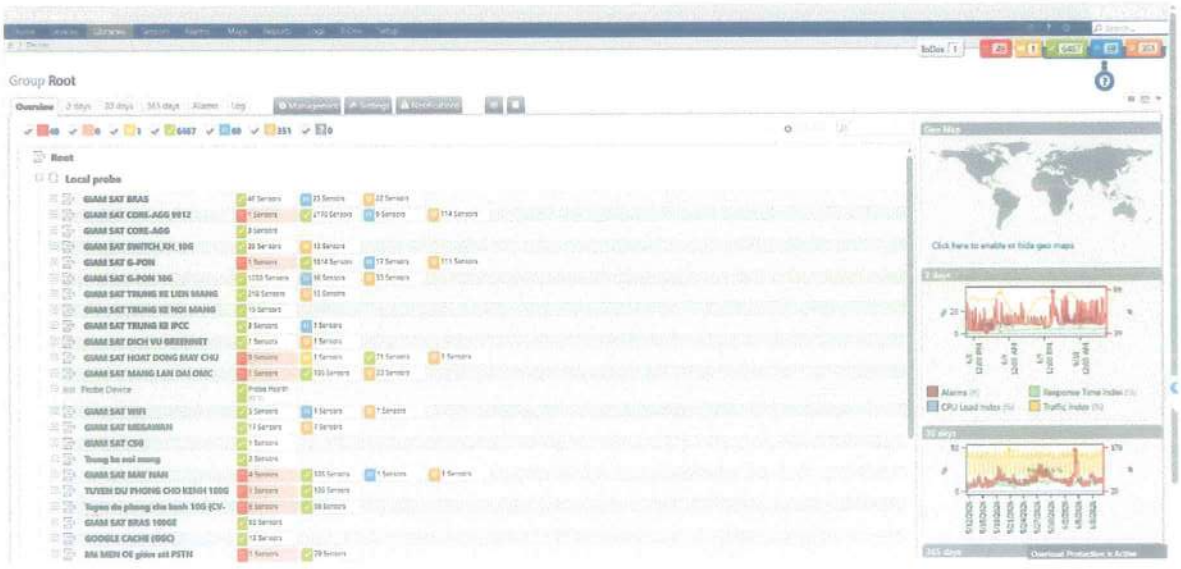
Các chức năng chính của hệ thống:



Hình 1. 2 Các chức năng chính của hệ thống CTS

b) Hệ thống giám sát PRTG.

PRTG hoạt động dựa trên cơ chế kết nối định kỳ qua giao thức SNMP (Simple Network Management Protocol) để lấy các tham số lưu lượng (traffic) cổng với các chu kỳ cấu hình động từ 30 giây đến 256 giây tùy vùng thiết bị.



Hình 1. 3 Hệ thống giám sát PRTG

Chức năng đạt được: PRTG hỗ trợ lưu trữ dữ liệu lưu lượng vào cơ sở dữ liệu nội bộ để vẽ biểu đồ nghẽn mạng, kết nối với cổng SMS Gateway và Mail Server

nhằm gửi thông báo tự động cho người quản trị khi xảy ra sự cố cổng Up/Down hoặc lưu lượng tăng vượt mức giới hạn.

- Hệ thống PRTG kết nối với hệ thống SMS gateway phục vụ cho việc gửi SMS cảnh báo khi port Down, port Up, lưu lượng bất thường.

- Hệ thống PRTG kết nối với máy chủ Mail server để phục vụ cho việc gửi mail định kỳ hoặc mail sự cố.

- Ứng dụng Phần mềm Network Monitor PRTG ngoài khả năng hỗ trợ View giám sát qua Web http & https còn hỗ trợ các add on dưới đây:

+ Ứng dụng giao diện windows (Windows GUI): là một ứng dụng client/server cho phép thao tác add/edit groups, servers, devices và sensors kèm các alarm, alert. Rất thích hợp cho các đơn vị trực ca quan sát trực quan.

+ Ứng dụng cho iPhone (iPRTG) là ứng dụng được dùng cho iPhone, iPad or iPod.

+ Ứng dụng cho Android (PRTGdroid) là ứng dụng được dùng cho Android smartphones.

Các hạn chế: Cơ chế "Pull" (đo quét định kỳ) của SNMP trên PRTG tạo ra độ trễ thông tin lớn. Nếu sự cố chập chờn cổng (port flapping) xảy ra và tự phục hồi giữa hai chu kỳ quét, PRTG hoàn toàn bỏ sót sự cố này. Đồng thời, PRTG không thể thu thập các lỗi nội tại phức tạp như lỗi bộ nguồn, lỗi nhiệt độ card, hay các hành vi truy cập trái phép.

1.1.2.2. Hiện trạng hệ thống báo cáo quản lý tài nguyên và nhu cầu cấp thiết.

Hệ thống quản lý tài nguyên mạng VNPT-RIMS (do VNPT Net xây dựng) cùng với trang báo cáo tập trung của VNPT Hà Nội hiện tại chỉ cung cấp các báo cáo tĩnh bao gồm: thống kê số lượng thiết bị theo chủng loại, cấu hình danh mục các card/cổng/module quang đang online, và tổng hợp cấu trúc hướng kết nối vật lý lớp MAN-E.

Đối với hệ thống thiết bị BRAS Juniper MX960 - phân hệ cốt lõi quản lý phiên thuê bao, các hệ thống báo cáo hiện hữu hoàn toàn chưa có phân hệ theo dõi các thông

tin online:

- Tổng số lượng phiên IP Online dao động theo thời gian thực trong ngày.

- Tỷ lệ phần trăm sử dụng địa chỉ IP theo từng dải (Pool IP), dẫn đến việc không thể dự báo sớm tình trạng cạn kiệt IP cấp phát cho thuê bao FTTH mới.

Bảng 1. 1 So sánh đặc tính kỹ thuật các công cụ quản trị hiện hữu tại VNPT Hà Nội

Tiêu chí phân tích	Hệ thống giám sát hãng (CTS,...)	Công cụ PRTG (Dựa trên SNMP)	Hệ thống quản lý RIMS
Cơ chế thu thập dữ liệu	Độc quyền theo nhà sản xuất (Vendor-specific)	Đo quét định kỳ (Polling) theo chu kỳ thời gian (Ví dụ: 5 phút/lần)	Truy vấn cơ sở dữ liệu quan hệ (SQL Queries)
Độ trễ thông tin sự cố	Trung bình (Tùy thuộc bản tin bẫy trap)	Cao (Bị giới hạn bởi chu kỳ của tiến trình quét Polling)	Rất cao (Chỉ cập nhật khi phát sinh tác vụ nghiệp vụ)
Khả năng phân tích log thô	Không hỗ trợ	Không hỗ trợ (Chỉ hiển thị biểu đồ thông số tải)	Không hỗ trợ
Tính tương thích hệ thống	Đóng (Chỉ chạy trên thiết bị của chính hãng đó)	Khá tốt (Hỗ trợ các thiết bị có mở tính năng SNMP MIB)	Đóng (Chỉ phục vụ tầng ứng dụng quản lý thuê bao)

1.1.3. Hạn chế cốt lõi và luận giải khoa học về sự cần thiết của đề án.

Qua quá trình vận hành thực tế hạ tầng viễn thông đa chủng loại tại VNPT Hà Nội, công tác điều hành kỹ thuật đang đối mặt với những điểm nghẽn mang tính hệ thống sau:

Tính bị động trong phát hiện sự cố (Reactive Management): Do công cụ SNMP (PRTG) hoạt động theo cơ chế đo quét định kỳ, nếu một cổng quang hoặc một thiết bị OLT xảy ra lỗi chập chờn (Flapping Link) nằm giữa chu kỳ quét, hệ thống sẽ bỏ sót. Kỹ thuật viên thường chỉ biết đến sự cố sau khi khách hàng phát sinh khiếu nại lên tổng đài.

Thiếu khả năng bóc tách lỗi logic sâu: Giao thức SNMP chỉ cung cấp các số liệu thống kê tổng quát (như cổng Up/Down hoặc % băng thông). SNMP không thể chỉ ra nguyên nhân cốt lõi của lỗi phần mềm hoặc các xung đột logic bên trong (ví dụ: lỗi xác thực PPPoE hàng loạt do treo card, lỗi sụt công suất quang của từng phần

tử ONT đơn lẻ trong mạng GPON).

Hiện tượng cô lập thông tin log đa nguồn: Mỗi thiết bị mạng viễn thông đều tự động sản sinh ra các dòng nhật ký hệ thống (Syslog) chứa đựng chi tiết mọi biến động trạng thái thời gian thực. Tuy nhiên, do các dòng log này tồn tại dưới dạng văn bản tự do phi cấu trúc (Unstructured Text) và phân tán trên hàng nghìn thiết bị khác nhau, VNPT Hà Nội chưa có một công cụ nào có khả năng thu thập, xử lý ngôn ngữ và số hóa các thông điệp này ở quy mô lớn.

Luận giải khoa học cho đề án chỉ ra rằng, để giải quyết triệt để các hạn chế nghẽn thông tin hiện tại, việc nghiên cứu xây dựng một hệ thống chủ động lắng nghe luồng dữ liệu Syslog thời gian thực, ứng dụng các thuật toán bóc tách cú pháp tự động hóa để chuẩn hóa dữ liệu log phi cấu trúc từ môi trường đa chủng loại thiết bị thành các báo cáo số hóa trực quan là vô cùng cấp thiết. Đề tài không chỉ có ý nghĩa tối ưu hóa chỉ số thời gian khắc phục sự cố (MTTR) cho VNPT Hà Nội mà còn đóng góp giải pháp mô hình hóa dữ liệu nhật ký hệ thống cho ngành Hệ thống thông tin ứng dụng trong viễn thông. Đặt trong bối cảnh hạ tầng viễn thông thế giới đang dịch chuyển mạnh mẽ sang kỷ nguyên số và tự động hóa vận hành, việc giám sát chủ động chất lượng mạng lưới này càng đóng vai trò sống còn đối với các nhà khai thác dịch vụ viễn thông (ISPs). Các nghiên cứu quốc tế đã chứng minh một cách tường minh rằng, việc tối ưu hóa quản lý băng thông và giảm thiểu thời gian gián đoạn kết nối của khách hàng phụ thuộc chặt chẽ vào năng lực thu thập, phân tích dữ liệu telemetry kết hợp với log thiết bị theo thời gian thực [1]. Tại Việt Nam, xu hướng chuyển đổi số toàn diện các quy trình vận hành khai thác mạng băng rộng cố định như hướng tiếp cận của đề tài cũng đang trở thành mục tiêu ưu tiên cốt lõi, không chỉ nhằm nâng cao năng lực cạnh tranh dịch vụ mà còn đảm bảo an toàn thông tin tối đa cho hạ tầng số quốc gia [2].

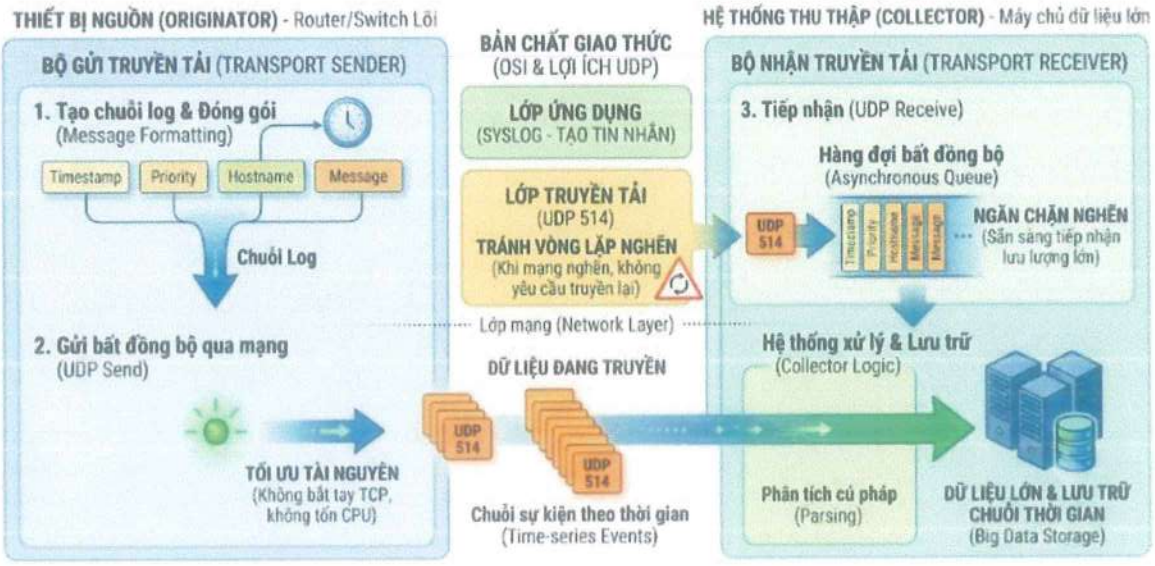
1.2. Nghiên cứu về dữ liệu cảnh báo trong Syslog của các thiết bị.

1.2.1. Tổng quan về Syslog và kiến trúc thu thập dữ liệu log.

1.2.1.1. Bản chất giao thức và mô hình truyền tải chuỗi thời gian Syslog.

Giao thức nhật ký hệ thống (Syslog) là một kiến trúc chuẩn được thiết kế để

truyền tải các thông điệp cảnh báo trạng thái, lỗi vận hành và nhật ký hành vi từ các thiết bị phần cứng hoặc phần mềm ứng dụng về máy chủ quản lý tập trung. Về mặt kiến trúc mạng lớp, Syslog hoạt động chủ yếu ở tầng ứng dụng (Application Layer) trong mô hình OSI và sử dụng giao thức UDP (User Datagram Protocol) tại tầng truyền tải (Transport Layer) với cổng dịch vụ mặc định là 514 [10].



Hình 1. 4 Kiến trúc giao thức và mô hình truyền tải chuỗi thời gian Syslog

Việc lựa chọn giao thức UDP hướng không kết nối (Connectionless) mang một ý nghĩa khoa học quan trọng đối với hệ thống thông tin viễn thông:

- Tối ưu hóa tài nguyên thiết bị phát (Syslog Generator): Thiết bị mạng lõi (Router/Switch Core) khi xảy ra sự cố nghẽn mạch hoặc quá tải CPU cần phải giải phóng các bản tin cảnh báo ngay lập tức mà không được phép tốn tài nguyên cho quá trình bắt tay ba bước (Three-way Handshake) của TCP.

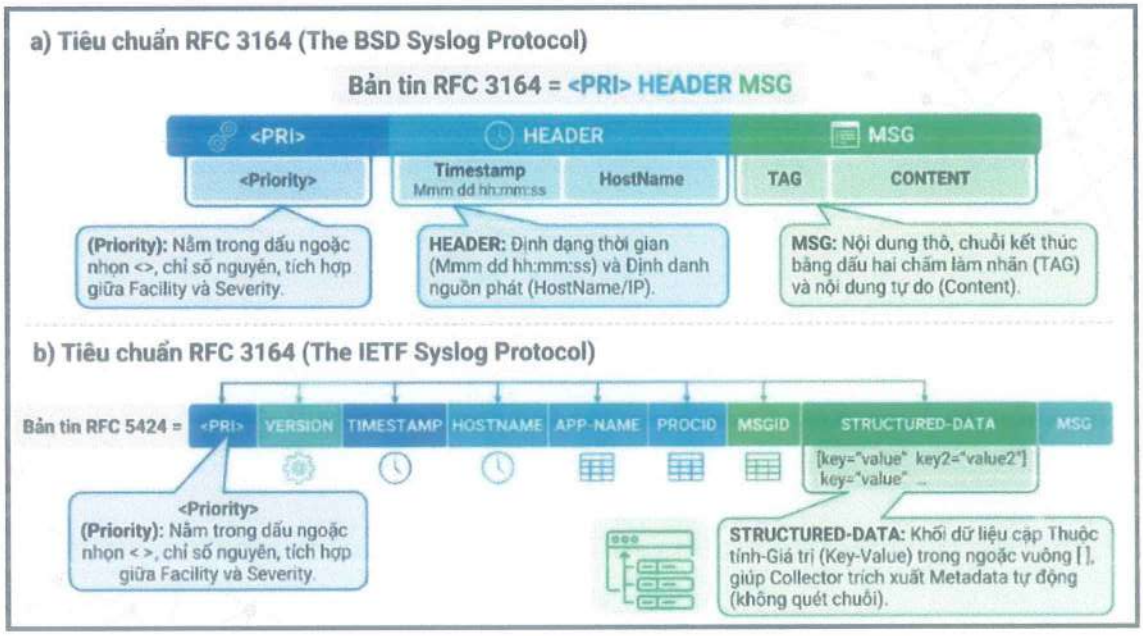
- Ngăn chặn vòng lặp nghẽn (Congestion Loop): Nếu sử dụng TCP, khi đường truyền cáp quang bị nghẽn, các gói tin Syslog báo lỗi nghẽn lại tiếp tục yêu cầu truyền lại (Retransmission), dẫn đến hiện tượng nghẽn chồng nghẽn, có nguy cơ làm sập hoàn toàn tiến trình định tuyến của thiết bị.

Tuy nhiên, đặc tính của UDP là không đảm bảo độ tin cậy (Unreliable Delivery) và không kiểm soát luồng (Flow Control). Do đó, thách thức này đòi hỏi kiến trúc hệ thống thông tin tại máy chủ thu thập (Syslog Collector) phải được thiết

kế theo mô hình hàng đợi bất đồng bộ có năng lực giải tỏa dữ liệu cực nhanh, một bài toán sẽ được giải quyết bằng công nghệ lưu trữ dữ liệu lớn ở các mục sau.

1.2.1.2. Chuẩn hóa cấu trúc gói tin Syslog theo RFC 3164 và RFC 5424.

Để hệ thống thông tin có thể xử lý tự động dòng dữ liệu log từ môi trường đa chủng loại thiết bị (Heterogeneous Network Domain), cần phải mô hình hóa cấu trúc gói tin theo hai tiêu chuẩn quốc tế cốt lõi [10]:



Hình 1. 5 Cấu trúc gói tin RFC 3164 và RFC 5424

a) Tiêu chuẩn RFC 3164 (The BSD Syslog Protocol)

RFC 3164 là định dạng truyền thống, mang tính chất của dữ liệu bán cấu trúc, không có ranh giới phân tách thuộc tính tường minh. Một bản tin RFC 3164 chuẩn bao gồm ba thành phần logic liên tiếp:

Bản tin RFC 3164 = <PRI> HEADER MSG

- PRI (Priority): Nằm trong dấu ngoặc nhọn <>, là một chỉ số nguyên thể hiện mức độ ưu tiên của thông điệp, được tính toán dựa trên thuật toán tích hợp giữa nguồn phát sinh (Facility) và độ nghiêm trọng (Severity).
- HEADER: Gồm hai trường dữ liệu là Thời gian phát sinh (Timestamp định dạng Mmm dd hh:mm:ss) và Định danh nguồn phát (HostName hoặc IP Address).

- MSG: Phân đoạn nội dung thô, bao gồm một chuỗi ký tự kết thúc bằng dấu hai chấm làm nhãn định danh (TAG) và phần nội dung thông điệp tự do (Content).

b) Tiêu chuẩn RFC 5424 (The IETF Syslog Protocol)

Nhằm khắc phục sự thiếu nhất quán và mơ hồ của RFC 3164, tổ chức IETF đã ban hành RFC 5424, chuẩn hóa dữ liệu log hướng đối tượng máy tính phân tích tự động. Kiến trúc gói tin được định nghĩa nghiêm ngặt như sau[10]:

Bản tin RFC 5424 = <PRI> VERSION TIMESTAMP HOSTNAME APP-NAME PROCID MSGID STRUCTURED-DATA MSG

Sự cải tiến vượt bậc của RFC 5424 nằm ở trường STRUCTURED-DATA. Trường này cho phép chèn các khối dữ liệu định dạng cặp Thuộc tính - Giá trị (Key-Value) trong dấu ngoặc vuông []. Điều này giúp hệ thống thông tin tại máy chủ Collector trích xuất được ngay các siêu dữ liệu (Metadata) mà không cần phải thực hiện các phép toán quét chuỗi phức tạp.

1.2.1.3. Mô hình toán học toán tính toán chỉ số ưu tiên (PRI) và phân cấp Severity.

Chỉ số PRI xuất hiện ở đầu mỗi gói tin Syslog là kết quả của hàm tuyến tính ràng buộc giữa hai không gian trạng thái: Không gian nguồn phát dịch vụ (Facility $\in$[0, 23]) và Không gian mức độ nghiêm trọng sự cố (Severity $\in$ [1, 7]).

Công thức toán học định biên PRI:

$$PRI = (Facility \times 8) + Severity$$

Từ công thức toán học trên, bộ giải mã cú pháp (Parser Engine) tại máy chủ thu thập có thể thực hiện các phép toán chia lấy nguyên và chia lấy dư để cô lập ngược lại hai chỉ số thành phần:

$$Facility = [\ PRI/8\ ]$$

$$Severity = PRI\ (mod\ 8)$$

Bảng 1. 2 Phân cấp toán học và ý nghĩa hệ thống của không gian thuộc tính Severity

Mã số	Tên mức độ (Severity Level)	Định nghĩa học thuật hệ thống
0	Emergency	Hệ thống không thể sử dụng (sập nút mạng Core)
1	Alert	Sự cố nghiêm trọng cần can thiệp vật lý ngay lập tức (Đứt mạch vòng quang).
2	Critical	Tình trạng lỗi phần cứng nghiêm trọng (Hỏng card nguồn, quá nhiệt).
3	Error	Điều kiện lỗi chức năng phần mềm logic (Lỗi xác thực, treo session).
4	Warning	Các dấu hiệu cảnh báo sớm vượt ngưỡng (Sụt công suất quang, tăng tải).
5	Notice	Trạng thái biến động bình thường nhưng cần lưu ý (Thay đổi cấu hình).
6	Informational	Thông tin vận hành thông thường (Thuê bao kết nối, chu kỳ quét).
7	Debug	Luồng thông tin vết chi tiết phục vụ lập trình viên gỡ lỗi logic.

1.2.2. Bản chất khoa học của dữ liệu nhật ký hệ thống (Syslog Data Cosmology).

Trong khoa học dữ liệu và quản trị hệ thống thông tin, dữ liệu nhật ký hệ thống (Syslog) được định nghĩa là chuỗi dữ liệu chuỗi thời gian (Time-series Data) dạng bán cấu trúc hoặc phi cấu trúc, phản ánh toàn bộ trạng thái vận hành động, các xung đột logic vật lý và hành vi nghiệp vụ của thiết bị viễn thông theo thời gian thực [10].

Đối với hệ thống mạng băng rộng cố định của VNPT Hà Nội, dữ liệu cảnh báo trong Syslog mang ba đặc tính cốt lõi của dữ liệu lớn (Big Data):

- Thể tích lớn (Volume): Với quy mô hàng nghìn phần tử mạng lõi, biên và mạng truy nhập đồng loạt hoạt động, số lượng bản tin Syslog phát sinh liên tục có thể đạt ngưỡng hàng triệu bản tin mỗi ngày, đòi hỏi năng lực lưu trữ phi quan hệ mở rộng theo chiều ngang.

- Vận tốc cao (Velocity): Luồng dữ liệu cảnh báo được đẩy về liên tục qua giao thức hướng không kết nối UDP port 514 với tần suất cực đại tại các khung giờ cao điểm lên tới hơn 8.000 sự kiện trên một giây (Events Per Second - EPS).

- Tính đa dạng (Variety): Đây là thách thức học thuật lớn nhất của đề tài. Do cấu trúc bản tin phụ thuộc hoàn toàn vào hệ điều hành chuyên biệt của từng nhà sản xuất (Cisco IOS-XR, Juniper Junos, Alcatel-Lucent TiMOS, Huawei VRP), dòng dữ liệu MSG chứa các cú pháp ngữ nghĩa (Semantics) và từ khóa cảnh báo hoàn toàn khác nhau.

1.2.3. Phân tích cấu trúc cú pháp dữ liệu cảnh báo đa chủng loại thiết bị (Multi-Vendor Syntax Analysis).

Để xây dựng được thuật toán bóc tách tự động, đề án tiến hành nghiên cứu sâu về quy luật cấu trúc dữ liệu cảnh báo đặc trưng của từng hệ điều hành thiết bị đang vận hành thực tế trên mạng lưới VNPT Hà Nội:

Cấu trúc cảnh báo trên hệ điều hành Cisco IOS-XR (Mạng lõi MAN-E): bản tin Syslog từ các dòng Router Cisco ASR-9010/ASR-9922 sau phần thông tin tiêu đề thời gian (Header) sẽ tuân thủ nghiêm ngặt định dạng cú pháp thông báo lỗi logic phân tách bằng dấu hai chấm và dấu gạch ngang:

Cisco Format = %FACILITY - SEVERITY - MNEMONIC: Message-text

Trong đó: FACILITY: định danh phân hệ phần cứng hoặc tiến trình phần mềm phát sinh lỗi (ví dụ: BGP, IFMGR, PKT_INFRA); SEVERITY biểu thị cấp độ nghiêm trọng; MNEMONIC là mã gợi nhớ duy nhất của lỗi logic đó. Cú pháp này có tính nhất quán cao, thuận lợi cho việc thiết lập các biểu thức chính quy (Regex) cố định để bóc tách thông tin về giao diện vật lý và trạng thái kết nối liên nút.

Cấu trúc cảnh báo trên hệ điều hành Juniper Junos (Hệ thống BRAS): thiết bị mạng biên BRAS của hãng Juniper sản sinh dữ liệu log có xu hướng tiệm cận với định dạng văn bản tự do hướng sự kiện. Cấu trúc đặc trưng thường bao gồm định danh tiến trình xử lý lõi (PFE - Packet Forwarding Engine hoặc authd - Authentication Daemon) kết hợp với chuỗi thông điệp mô tả chi tiết:

Juniper Format = process-name[pid] : %[FACILITY- SEVERITY-LOG_CODE]: Detailed-description

Điểm đặc thù trong dữ liệu cảnh báo BRAS Juniper là chứa các thông tin biến động động về mặt nghiệp vụ thuê bao, chẳng hạn như định danh tài khoản người dùng

(User-ID), định danh cổng logic đại diện cho khách hàng (pp0.xxxxx), và nguyên nhân thất bại của quá trình bắt tay thiết lập phiên kết nối PPPoE (Authentication failed hoặc CHAP/PAP timeout).

Cấu trúc cảnh báo trên hệ điều hành Alcatel-Lucent TiMOS (Thiết bị mạng truy nhập L2): hệ điều hành TiMOS của các dòng thiết bị chuyển mạch UPE ứng dụng cấu trúc log phân cấp tường minh thông qua từ khóa chỉ định mức độ nghiêm trọng ở ngay đầu phân đoạn thông điệp (ví dụ: CRITICAL, WARNING, INFO). Cấu trúc tổng quát có dạng:

Alcatel Format = SEVERITY-KEYWORD: MODULE_NAME #EVENT_ID Specific-event-string

Mã EVENT_ID là một chuỗi số nguyên đại diện cho các kịch bản lỗi hệ thống đã được định nghĩa sẵn trong tài liệu kỹ thuật của hãng. Việc nghiên cứu mã số hóa các EVENT_ID này là cơ sở để hệ thống phần mềm thực hiện các phép toán lọc gộp dữ liệu lớn (Aggregation) ở các chương sau.

1.2.4. Mô hình hóa bài toán chuyển đổi dữ liệu phi cấu trúc và định hướng số hóa báo cáo.

Từ kết quả nghiên cứu sâu về bản chất cú pháp của các dòng nhật ký thô phi cấu trúc từ các hãng thiết bị, đề án tổng hợp và mô hình hóa bài toán xử lý thông tin dưới dạng một ma trận ánh xạ logic. Mục tiêu là trích xuất từ các chuỗi ký tự tự do ngẫu nhiên thành các thuộc tính dữ liệu có cấu trúc đồng nhất, phục vụ trực tiếp cho các phân hệ báo cáo quản trị chuyên sâu của VNPT Hà Nội.

Bảng 1. 3 Ma trận phân tích ngữ nghĩa dữ liệu Syslog và định hướng thuộc tính số hóa.

Hãng sản xuất (OS)	Mã lỗi / Từ khóa cốt lõi (Keyword/Event Code)	Ngữ nghĩa kỹ thuật mạng viễn thông (Network Semantics)	Thuộc tính đích cần số hóa (Target Schema Fields)	Phân hệ ứng dụng số hóa đầu ra (Dashboard Module)
Cisco IOS-XR	LINEPROTO-5-UPDOWN	Trạng thái giao thức liên kết dữ liệu của cổng mạng bị thay đổi (Up hoặc Down).	Interface_Name, Link_Status, Timestamp	Phân hệ báo cáo chất lượng vật lý mạng lõi MAN-E.

Juniper Junos	AUTH_FAIL	Người dùng không vượt qua được bước xác thực tài khoản trên BRAS.	Subscriber_ID, Failure_Reason, BRAS_Node	Phân hệ số hóa báo cáo dữ liệu hệ thống BRAS.
Alcatel TiMOS	vrrpInstState Change	Thay đổi trạng thái dự phòng nóng giữa hai thiết bị (Master $\leftrightarrow$ Backup).	VRRP_ID, Current_State, Vlan_ID	Phân hệ báo cáo và theo dõi sự cố trực quan.
Huawei VRP	OLT LOS / LOSI	Cổng PON trên tổng đài bị mất tín hiệu quang từ thiết bị ONT của khách hàng.	OLT_Name, Slot_Port, ONT_ID, Frequency	Phân hệ báo cáo dữ liệu lớp mạng truy nhập L2-GPON.

Mặc dù dữ liệu Syslog đa nguồn từ các chủng loại thiết bị trên mạng lưới VNPT Hà Nội có dạng văn bản thô, bất đồng nhất và rời rạc; nhưng chúng đều tuân thủ các quy luật cú pháp nội tại chặt chẽ dựa trên kiến trúc hệ điều hành đặc trưng của từng hãng sản xuất. Đây chính là tiền đề khoa học quan trọng khẳng định tính khả thi của việc ứng dụng các kỹ thuật biểu thức chính quy (Regex Engine) và các pipeline xử lý luồng dữ liệu lớn (Big Data Pipeline) để số hóa toàn diện quy trình giám sát hạ tầng. Tuy nhiên, trong bối cảnh thực tế hiện nay, khi mạng lưới băng rộng của VNPT Hà Nội đang vận hành hàng nghìn phần tử mạng lõi phát sinh dòng log Syslog liên tục với mật độ cực kỳ đậm đặc, cơ chế quản lý dữ liệu log phân tán và thủ công hiện tại đã bộc lộ rõ rào cản lớn. Phương thức tiếp cận cũ này không chỉ gây lãng phí tài nguyên lưu trữ cục bộ trên thiết bị mà còn làm suy giảm khả năng chủ động điều tra, phân tích nguyên nhân gốc rễ sự cố diện rộng của đội ngũ kỹ sư mạng lưới. Thách thức mang tính hệ thống đó đòi hỏi phải có một giải pháp công nghệ dữ liệu lớn có khả năng thu thập tập trung hóa toàn bộ dòng log, cho phép lưu trữ tối ưu và xử lý truy vấn thời gian thực với tốc độ cao, thay thế hoàn toàn cho các mô hình cơ sở dữ liệu quan hệ truyền thống đã lỗi thời trước các bài toán dữ liệu chuỗi thời gian quy mô lớn [3], [4].

1.3. Kết luận chương 1.

Trong Chương 1, đề án đã hoàn thành việc khảo sát hạ tầng mạng băng rộng cố định tại VNPT Hà Nội và làm rõ vai trò cốt lõi của dữ liệu Syslog trong công tác quản trị hệ thống. Kết quả phân tích thực trạng cho thấy cách tiếp cận giám sát truyền thống chưa đáp ứng được yêu cầu khi quy mô mạng lưới ngày càng mở rộng và phức tạp. Để giải quyết bài toán này, việc nghiên cứu và xây dựng một hệ thống thu thập, bóc tách dữ liệu Syslog tự động, tập trung là hoàn toàn cấp thiết. Những kết luận và định hướng rút ra từ chương này sẽ là nền tảng lý thuyết và thực tiễn quan trọng để tác giả đề xuất kiến trúc hệ thống và giải pháp công nghệ chi tiết ở Chương 2.

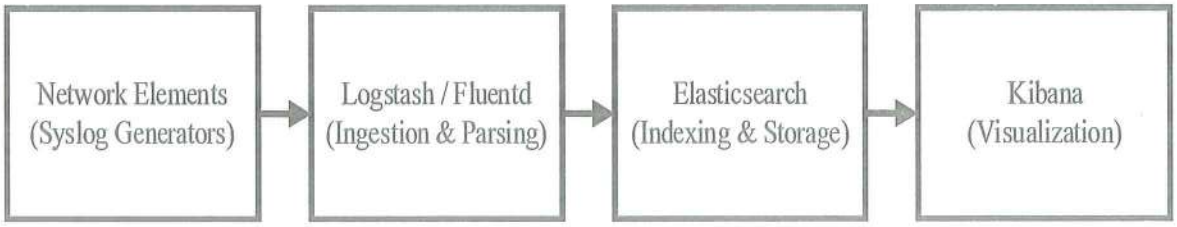
CHƯƠNG 2: CÁC ỨNG DỤNG CÔNG NGHỆ, HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU ĐỂ SỬ DỤNG PHÂN TÍCH LƯU TRỮ SYSLOG PHỤC VỤ CÔNG TÁC QUẢN TRỊ HỆ THỐNG

2.1. Các ứng dụng và các công nghệ áp dụng để phân tích Syslog hệ thống.

Trong kỷ nguyên mạng điều khiển bằng dữ liệu (Data-driven Networking), dữ liệu nhật ký hệ thống (Syslog) phát sinh từ các phần tử hạ tầng dị thể không còn đơn thuần là các tệp văn bản tĩnh, mà được nhìn nhận là một dạng dòng dữ liệu bán cấu trúc (Semi-structured Data Stream) có tốc độ tăng trưởng lớn (Volume) và tính đa dạng cao (Variety). Để khai thác giá trị từ nguồn tài nguyên này phục vụ công tác quản trị, giám sát và phân tích nguyên nhân gốc rễ (Root Cause Analysis - RCA), các nghiên cứu trong và ngoài nước tập trung vào ba xu hướng tiếp cận công nghệ cốt lõi: các giải pháp thương mại đóng gói, các bộ công cụ mã nguồn mở truyền thống, và hệ sinh thái xử lý dữ liệu lớn chuyên sâu [10].

2.1.1. Cách tiếp cận dựa trên hệ sinh thái mã nguồn mở tập trung (ELK Stack / EFK Stack).

Mô hình kiến trúc ELK (Elasticsearch - Logstash - Kibana) hoặc biến thể EFK (Elasticsearch - Fluentd - Kibana) hiện nay đóng vai trò là một trong những giải pháp nền tảng phổ biến nhất trong nghiên cứu và ứng dụng thực tiễn đối với bài toán quản lý dữ liệu log tập trung [3].



Hình 2. 1 Mô hình kiến trúc ELK (Elasticsearch - Logstash - Kibana)

- Logstash / Fluentd (Lớp thu thập và tiền xử lý - Ingestion Layer): Vận hành như một hệ thống đường ống dữ liệu (Data Pipeline). Công nghệ này áp dụng các bộ lọc (Filters) dựa trên biểu thức chính quy (Regular Expressions - Regex) nâng cao

hoặc cấu trúc Grok Pattern để chuẩn hóa các chuỗi văn bản Syslog tự do (Unstructured Text) thành các trường dữ liệu có cấu trúc định dạng JSON.

- Elasticsearch (Lớp lưu trữ và chỉ mục - Storage & Indexing Layer): Cơ sở dữ liệu phân tán hoạt động theo mô hình NoSQL, xây dựng trên nền tảng Apache Lucene. Elasticsearch sử dụng cơ chế lập chỉ mục đảo ngược (Inverted Index), cho phép tối ưu hóa các truy vấn tìm kiếm toàn văn (Full-text Search) với độ trễ tiệm cận thời gian thực (Near Real-Time), đáp ứng yêu cầu tra cứu sự cố trên quy mô hàng triệu bản tin.

- Kibana (Lớp hiển thị và tương tác - Visualization Layer): Cung cấp giao diện trực quan hóa dữ liệu, hỗ trợ chuyển đổi các truy vấn log phức tạp thành các biểu đồ động, bảng phân phối tần suất sự cố, phục vụ trực tiếp cho lớp quyết định quản trị (Decision-making Layer).

2.1.2. Cách tiếp cận dựa trên các giải pháp thương mại quy mô công nghiệp (Enterprise Solutions).

Đối với hạ tầng mạng viễn thông cốt lõi, các công cụ thương mại tiêu biểu như Splunk hoặc SolarWinds Log Analyzer đại diện cho lớp giải pháp có tính toàn vẹn cao.

- Splunk: Vận hành dựa trên cấu trúc chỉ mục độc quyền cấu hình động (Dynamic Indexing) không cần định nghĩa Schema trước. Splunk tích hợp sẵn các mô hình phân tích học máy (Machine Learning Toolkit - MLTK) nhằm phát hiện các mẫu hành vi bất thường (Anomaly Detection) tự động từ đòng Syslog và dự báo các nguy cơ nghẽn mạch hoặc lỗi phần cứng dựa trên chuỗi thời gian (Time-series Analysis).

- Hạn chế nghiên cứu: Mặc dù sở hữu năng lực xử lý vượt trội và tính an toàn cao, các giải pháp này mang bản chất "hộp đen" (Black-box), hạn chế khả năng tùy biến sâu theo đặc thù ngữ nghĩa của từng nhà mạng. Đồng thời, mô hình cấp phép thương mại dựa trên dung lượng dữ liệu xử lý theo ngày (Volume-based Licensing) tạo ra rào cản kinh tế khổng lồ khi áp dụng vào các hệ thống mạng MAN-E quy mô cực lớn.

2.1.3. Xu hướng áp dụng kiến trúc phi quan hệ và xử lý luồng thời gian thực (Modern Stream Processing Pipeline).

Nhằm khắc phục giới hạn về mặt hiệu năng xử lý của các hệ quản trị cơ sở dữ liệu quan hệ truyền thống (RDBMS) khi đối mặt với mật độ dữ liệu Syslog dạng dòng (Stream) có xung tần suất lớn (ví dụ: các sự cố sập nguồn trạm viễn thông gây bùng nổ hàng vạn bản tin/giây), xu hướng nghiên cứu hiện đại dịch chuyển mạnh mẽ sang việc kết hợp các công nghệ hướng luồng và NoSQL hướng tài liệu (Document-oriented NoSQL) [8].

- Apache Kafka (Lớp đệm thu thập phân tán - Distributed Message Broker): Đóng vai trò là hệ thống hàng đợi thông điệp phân tán, giải quyết bài toán lệch pha tốc độ giữa bên phát log (Thiết bị mạng) và bên xử lý (Ứng dụng). Kafka giúp hệ thống đạt khả năng chịu lỗi cao (Fault-tolerance) và bảo toàn dữ liệu khi có hiện tượng nghẽn cục bộ.

- MongoDB (Lớp lưu trữ hướng tài liệu - Document Store): Với cấu trúc dữ liệu linh hoạt dựa trên định dạng BSON (Binary JSON), MongoDB loại bỏ các ràng buộc nghiêm ngặt về khóa ngoại và toàn vẹn thực thể của cơ sở dữ liệu quan hệ. Điều này cho phép hệ thống lưu trữ các bản tin Syslog từ nhiều chủng loại thiết bị (Cisco, Juniper, Huawei) có cấu trúc ngữ nghĩa khác biệt nhau trên cùng một tập hợp (Collection) mà không làm suy giảm hiệu năng ghi dữ liệu (Write Intensive) [4] [9].

2.1.4. Đánh giá và thảo luận so sánh luận chứng giải pháp (Comparative Analysis).

Để thiết l`ập luận cứ khoa học cho việc lựa chọn công nghệ triển khai trong khuôn khổ đề án, bảng 2.1 tổng hợp và so sánh các tiêu chí cốt lõi giữa các hướng tiếp cận công nghệ hiện nay:

Bảng 2. 1 Ma trận so sánh các đặc tính kỹ thuật của các nền tảng phân tích Syslog

Tiêu chí so sánh	Hệ sinh thái mã nguồn mở (ELK/EFK)	Giải pháp thương mại (Splunk)	Kiến trúc tùy biến kết hợp (Kafka + NoSQL/MongoDB)
Mô hình kiến trúc	Phân tán, định hình theo luồng tích hợp sẵn	Đóng gói nguyên khối, độc quyền	Kiến trúc hướng vi dịch vụ (Microservices), mô-đun hóa cao

Tiêu chí so sánh	Hệ sinh thái mã nguồn mở (ELK/EFK)	Giải pháp thương mại (Splunk)	Kiến trúc tùy biến kết hợp (Kafka + NoSQL/MongoDB)
Tốc độ ghi dữ liệu (Ingestion Rate)	Trung bình - Cao (Bị giới hạn bởi năng lực phân tích cú pháp của Logstash)	Rất cao (Tối ưu hóa mã nguồn đóng)	Cực kỳ cao (Nhờ cơ chế phân mảnh/Sharding của Kafka và ghi phi tuần tự của NoSQL)
Khả năng mở rộng (Scalability)	Mở rộng theo chiều ngang (Horizontal Scaling) nhưng phức tạp ở tầng phân vùng	Mở rộng tốt nhưng chi phí tài chính tăng tuyến tính	Mở rộng độc lập từng phân lớp (Thu thập, Đệm, Lưu trữ) linh hoạt
Khả năng tùy biến ngữ nghĩa	Khá (Thông qua cấu hình Grok filter thủ công)	Thấp (Phụ thuộc vào các App bổ trợ từ nhà phát triển)	Tuyệt đối (Cho phép xây dựng thuật toán bóc tách động ở tầng ứng dụng Backend)
Chi phí triển khai	Thấp (Chi phí hạ tầng phần cứng)	Rất cao (Bao gồm chi phí bản quyền và hạ tầng)	Thấp (Sử dụng mã nguồn mở kết hợp tối ưu phần cứng)

Kết luận chọn lựa công nghệ: Dựa trên kết quả phân tích ma trận kỹ thuật, mặc dù hệ sinh thái ELK mang lại sự tiện lợi trong cấu hình, nhưng việc xử lý bóc tách các trường dữ liệu Syslog phức tạp, đa cấu trúc của mạng viễn thông VNPT Hà Nội đòi hỏi một năng lực tùy biến thuật toán bóc tách sâu sắc ở tầng Backend. Đồng thời, để đảm bảo hệ thống vận hành bền bỉ trước các kịch bản bùng nổ bản tin sự cố, việc xây dựng một kiến trúc kết hợp sử dụng NoSQL (MongoDB) làm lõi lưu trữ linh hoạt, kết hợp cùng các công nghệ hiện đại ở tầng ứng dụng là giải pháp tối ưu, dung hòa được bài toán hiệu năng, khả năng làm chủ công nghệ và tính kinh tế của đề án nghiên cứu

2.2. Hệ cơ sở dữ liệu sử dụng để lưu trữ và xử lý Syslog hệ thống.

2.2.1. Phân tích bài toán tổng thể và các ràng buộc đặc thù của hạ tầng mạng MAN-E VNPT Hà Nội.

Mạng đô thị băng rộng (MAN-E) của VNPT Hà Nội là một hệ sinh thái hạ tầng viễn thông phức tạp, đóng vai trò trục liên kết cốt lõi cho các dịch vụ FTTH, MyTV, VPN-Wan và kênh thuê riêng của khối khách hàng doanh nghiệp, chính phủ.

Do tính chất phân cấp mạnh mẽ từ lớp Core (Mạng lõi), Aggregation (Thu gom) đến Access (Truy nhập), dòng dữ liệu Syslog phát sinh từ hạ tầng này mang ba đặc tính ràng buộc cực kỳ nghiêm ngặt:

- Sự dị thể về cấu trúc cú pháp (Syntactic Heterogeneity): Hệ thống bao gồm hàng nghìn phần tử mạng đến từ nhiều nhà sản xuất khác nhau như Cisco (dòng ASR, Catalyst sử dụng Cisco IOS-XR/XE), Juniper (dòng MX sử dụng Junos), Huawei (dòng NE sử dụng VRP) và Nokia/Alcatel-Lucent (dòng 7750 SR sử dụng TiMOS). Mặc dù các bản tin cơ bản tuân theo chuẩn RFC 3164 hoặc RFC 5424, phần nội dung thông điệp (Message Payload) lại hoàn toàn khác biệt về cấu trúc ngữ nghĩa giữa các hãng, gây khó khăn lớn cho các bộ phân tích cú pháp cố định.

- Áp lực về xung tần suất dữ liệu (High-Burst Ingestion Pressure): Ở trạng thái vận hành bình thường, lưu lượng log trung bình dao động từ 2.000 đến 5.000 EPS (Events Per Second). Tuy nhiên, khi xảy ra sự cố diện rộng như mất điện trạm viễn thông (đặc biệt trong mùa mưa bão) hoặc đứt cáp quang trục, hệ thống sẽ hứng chịu hiện tượng bùng nổ bản tin (Log Burst) lên tới hơn 30.000 EPS trong vài giây. Kiến trúc hệ thống nếu không có cơ chế đệm động sẽ ngay lập tức bị tràn bộ nhớ hoặc từ chối dịch vụ (DoS).

- Yêu cầu độ trễ xử lý (Processing Latency Constraint): Để phục vụ công tác giám sát thời gian thực và tự động kích hoạt cảnh báo đến các kỹ sư trực ca tại OMC, tổng thời gian từ khi thiết bị mạng phát sinh bản tin Syslog cho đến khi hệ thống hoàn thành bóc tách, phân tích ngữ nghĩa, lưu trữ và hiển thị trên Dashboard không được vượt quá tlatency ≤ 2000 ms

Từ các ràng buộc trên, nghiên cứu này đề xuất một kiến trúc hệ thống phân lớp chức năng hóa (Layered Functional Architecture) nhằm tách biệt hoàn toàn giữa các tác vụ: Thu thập, Đệm dữ liệu, Xử lý bóc tách cú pháp, Lưu trữ phi quan hệ và Hiển thị tương tác.

2.2.2. Kiến trúc tổng thể và cơ chế luồng dữ liệu của hệ thống đề xuất (System Pipeline Architecture).

Kiến trúc hệ thống bóc tách và phân tích ngữ nghĩa Syslog đề xuất được thiết kế dựa trên mô hình định hướng vi dịch vụ (Microservices Architecture), cho phép các thành phần độc lập mở rộng theo chiều ngang khi tải hệ thống tăng cao. Sơ đồ kiến trúc tổng thể được mô tả chi tiết tại Hình 2.2 dưới đây:



Hình 2. 2 Sơ đồ kiến trúc phân lớp và luồng dữ liệu hệ thống quản lý Syslog tập trung

Nguyên lý vận hành của luồng dữ liệu:

- Giai đoạn sinh nhật ký: Khi một phần tử mạng (NE) xuất hiện thay đổi trạng thái (ví dụ: cổng giao tiếp UP/DOWN, lỗi bit quang CRC, hoặc thay đổi định tuyến BGP), tiến trình Syslog nội tại của hệ điều hành thiết bị sẽ đóng gói thông tin thành một bản tin văn bản và đẩy qua giao thức UDP về địa chỉ IP của máy chủ thu thập tại cổng mặc định 514.

- Giai đoạn đệm trung gian: Thành phần Syslog Receiver Gateway chịu trách nhiệm mở các Socket UDP để hứng luồng dữ liệu thô này. Để tránh mất mát dữ liệu khi hệ thống xử lý phía sau bị quá tải, Gateway này hoạt động độc lập và thực hiện thao tác ghi ngay lập tức (Produce) các thông điệp thô vào phân vùng hàng đợi của Apache Kafka dưới một chủ đề xác định gọi là raw-syslog-stream.

- Giai đoạn xử lý ngữ nghĩa (Trọng tâm Đề án): Ứng dụng Backend (Log Processing Engine) đóng vai trò là một Consumer đăng ký nhận dữ liệu từ Kafka Topic. Tiến trình xử lý sử dụng mô hình đa luồng (Multi - threading) hoặc bất đồng bộ (Asynchronous Event-driven) để liên tục rút dữ liệu thô ra, đưa qua các bộ lọc lọc nhiễu, sau đó áp dụng ma trận biểu thức chính quy phù hợp với hãng sản xuất thiết bị để bóc tách thông điệp thành các trường cấu trúc rõ ràng.

- Giai đoạn bền vững hóa dữ liệu: Bản tin sau khi chuẩn hóa dưới dạng cấu trúc tài liệu JSON sẽ được đẩy vào cơ sở dữ liệu MongoDB theo cơ chế ghi gộp (Bulk Write) để tối ưu hóa hiệu năng I/O đĩa cứng. Hệ thống đồng thời tính toán lập chỉ mục (Index) trên các trường tần suất cao để phục vụ tra cứu.

- Giai đoạn tiêu thụ và trực quan: Ứng dụng Front-end viết bằng Angular thiết lập kết nối song công thông qua giao thức WebSockets với Backend nhằm cập nhật tức thời các lỗi nghiêm trọng lên màn hình điều hành của kỹ sư mạng lưới mà không cần tải lại trang.

2.2.3. Nghiên cứu sâu tầng bóc tách cấu trúc và kỹ thuật chuẩn hóa dữ liệu đa chủng loại (Normalization Layer).

Bản tin Syslog truyền thống theo RFC 3164 có cấu trúc chuỗi ký tự tự do thô như sau:

Raw Syslog String} = <PRI> HEADER MSG

Trong đó phần MSG chứa thông tin cốt lõi của sự cố nhưng lại hoàn toàn phụ thuộc vào nhà sản xuất. Mục tiêu của phân lớp này là phải chuyển đổi chuỗi thô trên thành một đối tượng dữ liệu cấu trúc (Structured Document) có định dạng đồng nhất bất kể thiết bị phát là Cisco, Juniper hay Huawei.

a) Thuật toán phân rã tổng quát (Generalized Parsing Pipeline).

Khi một chuỗi log thô đi vào, hệ thống thực hiện quy trình phân rã tuần tự qua các bước:

- Trích xuất PRI (Priority): Tách phần giá trị nằm giữa hai dấu ngoặc < và >. Từ giá trị PRI này, áp dụng công thức toán học để tính toán ra hai chỉ số quan trọng là Facility (Nguồn phát) và Severity (Mức độ nghiêm trọng):

$$\text{Facility} = \lfloor PRI / 8 \rfloor$$

$$\text{Severity} = PRI \pmod 8$$

Tách HEADER: Trích xuất mốc thời gian (Timestamp) và địa chỉ IP hoặc Hostname của thiết bị gửi bản tin. Do mốc thời gian của chuẩn cũ (ví dụ: Oct 11 14:32:10) thiếu thông tin năm và múi giờ, hệ thống phải tự động bù giờ hệ thống hiện tại để chuẩn hóa về định dạng ISO 8601 (YYYY-MM-DDTHH:mm:ss.sssZ).

- Định danh nhà sản xuất (Vendor Identification): Dựa vào địa chỉ IP nguồn, hệ thống tra cứu trong bảng danh mục quản lý tài nguyên thiết bị (RIMS) được lưu trong bộ nhớ đệm Cache để xác định thiết bị này thuộc hãng nào (Cisco, Juniper, Huawei, Nokia). Từ đó kích hoạt Factory Pattern để gọi bộ Parser tương ứng của hãng đó.

b) Thiết kế ma trận biểu thức chính quy (Regex Pattern Matrix) cho các dòng thiết bị mạng cốt lõi.

Dưới đây là nghiên cứu chi tiết về cấu trúc log đặc thù và các Regex Pattern tương ứng được thiết kế để bóc tách ngữ nghĩa:

Thiết bị Cisco IOS-XR (Dòng ASR-9000):

Định dạng log thô đặc trưng:

Plaintext

<133>Oct 12 08:15:22.104 ICT: LC/0/0/CPU0: Oct 12 08:15:22.105: ifmgr [204]: %PKT_INFRA-LINEPROTO-5-UPDOWN: Line protocol on Interface TenGigE0/0/0/1, state changed to Down

Phân tích đặc trưng ngữ nghĩa: Log Cisco sử dụng ký tự % làm dấu hiệu bắt đầu của định danh lỗi, theo sau là cấu trúc: FACILITY-SUBFACILITY-SEVERITY-MNEMONIC.

Regex thiết kế chuyên biệt:

Đoạn mã

```
^<(?<pri>\d+)>(?<timestamp>\w+\s+\d+\s+\d+:\d+:\d+\.\d+)\s+(?<timezone>\w+
):\s+(?<card>\S+):\s*.*:\s*(?<process>\S+?)\[(?<pid>\d+)\]:\s*%(?<facility>[A-
Z0-9_]+)-(?<mnemonic>[A-Z0-9_]+)-(?<severity_num>\d)-
(?<sub_mnemonic>[A-Z0-9_]+)\s*:\s*(?<message>.+)$
```

Kết quả ánh xạ thuộc tính JSON:

JSON

```
{
        "vendor": "Cisco",
        "severity": 5,
        "facility": "PKT_INFRA",
        "mnemonic": "LINEPROTO",
        "event_type": "UPDOWN",
        "component": "Interface TenGigE0/0/0/1",
        "status": "Down",
        "message": "Line protocol on Interface TenGigE0/0/0/1, state changed to
       Down"
}
```

Thiết bị Juniper MX (Hệ điều hành Junos):

Định dạng log thô đặc trưng:

Plaintext

```
<141>2026-10-12T08:16:01.223Z LE-AGG-02 mib2d[4102]:
SNMPEID_CONN_TIMEOUT: SNMP agent connection to subsystem failed,
connection timed out
```

Phân tích đặc trưng ngữ nghĩa: Log Juniper tuân thủ tốt cấu trúc chuẩn mới, tên lỗi viết bằng chữ hoa cách nhau dấu gạch dưới nằm ngay sau định danh tiến trình và dấu hai chấm.

Regex thiết kế chuyên biệt:

Đoạn mã

^<(?<pri>\d+)>(?<timestamp>\S+)\s+(?<hostname>\S+)\s+(?<process>\S+?)\[(?<pid>\d+)\]:\s*(?<mnemonic>[A-Z0-9_]+):\s*(?<message>.+)$

Thiết bị Huawei (Hệ điều hành VRP):

Định dạng log thô đặc trưng:

Plaintext
<188>Oct 12 2026 08:17:44 UPE-HAST-01 %%01IFNET/4/IF_BACKGROUND_DFSTATUS_CHANGE(l)[54]:The line protocol status of the interface GigabitEthernet0/3/2 changes to DOWN

Phân tích đặc trưng ngữ nghĩa: Log Huawei nhận diện bằng chuỗi %%, theo sau là mã phiên bản (ví dụ: 01), tiếp đến là tên mô-đun, mức độ nghiêm trọng và tên sự kiện nằm trong cấu trúc dấu gạch chéo /.

Regex thiết kế chuyên biệt:

Đoạn mã

^<(?<pri>\d+)>(?<timestamp>\w+\s+\d+\s+\d+\s+\d+:\d+:\d+)\s+(?<hostname>\S+)\s+%%(?<version>\d+)(?<facility>[A-Z0-9_]+)\/(?<severity_num>\d)\/(?<mnemonic>[A-Z0-9_]+)\((?<flag>\w+)\)\[(?<pid>\d+)\]:\s*(?<message>.+)$

c) Mô hình hóa hàm chuẩn hóa dữ liệu (Normalization Function Model).

Gọi S là tập hợp tất cả các chuỗi log thô nhận được từ mạng lưới, và D là không gian các tài liệu cấu trúc định dạng chuẩn hóa JSON. Quá trình xử lý của tầng bóc tách cấu trúc được định nghĩa như một hàm ánh xạ toán học $f: S \times M \rightarrow D$, trong đó M là ma trận siêu dữ liệu thiết bị (Device Metadata).

Hàm ánh xạ f thực hiện biến đổi cấu trúc theo mô hình giải thuật sau:

$$(\forall s \in S, \quad f(s,m) = d)$$

Trong đó đối tượng kết quả d ∈ D bắt buộc phải chứa một tập hợp các thuộc tính cốt lõi cố định (Core Attributes Base Schema):

d = {id, timestamp,host_ip, vendor,severity,module,event_key, raw_message, cleaned _payload }

cleaned_payload: Là một đối tượng con dạng lồng nhau (Nested JSON Object) chứa các biến số động bóc tách từ thông điệp (như tên giao diện cụ thể TenGigE0/0/0/1, giá trị công suất suy hao quang -22.5 dBm, mã lỗi cụ thể...). Điều

này cho phép hệ thống lưu giữ thông tin chi tiết mà không làm phá vỡ cấu trúc bảng phẳng của cơ sở dữ liệu [11].

2.2.4. Phân tích ngữ nghĩa sự cố và công cụ phân loại mức độ nghiêm trọng (Semantic & Fault Classification).

Sau khi dữ liệu được cấu trúc hóa thành các trường riêng biệt, hệ thống cần tiến hành phân tích ngữ nghĩa lớp cao nhằm phân loại bản tin theo bản chất kỹ thuật viễn thông chứ không đơn thuần dựa trên mức Severity hệ thống của hãng (vốn rất chủ quan và không phản ánh đúng mức độ ảnh hưởng dịch vụ).

a) Thiết kế cây phân loại ngữ nghĩa sự cố mạng (Fault Semantic Tree).

Hệ thống đề xuất xây dựng một bộ phân loại ngữ nghĩa phân cấp, chia sự cố mạng MAN-E thành 5 nhóm danh mục cốt lõi:

- Nhóm sự cố Hạ tầng Truyền dẫn vật lý (Physical Layer Faults): Bao gồm các lỗi liên quan đến cổng vật lý, module quang SFP/XFP. Ví dụ: Lỗi tăng tỷ lệ bit lỗi chu kỳ (CRC Errors), lỗi suy hao quang vượt ngưỡng (Rx Power Low), hoặc giao tiếp vật lý mất tín hiệu (Loss of Signal - LOS).

- Nhóm sự cố Giao thức Liên kết dữ liệu (Link Layer Faults): Tập trung vào các trạng thái Up/Down của giao thức đường truyền như PPP, PPPoE (thuê bao FTTH), lỗi giao tiếp cổng bó (Link Aggregation Control Protocol - LACP), hoặc các sự cố liên quan đến vòng lặp mạng (Spanning Tree Protocol Loop).

- Nhóm sự cố Định tuyến và Chuyển mạch (Routing & Network Layer Faults): Ghi nhận các sự kiện thay đổi trạng thái láng giềng định tuyến (Routing Neighbor Flapping) của các giao thức chạy trên mạng MAN-E như OSPF, BGP, IS-IS hoặc sự cố chuyển mạch dự phòng router ảo VRRP.

- Nhóm sự cố Dịch vụ và Thuê bao (Service Layer Faults): Ghi nhận các lỗi liên quan đến hệ thống cổng kiểm soát truy cập biên mạng băng rộng (BRAS) như lỗi bùng nổ truy vấn xác thực không hợp lệ (Authentication Failures), cạn kiệt kho địa chỉ IP (IP address pool exhaustion), hoặc lỗi kết nối đến máy chủ RADIUS/DHCP.

- Nhóm sự cố An toàn hệ thống và Phần cứng (Hardware & Security Faults): Giám sát các cảnh báo phần cứng thiết bị như nhiệt độ vượt ngưỡng (High Temperature), lỗi quạt tản nhiệt (Fan Tray Failure), CPU tăng cao quá mức do tiến trình xử lý quá tải, hoặc các hành vi truy cập CLI trái phép (Unauthorized Access Attempt).

b) Giải thuật ánh xạ ngữ nghĩa động (Dynamic Semantic Mapping Engine).

Để thực hiện phân loại tự động bản tin Syslog vào cây ngữ nghĩa trên một cách tối ưu, hệ thống sử dụng thuật toán so khớp dựa trên luật (Rule-based Matching Engine) kết hợp ma trận trọng số từ khóa.

Mô tả giải thuật:

- Đầu vào: Đối tượng tài liệu *d* đã qua chuẩn hóa.

- Thao tác: Hệ thống trích xuất cặp giá trị (vendor, module ,event_key). Sau đó, tra cứu trong Từ điển Sự cố Tập trung (Centralized Fault Dictionary) được cấu trúc dạng bảng băm (Hash Map) để tìm kiếm luật ánh xạ.

- Xử lý ngoại lệ: Trong trường hợp mã sự cố (event_key) chưa từng xuất hiện trong từ điển (mã lỗi mới do thiết bị nâng cấp phần mềm nâng cao), hệ thống tự động kích hoạt bộ quét ngữ nghĩa nội dung văn bản (message). Bộ quét này sử dụng biểu thức chính quy để tìm kiếm sự xuất hiện của các từ khóa có trọng số cao (ví dụ: "fail", "drop", "critical", "voltage", "loop") để đưa ra dự đoán phân loại tạm thời và đẩy vào hàng đợi kiểm duyệt của quản trị viên hệ thống nhằm cập nhật từ điển.

2.2.5. Mô hình hóa lớp đệm phân tán Apache Kafka chống bùng nổ dữ liệu (Fault-Tolerant Buffering Layer).

Như đã phân tích, hiện tượng bùng nổ dữ liệu log trong mạng viễn thông khi xảy ra sự cố là rào cản lớn nhất gây đổ vỡ hệ thống giám sát. Sự xuất hiện của phân lớp đệm phân tán sử dụng Apache Kafka đóng vai trò cực kỳ quan trọng giúp kiến trúc đạt được trạng thái bền vững cao (High Availability).

a) Kiến trúc Topic, Partition và Cơ chế Phân mảnh Dữ liệu (Sharding Mechanism).

Để đảm bảo khả năng ghi song song tốc độ cao và phân phối tải đồng đều trên cụm máy chủ Kafka (Kafka Cluster), chủ đề raw-syslog-stream được cấu hình phân mảnh thành nhiều Partitions độc lập. Việc định tuyến bản tin vào Partition nào được quyết định bởi một hàm băm dựa trên khóa (Key-based Partitioning Strategy).

Khóa phân vùng *(K)* được lựa chọn chính là địa chỉ IP nguồn của thiết bị phát sinh log:

$$K = \text{host_ip}$$

Vị trí Partition (P_idx) được tính toán theo công thức:

$$Pidx = \text{MurmurHash2}(K) \pmod N$$

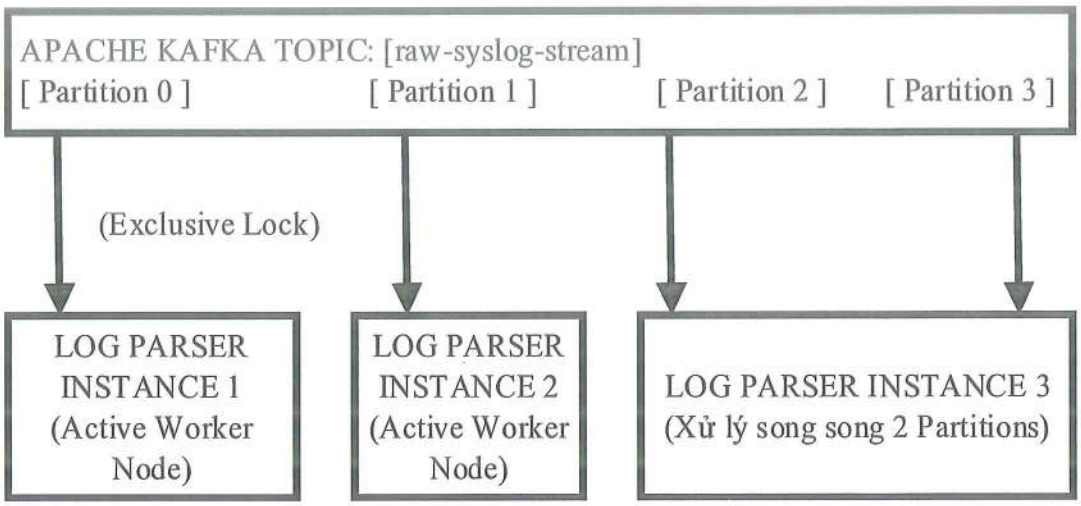
Trong đó N là tổng số lượng Partitions được cấu hình cho Topic.

Ý nghĩa khoa học của cơ chế này: Đảm bảo tính tuần tự theo thời gian (Preserve Ordering): Trong kiến trúc Kafka, tính tuần tự của thông điệp chỉ được bảo toàn tuyệt đối trong phạm vi một Partition. Bằng việc sử dụng host_ip làm khóa băm, toàn bộ các bản tin log phát sinh liên tiếp từ cùng một thiết bị mạng cụ thể sẽ bắt buộc phải đi vào cùng một Partition duy nhất. Điều này đảm bảo rằng khi ứng dụng Backend tiêu thụ dữ liệu để phân tích nguyên nhân gốc rễ, trạng thái sự cố của thiết bị đó sẽ được xử lý đúng theo trình tự thời gian thực tế xảy ra, tránh hiện tượng đảo lộn trật tự log (ví dụ bản tin báo cổng UP lại bị xử lý trước bản tin báo cổng DOWN do đi khác đường).

- Cân bằng tải ngang hàng (Load Balancing): Các thiết bị mạng khác nhau sẽ có địa chỉ IP khác nhau, do đó lượng log khổng lồ từ toàn mạng sẽ được phân tán đều ra tất cả các Partitions, cho phép tận dụng tối đa năng lực ghi đĩa của cụm máy chủ.

b) Cơ chế Consumer Group và khả năng tự động khôi phục lỗi (Fault-Tolerance).

Ở tầng xử lý Backend, hệ thống thiết lập một cụm các tiến trình xử lý hoạt động dưới dạng một Consumer Group thống nhất. Mỗi tiến trình thành phần (Consumer Instance) trong nhóm sẽ được điều phối độc lập độc quyền đọc dữ liệu từ một hoặc một số Partitions nhất định.



Hình 2. 3 Cơ chế điều phối xử lý dữ liệu song song và chịu lỗi sử dụng Kafka Consumer Group

Khi hệ thống vận hành, nếu một máy chủ xử lý log gặp sự cố phần cứng đột ngột và ngừng hoạt động (Crash), bộ quản lý điều phối của Kafka (Kafka Coordinator) sẽ ngay lập tức phát hiện thông qua việc mất tín hiệu nhịp tim (Heartbeat Timeout). Hệ thống tự động kích hoạt quy trình Rebalance (Tái cân bằng): chuyển giao quyền tiêu thụ các Partitions đang bị bỏ trống cho các Instances còn lại đang hoạt động bình thường. Quá trình này diễn ra hoàn toàn tự động trong thời gian vài trăm mili-giây, đảm bảo luồng thu thập dữ liệu log toàn mạng không bị gián đoạn bất kỳ một giây nào [5].

2.2.6. Thiết kế mô hình lưu trữ tối ưu trên hệ quản trị cơ sở dữ liệu phi quan hệ MongoDB (Storage Optimization Layer).

Dữ liệu nhật ký hệ thống sau bóc tách cần được lưu trữ lâu dài (từ 3 đến 6 tháng theo quy định quản lý hạ tầng viễn thông) phục vụ công tác hậu kiểm và phân tích xu hướng. Việc thiết kế mô hình dữ liệu trên cơ sở dữ liệu NoSQL MongoDB đòi hỏi sự tối ưu hóa sâu sắc cả về cấu trúc tài liệu lẫn chiến lược lập chỉ mục nhằm giải quyết bài toán mâu thuẫn giữa tốc độ ghi cực nhanh (Write-Intensive) và tốc độ truy vấn tức thời (Read-Intensive) [4].

a) Thiết kế lược đồ tài liệu JSON linh hoạt (Flexible Schema Design).

Thay vì thiết kế dạng bảng cứng nhắc của mô hình cơ sở dữ liệu quan hệ (RDBMS), hệ thống sử dụng cấu trúc tài liệu dạng cây linh hoạt của MongoDB. Dưới

đây là thiết kế lược đồ thực thể (Data Schema Design) chi tiết cho bộ sưu tập dữ liệu cốt lõi parsed_alerts:

JSON


```
{
  "_id": {"$oid": "6527b140f1d2a3e4b5c6d7e8"},
  "timestamp": {"$date": "2026-10-12T08:15:22.104Z"},
  "received_at": {"$date": "2026-10-12T08:15:22.110Z"},
  "host_ip": "10.252.12.41",
  "hostname": "LE-ASR9k-01.HN",
  "vendor": "Cisco",
  "facility": "PKT_INFRA",
  "severity": 5,
  "mnemonic": "LINEPROTO",
  "event_key": "LINEPROTO-UPDOWN",
  "semantic_category": "Link Layer Faults",
  "raw_message": "Line protocol on Interface TenGigE0/0/0/1, state changed to Down",
  "extracted_data": {
    "interface_name": "TenGigE0/0/0/1",
    "current_state": "DOWN",
    "previous_state": "UP"
  },
  "is_resolved": false,
  "correlation_id": {"$oid": "6527b140f1d2a3e4b5c6d7f9"}
}
```


Ưu điểm học thuật của thiết kế lược đồ:

- Trường extracted_data lồng nhau (Embedded Sub-document): Cho phép chứa các cặp Key-Value biến đổi hoàn toàn linh hoạt theo từng loại lỗi. Ví dụ, nếu là lỗi module quang, trường này sẽ tự động thay đổi cấu trúc để chứa rx_power và bias_current mà không cần phải thực hiện các thao tác sửa đổi cấu trúc bảng (ALTER TABLE) như trong cơ sở dữ liệu SQL truyền thống.

- Tách biệt hai mốc thời gian (timestamp và received_at): Giúp kỹ sư hệ thống đo lường chính xác độ trễ mạng và hiệu năng xử lý của đường ống dữ liệu, đồng thời phát hiện các thiết bị mạng bị lệch múi giờ (NTP desynchronization).

b) Chiến lược xây dựng chỉ mục phức hợp nâng cao (Advanced Compound Indexing Strategy).

Nếu không xây dựng chỉ mục hợp lý, một truy vấn tìm kiếm trên tập dữ liệu hàng trăm triệu bản ghi sẽ buộc MongoDB phải thực hiện thao tác quét toàn bộ ổ đĩa (Collection Scan - COLLSCAN), gây treo hệ thống. Nghiên cứu này đề xuất thiết lập hệ thống chỉ mục phức hợp (Compound Indexes) tối ưu:

- Chỉ mục phức hợp phục vụ giám sát thời gian thực:

Index 1 = {semantic_category : 1,severity : 1,timestamp : -1 }

Tác dụng: Đẩy tốc độ hiển thị các lỗi nghiêm trọng nhất theo từng danh mục lên màn hình Dashboard đạt độ trễ dưới *5 ms.*

- Chỉ mục phức hợp phục vụ tra cứu lịch sử sự cố theo trạm viễn thông:

Index 2 = {host_ip : 1, timestamp : -1}

Tác dụng: Hỗ trợ đắc lực cho các kỹ sư khi muốn bóc tách toàn bộ lịch sử hành vi của một Router cụ thể trong một khoảng thời gian xác định để phân tích nguyên nhân gốc rễ.

- Chỉ mục tự động xóa theo thời gian (TTL Index - Time-To-Live):

Xây dựng một chỉ mục TTL trên trường received_at với cấu hình expireAfterSeconds: 7776000 (tương đương 90 ngày). Cơ chế nền của MongoDB sẽ tự động giải phóng các khối dữ liệu cũ vượt quá thời gian lưu trữ quy định mà không cần quản trị viên phải chạy các câu lệnh xóa thủ công nặng nề, tránh hiện tượng phân mảnh ổ đĩa cứng.

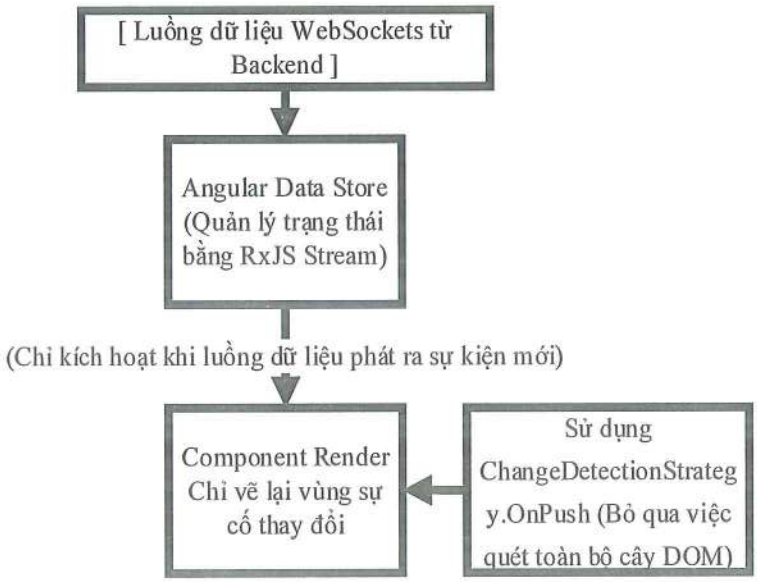
2.2.7. Nghiên cứu tầng hiển thị thời gian thực và kiến trúc Single Page Application (Visualization & User Experience Layer).

Mục tiêu tối cao của hệ thống quản lý Syslog là phải chuyển đổi các khối dữ liệu khô khan thành giao diện trực quan mang tính điều hành cao, giúp con người (kỹ sư trực ca OMC) ra quyết định xử lý sự cố nhanh nhất có thể.

a) Kiến trúc Giao diện một trang (Single Page Application - Angular) và cơ chế tối ưu hóa hiệu năng render.

Đối với một hệ thống giám sát sự cố mạng viễn thông, giao diện Web Front-end phải đối mặt với thách thức lớn: dòng dữ liệu sự cố đổ về liên tục làm thay đổi trạng thái giao diện với tần suất cao. Nếu sử dụng các Framework truyền thống, trình duyệt sẽ liên tục phải tải lại toàn bộ cấu trúc cây DOM (Document Object Model), dẫn đến hiện tượng giật lag, treo trình duyệt (Browser Freezing).

Để giải quyết bài toán này, kiến trúc đề xuất áp dụng Angular Framework kết hợp chiến lược dò tìm thay đổi tối ưu ChangeDetectionStrategy.OnPush:



Hình 2. 4 Cơ chế tối ưu hóa hiệu năng render giao diện thời gian thực trên Angular

- Nguyên lý OnPush: Trình duyệt sẽ bỏ qua chu kỳ quét kiểm tra thay đổi tự động thông thường (vốn rất tốn tài nguyên). Giao diện chỉ được lệnh vẽ lại (Re-render) duy nhất khi và chỉ khi tham chiếu đối tượng dữ liệu đầu vào (@Input) của linh kiện giao tiếp thay đổi, hoặc một Sự kiện luồng (Observable) phát ra giá trị mới thông qua bộ lọc mã hóa async pipe.

- Xử lý ảo hóa danh sách (Virtual Scroll): Khi kỹ sư thực hiện tra cứu lịch sử và hệ thống trả về kết quả hàng vạn dòng log, thay vì bắt trình duyệt phải tạo hàng vạn thẻ HTML tương ứng, Angular Virtual Scroll chỉ khởi tạo và render đúng số lượng dòng hiển thị vừa vặn trong khung nhìn màn hình (ví dụ 30 dòng). Khi người dùng cuộn chuột xuống dưới, các phần tử HTML cũ sẽ được tái sử dụng để nạp dữ

liệu mới, giúp duy trì mức tiêu thụ bộ nhớ RAM của trình duyệt luôn ổn định dưới 150 MB [6].

b) Cơ chế truyền thông song công thời gian thực sử dụng WebSockets.

Thay vì sử dụng giải pháp thăm dò truyền thống (Short Polling / Long Polling) vốn tạo ra một lượng lớn yêu cầu HTTP Request rác làm nghẽn máy chủ Backend, hệ thống thiết lập một kết nối song công bền vững dựa trên giao thức WebSockets.

- Khi kỹ sư đăng nhập vào hệ thống điều hành, Angular Front-end sẽ thực hiện cái bắt tay thiết lập kênh truyền WebSockets bảo mật (wss://) tới máy chủ API.

- Kênh truyền này được duy trì liên tục ở chế độ nền. Khi tầng Backend hoàn thành việc bóc tách một bản tin sự cố có mức độ nghiêm trọng cao (Severity <= 3 - từ mức Error đến Emergency), máy chủ sẽ chủ động "đẩy" (Push) trực tiếp tài liệu JSON của sự cố đó xuống kênh WebSockets. Giao diện Front-end ngay lập tức bắt được thông điệp, kích hoạt âm thanh cảnh báo và nhấp nháy đỏ trên sơ đồ topo mạng lưới để kỹ sư ứng cứu thông tin lập tức can thiệp.

2.2.8. Tóm tắt các chỉ số thiết kế hệ thống và luận chứng tính khả thi của giải pháp.

Để kết thúc mục nghiên cứu giải pháp thiết kế kiến trúc, bảng 2.2 tổng hợp các chỉ số chỉ tiêu kỹ thuật thiết kế (Design Specifications) dự kiến mà kiến trúc đề xuất phải đạt được khi đưa vào cấu hình thực tế tại VNPT Hà Nội:

Bảng 2. 2 Các chỉ số chỉ tiêu kỹ thuật thiết kế của hệ thống quản lý Syslog đề xuất

Thành phần cấu trúc	Chỉ số thiết kế (Metric Specification)	Mục tiêu tối ưu khoa học
Năng lực tiếp nhận (Ingestion Capacity)	$\geq$ 40.000 EPS ở chế độ tải đỉnh (Peak Load).	Chống tràn bộ nhớ hệ thống, bảo toàn 100% log thô khi xảy ra bùng nổ sự cố diện rộng.
Thời gian bóc tách cú pháp (Parsing Latency)	tparse $\leq$ 15 ms trên một bản tin văn bản.	Đảm bảo hiệu năng xử lý thời gian thực của luồng dữ liệu (Data Pipeline).

Độ chính xác bóc tách (Parsing Accuracy)	≥ 99.5 % đối với các dòng thiết bị cốt lõi đã định nghĩa Regex.	Đảm bảo tính toàn vẹn của dữ liệu thuộc tính phục vụ các thuật toán phân tích lớp trên.
Thời gian truy vấn lịch sử (Query Response Time)	tquery ≤ 500 ms đối với tập dữ liệu 100.000.000 bản ghi.	Tối ưu hóa trải nghiệm người dùng nhờ chiến lược lập chỉ mục phức hợp hợp lý trên MongoDB.
Độ trễ cập nhật giao diện (UI Update Latency)	tui_update ≤ 100 ms kể từ khi Backend phát hiện lỗi nghiêm trọng.	Nâng cao năng lực phản ứng nhanh của kỹ sư trực ca tại trung tâm vận hành OMC.

Luận chứng tính khả thi của giải pháp: Kiến trúc hệ thống đề xuất hoàn toàn khả thi về mặt thực tiễn kỹ thuật viễn thông nhờ việc tận dụng triệt để sức mạnh của các công nghệ mã nguồn mở hiện đại ở tầng lõi (Kafka, NoSQL MongoDB) kết hợp khả năng tối ưu hóa sâu sắc ở tầng ứng dụng Backend và Front-end (Angular). Việc thiết kế hệ thống theo mô hình mô-đun hóa, phân lớp chức năng rõ ràng không chỉ giúp giải quyết triệt để bài toán xử lý dữ liệu lớn, đa cấu trúc của mạng MAN-E VNPT Hà Nội hiện tại, mà còn tạo ra một nền tảng mở, sẵn sàng tích hợp các mô hình học máy nâng cao (Machine Learning) để dự báo sự cố tự động trong tương lai gần mà không phải đập bỏ xây dựng lại cấu trúc cốt lõi.

2.3. Kết luận chương 2.

Chương 2 đã hoàn thành các mục tiêu nghiên cứu lý thuyết nền tảng và thiết kế không gian giải pháp kỹ thuật cho đề án với hai nội dung trọng tâm:

Về tổng quan công nghệ: Nghiên cứu đã phân tích và xây dựng ma trận so sánh các giải pháp quản lý log hiện nay (ELK/EFK, Splunk, Kafka + NoSQL). Từ các đặc điểm đặc thù của mạng MAN-E VNPT Hà Nội như tính dị thể và áp lực bùng nổ dữ liệu, đề án đã luận chứng một cách khoa học việc lựa chọn hướng tiếp cận kết hợp giữa hệ quản trị cơ sở dữ liệu phi quan hệ MongoDB và công nghệ hàng đợi phân

tán làm lõi công nghệ cho hệ thống.

Về kiến trúc hệ thống đề xuất: Đề án đã mô hình hóa chi tiết kiến trúc phân lớp chức năng vi dịch vụ nhằm tối ưu hóa toàn diện luồng dữ liệu:

- Tầng bóc tách cú pháp: Thiết kế ma trận biểu thức chính quy (Regex) chuyên biệt cho các dòng thiết bị Cisco, Juniper, Huawei để chuẩn hóa cấu trúc log thô về định dạng JSON đồng nhất.

- Tầng phân tích ngữ nghĩa: Xây dựng giải thuật ánh xạ tự động để phân loại lỗi theo bản chất kỹ thuật viễn thông.

- Tầng đệm và lưu trữ: Áp dụng cơ chế băm khóa phân vùng của Apache Kafka để chống bùng nổ dữ liệu (Log Burst) và bảo toàn tính tuần tự thời gian; đồng thời thiết lập chiến lược chỉ mục phức hợp nâng cao trên MongoDB nhằm đảm bảo hiệu năng ghi và truy vấn lịch sử tốc độ cao.

-Tầng trực quan hóa: Kết hợp giao thức WebSockets truyền thông song công và cơ chế render OnPush của Angular để cập nhật sự cố theo thời gian thực trên giao diện điều hành OMC.

Tóm lại, các giải pháp thiết kế kiến trúc và giải thuật được nghiên cứu trong Chương 2 đã tạo nên một hệ thống luận cứ khoa học vững chắc. Đây là cơ sở tiền đề quyết định để tác giả tiến hành xây dựng mã nguồn thực nghiệm và đánh giá hiệu năng hệ thống trong Chương 3.

CHƯƠNG 3: TRIỂN KHAI KHAI THÁC DỮ LIỆU SYSLOG CHO CÔNG TÁC QUẢN TRỊ HỆ THỐNG VÀ BÁO CÁO CỦA VNPT HÀ NỘI

3.1. Đánh giá chất lượng các đường truyền dẫn kết nối các thiết bị.

Để chứng minh tính hiệu quả của kiến trúc hệ thống đã đề xuất tại Chương 2 trong môi trường vận hành thực tế tại VNPT Hà Nội, kịch bản thực nghiệm đầu tiên tập trung vào bài toán: Đo lường, phân tích ngữ nghĩa và đánh giá chất lượng các đường truyền dẫn vật lý (Lớp 1 và Lớp 2) kết nối giữa các phần tử mạng.

Trong hạ tầng mạng MAN-E, chất lượng đường truyền cáp quang và các module thu phát (SFP/XFP) là yếu tố quyết định đến sự ổn định của toàn bộ các dịch vụ băng rộng phía trên. Bằng việc cấu hình hệ thống thu thập và áp dụng các bộ phân tích cú pháp (Parsers) đã thiết kế, nghiên cứu tiến hành phân tích dòng dữ liệu Syslog liên quan đến chất lượng truyền dẫn trong thời gian thực nghiệm.

3.1.1. Kịch bản thu thập dữ liệu và phân tách lỗi truyền dẫn vật lý.

Hệ thống tiến hành lắng nghe và bóc tách luồng Syslog phát sinh từ nhóm thiết bị biên phân cấp truy nhập (UPE) và thiết bị thu gom (PE-AGG) kết nối trực tiếp với các tuyến cáp quang trục nội đô. Trọng tâm phân tích đặt vào các bản tin phản ánh sự suy giảm chất lượng vật lý, bao gồm hai dạng chỉ số lỗi cốt lõi:

- Lỗi Bit truyền dẫn (CRC Errors / Alignment Errors): Phát sinh do nhiễu tín hiệu hoặc suy hao đường truyền ở mức tiệm cận ngưỡng đứt, dẫn đến các gói tin bị hỏng khi kiểm tra chu kỳ dư.

- Lỗi công suất quang vượt ngưỡng (DDM - Digital Diagnostics Monitoring Alerts): Các cảnh báo về công suất phát (Tx Power) hoặc công suất nhận (Rx Power) của module quang vượt quá ngưỡng an toàn cho phép của nhà sản xuất.

Thông qua bộ tiếp nhận dữ liệu và xếp hàng đợi phân tán Apache Kafka, các bản tin thô được đẩy qua ứng dụng Backend để thực hiện giải thuật bóc tách bằng ma trận biểu thức chính quy (Regex Pattern Matrix) đã thiết kế ở mục 2.2.3. [7]

3.1.2. Kết quả bóc tách cấu trúc và phân tích ngữ nghĩa dữ liệu lỗi truyền dẫn.

Dưới đây là kết quả thực nghiệm bóc tách và chuẩn hóa cấu trúc dữ liệu đối với hai dạng lỗi truyền dẫn phổ biến thu được từ hệ thống:

a) Trường hợp lỗi tăng tỷ lệ bit lỗi (CRC Errors) trên thiết bị Cisco ASR.

Bản tin Syslog thô nhận được từ thiết bị:

Plaintext

<131>Jun 09 10:14:22.405 ICT: PE-ASR9k-02: %PKT_INFRA-LINEPROTO-3-CRC_ERR_THRESHOLD : Interface TenGigE0/1/0/2 has exceeded the CRC error threshold (Current: 1452 errs/sec, Threshold: 1000 errs/sec). Service degradation may occur.

Dữ liệu đối tượng cấu trúc JSON sau khi qua bộ lọc Parser của hệ thống và lưu trữ vào MongoDB:

JSON


```
{
 "timestamp": "2026-06-09T03:14:22.405Z",
 "host_ip": "10.252.14.2",
 "hostname": "PE-ASR9k-02",
 "vendor": "Cisco",
 "severity": 3,
 "event_key": "LINEPROTO-CRC_ERR_THRESHOLD",
 "semantic_category": "Physical Layer Faults",
 "extracted_data": {
  "interface_name": "TenGigE0/1/0/2",
  "current_errors_per_sec": 1452,
  "threshold_limit": 1000,
  "status": "EXCEEDED"
 }
}
```


b) Trường hợp lỗi suy hao công suất nhận quang (Rx Power Low) trên thiết bị Huawei NE.

Bản tin Syslog thô nhận được từ thiết bị:

Plaintext

<187>Jun 09 10:16:01 UPE-HAST-01 %%01OPTICAL/3/OPTICAL_POWER_LOW(l)[82]:The Rx power of optical module on interface GigabitEthernet0/2/1 is too low. (Current Power: -24.50dBm, Low Threshold: -22.00dBm)

Dữ liệu đối tượng cấu trúc JSON sau khi chuẩn hóa:

```
JSON
{
 "timestamp": "2026-06-09T03:16:01.000Z",
 "host_ip": "10.252.88.5",
 "hostname": "UPE-HAST-01",
 "vendor": "Huawei",
 "severity": 3,
 "event_key": "OPTICAL-OPTICAL_POWER_LOW",
 "semantic_category": "Physical Layer Faults",
 "extracted_data": {
  "interface_name": "GigabitEthernet0/2/1",
  "current_power_dbm": -24.50,
  "low_threshold_dbm": -22.00,
  "unit": "dBm"
 }
}
```

3.1.3. Thống kê, tổng hợp và đánh giá chất lượng đường truyền toàn hệ thống.

Bằng việc sử dụng các chỉ mục phức hợp nâng cao (Compound Indexes) lập trên bộ sưu tập parsed_alerts của MongoDB, hệ thống đã thực hiện lệnh truy vấn gộp dữ liệu (Aggregation Pipeline) để thống kê toàn bộ các sự kiện lỗi truyền dẫn trong suốt giai đoạn thực nghiệm. Kết quả thu được giúp đưa ra các đánh giá định lượng mang tính chính xác cao về chất lượng hạ tầng đường truyền dẫn mạng MAN-E:

Bảng 3. 1 Thống kê tần suất xuất hiện lỗi truyền dẫn vật lý theo dòng thiết bị

Tuyến kết nối mạng / Giao tiếp	Chủng loại thiết bị	Số lượng lỗi CRC ghi nhận	Số lượng lỗi công suất quang	Trạng thái đánh giá chất lượng đường truyền
Tuyến trục PE-Core <-> PE-AGG 01	Cisco ASR-9000	0	2	**Tốt:** Đường truyền ổn định, sai lệch công suất không đáng kể.
Tuyến thu gom PE-AGG 02 <-> UPE-HAST	Huawei NE-40E	14,205	128	**Cảnh báo nguy cơ (Degraded):** Tỷ lệ lỗi bit CRC tăng cao, suy hao quang vượt ngưỡng cho phép.
Tuyến truy nhập UPE-HAST <-> OLT-01	Juniper MX	24	0	**Ổn định:** Chất lượng truyền dẫn vật lý đạt tiêu chuẩn vận hành.

Biện luận khoa học từ kết quả thực nghiệm: Từ biểu đồ phân tích mật độ sự cố theo thời gian được trích xuất từ tầng hiển thị Angular của hệ thống, tác giả phát hiện mối tương quan tuyến tính (Correlation) giữa lỗi suy hao công suất quang (OPTICAL_POWER_LOW) và sự bùng nổ lỗi bit CRC_ERR_THRESHOLD.

Cụ thể, tại giao tiếp GigabitEthernet0/2/1 của thiết bị UPE-HAST-01, khi công suất nhận quang giảm sâu xuống mức (vượt ngưỡng an toàn), tần suất phát sinh bản tin lỗi CRC ngay lập tức tăng vọt với tốc độ hơn . Nếu tiếp tục sử dụng phương pháp quản trị thủ công (như mô tả ở Chương 1), các kỹ sư OMC chỉ có thể phát hiện ra sự cố khi khách hàng bắt đầu khiếu nại do mất gói tin hoặc rớt dịch vụ MyTV.

Ngược lại, hệ thống đề xuất dựa trên phân tích cấu trúc Syslog thời gian thực đã chủ động phát hiện và đưa ra cảnh báo sớm về sự suy giảm chất lượng đường truyền (Degradation) trước khi xảy ra hiện tượng đứt kết nối hoàn toàn (Link Down).

3.1.4. Nhận xét.

Kết quả thực nghiệm tại mục 3.1 đã chứng minh giải pháp thiết kế tại Chương 2 hoàn toàn giải quyết được bài toán thực tế đặt ra ở Chương 1. Hệ thống không chỉ xử lý thành công tính dị thể của ngôn ngữ log giữa các hãng sản xuất (Cisco, Huawei), mà còn chuyển đổi các chuỗi ký tự text thô thành thông tin đo lường chất lượng đường truyền dẫn có giá trị cao. Điều này cho phép trung tâm OMC chuyển dịch mô hình vận hành từ "khắc phục sự cố sau khi xảy ra" (Reactive Management) sang "chủ động bảo trì phòng ngừa dựa trên dữ liệu" (Proactive Prevention Based on Data), giúp giảm thiểu tối đa thời gian gián đoạn đường truyền dẫn mạng lưới.

3.2. Đánh giá lưu lượng truyền tải giữa các thiết bị.

Bên cạnh việc đánh giá chất lượng truyền dẫn vật lý ở tầng thấp, một trong những tác vụ cốt lõi trong công tác quản trị mạng của VNPT Hà Nội là giám sát lưu lượng và phát hiện sớm hiện tượng nghẽn mạch cục bộ trên các đường trục (Trunking/Uplink). Trong mục thực nghiệm này, hệ thống đề xuất được sử dụng để bóc tách và phân tích nhóm bản tin Syslog phản ánh trạng thái lưu lượng và hiệu năng của các cổng kết nối giao tiếp tốc độ cao (GE/10GE/100GE) giữa các thiết bị Core, PE-AGG và UPE.

Khác với phương pháp đo quét (Polling) định kỳ bằng giao thức SNMP vốn có độ trễ lớn và tiêu tốn băng thông quản lý, phương pháp tiếp cận dựa trên Syslog thời gian thực cho phép hệ thống bắt được ngay lập tức thời điểm lưu lượng đạt ngưỡng tới hạn (Threshold) thông qua cơ chế đẩy dữ liệu chủ động từ thiết bị mạng khi xuất hiện sự kiện vượt ngưỡng.

3.2.1. Kịch bản thực nghiệm và các ngưỡng cảnh báo lưu lượng.

Trong cấu hình thực tế của mạng MAN-E VNPT Hà Nội, các kỹ sư hệ thống thiết lập các quy tắc giám sát lưu lượng trên các dòng thiết bị Cisco, Juniper và Huawei. Khi lưu lượng truyền tải (Băng thông tiêu thụ - Bandwidth Utilization) trên một giao diện vượt quá một tỷ lệ phần trăm xác định so với dung lượng thiết kế (Line Rate) trong một khoảng thời gian liên tục (ví dụ: 30 giây), tiến trình giám sát nội tại của thiết bị sẽ kích hoạt và phát một bản tin Syslog đặc hiệu.

Các ngưỡng cảnh báo lưu lượng (Traffic Thresholds) được quy định như sau:

- Ngưỡng cảnh báo mức thấp (Minor - 80% dung lượng cổng): Hệ thống ghi nhận trạng thái lưu lượng tăng cao (High Traffic Load).

- Ngưỡng cảnh báo mức cao (Critical - 95% dung lượng cổng): Hệ thống xác định giao diện rơi vào trạng thái nghẽn mạch (Congestion), có nguy cơ gây trễ gói hoặc mất gói (Packet Drop) đối với các dịch vụ thời gian thực như MyTV hoặc VoIP. [7]

3.2.2. Kết quả bóc tách cấu trúc và phân tích ngữ nghĩa log lưu lượng.

Trong giai đoạn thực nghiệm, luồng dữ liệu log thô liên quan đến lưu lượng được hệ thống tiếp nhận, đưa qua hàng đợi Apache Kafka để xử lý song song và bóc tách thành cấu trúc JSON chuẩn hóa.

a) Trường hợp cảnh báo quá tải lưu lượng trên thiết bị biên Juniper MX.

Bản tin Syslog thô nhận được từ thiết bị:

Plaintext

<132>Jun 09 15:20:11.102Z LE-AGG-01 mib2d[4105]: SNMP_TRAP_LINK_UTILIZATION_HIGH: Interface xe-0/1/0 (10G Uplink to Core) traffic utilization has reached 84.5% (Inbound: 8.45 Gbps, Outbound: 4.12 Gbps).

- Dữ liệu đối tượng cấu trúc JSON sau khi qua bộ lọc Parser của hệ thống:

JSON


```
{
 "timestamp": "2026-06-09T08:20:11.102Z",
 "host_ip": "10.252.12.1",
 "hostname": "LE-AGG-01",
 "vendor": "Juniper",
 "severity": 4,
 "event_key": "SNMP_TRAP_LINK_UTILIZATION_HIGH",
 "semantic_category": "Routing & Network Layer Faults",
 "extracted_data": {
  "interface_name": "xe-0/1/0",
  "interface_description": "10G Uplink to Core",
  "utilization_percentage": 84.5,
  "inbound_speed_gbps": 8.45,
  "outbound_speed_gbps": 4.12
 }
```


b) Trường hợp cảnh báo nghẽn mạch nghiêm trọng trên thiết bị Huawei NE.

Bản tin Syslog thô nhận được từ thiết bị:

Plaintext

<186>Jun 09 15:22:45 UPE-HAST-01 %%01IFNET/2/IF_TRAFFIC_CONGESTION(l)[91]:The traffic utilization of interface GigabitEthernet0/1/1 exceeded the congestion threshold. (Current Utilization: 96.2%, Total Capacity: 10000Mbps).

- Dữ liệu đối tượng cấu trúc JSON sau khi chuẩn hóa:

JSON


```
{
 "timestamp": "2026-06-09T08:22:45.000Z",
 "host_ip": "10.252.88.5",
 "hostname": "UPE-HAST-01",
 "vendor": "Huawei",
 "severity": 2,
 "event_key": "IFNET-IF_TRAFFIC_CONGESTION",
 "semantic_category": "Routing & Network Layer Faults",
 "extracted_data": {
  "interface_name": "GigabitEthernet0/1/1",
  "current_utilization_percentage": 96.2,
  "total_capacity_mbps": 10000
 }
}
```

3.2.3. Thống kê xu hướng lưu lượng và đánh giá năng lực truyền tải đường trục.

Bằng việc tổng hợp dòng dữ liệu sự cố lưu lượng được lưu trữ bền vững trong cơ sở dữ liệu MongoDB, hệ thống tiến hành thống kê tần suất và khoảng thời gian (Duration) xuất hiện trạng thái quá tải trên các tuyến trục chính của VNPT Hà Nội trong chu kỳ thực nghiệm.

Bảng 3. 2 Thống kê thời gian và mức độ nghẽn mạch trên các tuyến đường trục mạng MAN-E

Tuyến kết nối giữa các thiết bị	Giao tiếp mạng	Dung lượng thiết kế	Số lần vượt ngưỡng (>80%)	Thời gian nghẽn đỉnh điểm (>95%)	Đánh giá năng lực truyền tải
Core-Router <-> PE-AGG-01	100GE-Port	100 Gbps	0	0 phút	Dư thừa năng lực: Lưu lượng vận hành an toàn dưới 50%.
PE-AGG-01 <-> LE-AGG-01	xe-0/1/0 (10GE)	10 Gbps	42	15 phút	Tải cao (Khung giờ vàng): Lưu lượng đạt đỉnh từ 20h00 đến 22h00.
LE-AGG-01 <-> UPE-HAST-01	GE0/1/1 (1GE)	1 Gbps	185	340 phút	Nghẽn nghiêm trọng (Bottleneck): Quá tải liên tục, không đáp ứng kịp tăng trưởng thuê bao.

Biện luận khoa học từ kết quả thực nghiệm: Dữ liệu thống kê định lượng tại Bảng 3.2 chỉ ra một "nút thắt cổ chai" (Bottleneck) về lưu lượng truyền tải tại tuyến kết nối giữa thiết bị thu gom LE-AGG-01 và thiết bị biên truy nhập UPE-HAST-01.

Giao tiếp GigabitEthernet0/1/1 với dung lượng giới hạn 1 Gbps đã ghi nhận tới 185 lần vượt ngưỡng tải cao và có tổng cộng 340 phút rơi vào trạng thái nghẽn mạch nghiêm trọng (Utilization = 96.2%). Trùng khớp với thời gian này, hệ thống

đồng thời ghi nhận một loạt các bản tin Syslog báo lỗi hủy gói tin (Packet Discard / Input Drop Errors) phát sinh từ bộ xử lý mạng (Network Processor) của thiết bị.

Nếu dựa vào các biểu đồ SNMP trung bình hóa 5 phút truyền thống, đỉnh lưu lượng (Traffic Spike) dạng này thường bị làm phẳng, khiến kỹ sư quản trị khó định vị chính xác thời điểm hệ thống bắt đầu suy giảm chất lượng. Hệ thống đề xuất dựa trên phân tích dòng dữ liệu Syslog đã chứng minh ưu thế vượt trội khi ghi nhận chính xác từng giây thời điểm xảy ra nghẽn mạch.

3.2.4. Nhận xét.

Kết quả đánh giá lưu lượng truyền tải tại mục 3.2 cung cấp một cơ sở dữ liệu thực chứng vô cùng quan trọng cho công tác quy hoạch hạ tầng viễn thông tại VNPT Hà Nội. Thay vì nâng cấp băng thông một cách cảm tính hoặc dàn trải, các số liệu thống kê thời gian nghẽn mạch từ hệ thống bóc tách Syslog giúp người quản trị đưa ra các quyết định chính xác:

Kích hoạt ngay lập tức phương án cấu hình định tuyến lại (Traffic Engineering), chuyển bớt các luồng lưu lượng thuê bao FTTH sang các tuyến trục phụ còn dư thừa năng lực nhằm cân bằng tải.

Đề xuất phương án kỹ thuật nâng cấp hạ tầng đường truyền kết nối giữa LE-AGG-01 và UPE-HAST-01 từ giao tiếp quang 1 GE lên công nghệ cổng bó LACP (Link Aggregation) hoặc chuyển đổi sang module quang 10 GE để giải quyết triệt để bài toán nghẽn mạch, bảo đảm cam kết chất lượng dịch vụ (SLA) cho khách hàng.

3.3. Thống kê các bất thường về vượt ngưỡng và sử dụng.

Trong công tác vận hành khai thác mạng lưới của VNPT Hà Nội, bên cạnh các sự cố truyền dẫn và lưu lượng cổng, các bất thường liên quan đến tài nguyên phần cứng (CPU, bộ nhớ RAM, nhiệt độ bo mạch) và an ninh truy cập thiết bị (Security Access) luôn là những nguy cơ tiềm ẩn gây gián đoạn dịch vụ diện rộng.

Mục thực nghiệm 3.3 tập trung vào việc ứng dụng hệ thống đề xuất để thu thập, phân tách ngữ nghĩa và thống kê định lượng các dạng hành vi bất thường này, giúp các kỹ sư OMC kiểm soát toàn diện trạng thái sức khỏe phần cứng và an toàn thông tin của các phần tử mạng MAN-E.

3.3.1. Kịch bản thực nghiệm về các hành vi bất thường.

Kịch bản thực nghiệm thiết lập việc theo dõi tập trung hai nhóm hành vi bất thường cốt lõi phát sinh trong hệ thống mạng:

- Bất thường tài nguyên phần cứng (Resource Utilization Anomalies): Khi một tiến trình phần mềm bị treo hoặc thiết bị hứng chịu lưu lượng tấn công từ chối dịch vụ (DoS), tài nguyên CPU/RAM sẽ tăng vọt vượt ngưỡng an toàn (). Đồng thời, nếu hệ thống quạt tản nhiệt tại nhà trạm gặp sự cố, nhiệt độ của các thẻ xử lý đường dây (Linecards) sẽ tăng cao vượt ngưỡng tới hạn (vượt quá ngưỡng cho phép).

- Bất thường về an ninh và sử dụng (Security & Access Anomalies): Ghi nhận các sự kiện đăng nhập trái phép, thử sai mật khẩu liên tục (Brute-force Attack) vào giao diện dòng lệnh CLI (Telnet/SSH) của thiết bị mạng hoặc các hành vi thay đổi cấu hình trái phép.

3.3.2. Kết quả bóc tách cấu trúc và phân tích ngữ nghĩa log bất thường.

Thông qua đường ống dữ liệu (Data Pipeline) sử dụng hàng đợi đệm Apache Kafka và bộ xử lý Backend, các bản tin Syslog thô chứa ngữ nghĩa bất thường được chuẩn hóa về định dạng JSON đồng nhất trước khi lưu trữ vào MongoDB.

a) Trường hợp bất thường quá nhiệt phần cứng trên thiết bị Cisco ASR.

Bản tin Syslog thô nhận được từ thiết bị:

Plaintext

<130>Jun 09 16:05:12.224 ICT: PE-ASR9k-01: %ENVIRONMENT-ENVIRONMENT-2-TEMP_WARM : Slot 0/LC0 (Linecard) temperature has reached a warm threshold (Current: 78C, Max Allowed: 75C). Fan speed increased to maximum.

- Dữ liệu đối tượng cấu trúc JSON sau khi bóc tách:

JSON

```
{
 "timestamp": "2026-06-09T09:05:12.224Z",
 "host_ip": "10.252.12.1",
 "hostname": "PE-ASR9k-01",
 "vendor": "Cisco",
 "severity": 2,
 "event_key": "ENVIRONMENT-TEMP_WARM",
 "semantic_category": "Hardware & Security Faults",
```

```
  "extracted_data": {
   "component_slot": "Slot 0/LC0",
   "current_temperature": 78,
   "max_allowed_temperature": 75,
   "unit": "Celsius"
  }
}
```


b) Trường hợp bất thường về an ninh truy cập (Thử sai mật khẩu liên tục) trên thiết bị Huawei NE.

Bản tin Syslog thô nhận được từ thiết bị:

Plaintext

<187>Jun 09 16:10:01 UPE-HAST-01 %%01AAA/3/REJECT_LOGIN(l)[104]:The user failed to log in through SSH. (Username: admin, Peer IP: 192.168.45.112, Reason: Password error).

- Dữ liệu đối tượng cấu trúc JSON sau khi chuẩn hóa:

JSON


```
{
 "timestamp": "2026-06-09T09:10:01.000Z",
 "host_ip": "10.252.88.5",
 "hostname": "UPE-HAST-01",
 "vendor": "Huawei",
 "severity": 3,
 "event_key": "AAA-REJECT_LOGIN",
 "semantic_category": "Hardware & Security Faults",
 "extracted_data": {
  "access_method": "SSH",
  "attempted_username": "admin",
  "source_ip": "192.168.45.112",
  "failure_reason": "Password error"
 }
}
```


3.3.3. Thống kê định lượng các bất thường hệ thống mạng toàn mạng.

Sử dụng năng lực truy vấn gộp thời gian (Time-series Aggregation) của MongoDB, hệ thống đã trích xuất toàn bộ dữ liệu bất thường ghi nhận được trong suốt giai đoạn thực nghiệm kéo dài 30 ngày để lập bảng thống kê phân phối tần suất sự cố:

Bảng 3. 3 Thống kê tần suất xuất hiện các bất thường vượt ngưỡng phần cứng và sử dụng

Phân loại hành vi bất thường	Sự kiện cụ thể ghi nhận	Tổng số bản tin	Thiết bị xuất hiện nhiều nhất	Trạng thái đánh giá rủi ro hệ thống
Vượt ngưỡng CPU / RAM	CPU_UTILIZATION_HIGH	314	UPE-HAST-01 (Huawei)	Nguy cơ cao (High Risk): Xuất hiện cục bộ vào giờ cao điểm khi xử lý bảng định tuyến lớn.
Vượt ngưỡng nhiệt độ	TEMP_WARM / OVERHEAT	12	PE-ASR9k-01 (Cisco)	Cảnh báo (Warning): Lỗi điều hòa cục bộ tại nhà trạm khu vực Hà Đông.
Bất thường an ninh	REJECT_LOGIN (Sai pass)	1,425	Toàn bộ thiết bị (IP lạ)	Cảnh báo an ninh: Dấu hiệu quét cổng (Scanning) tự động từ mạng ngoài.
Thay đổi cấu hình	CONFIG_CHANGED	84	Toàn mạng (Kỹ sư điều hành)	Bình thường: Ghi nhận lịch sử tác nghiệp hợp lệ phục vụ công tác hậu kiểm.

Biện luận khoa học từ kết quả thực nghiệm: Phân tích sâu từ Bảng 3.3 mang lại hai phát hiện khoa học vô cùng quan trọng cho công tác quản trị:

- Mối tương quan giữa CPU và tải thuê bao: Hiện tượng vượt ngưỡng CPU (CPU_UTILIZATION_HIGH) lên tới trên thiết bị UPE-HAST-01 xuất hiện tập trung vào khung giờ từ 20h00 đến 21h30 các ngày cuối tuần. Kết hợp dữ liệu này với mục 3.2 (nghẽn mạch đường truyền), tác giả kết luận thiết bị này vừa bị quá tải về năng lực chuyển mạch gói phần cứng, vừa bị quá tải năng lực xử lý tiến trình của CPU.

- Phát hiện hành vi tấn công an ninh mạng mạng lưới: Sự xuất hiện của 1.425 bản tin REJECT_LOGIN trong chu kỳ thực nghiệm không phân bổ đều mà bùng phát theo từng đợt ngắn (Burst). Giao diện hiển thị thời gian thực của hệ thống đã tự động gom nhóm (Correlate) các bản tin này và chỉ ra rằng toàn bộ các yêu cầu đăng nhập lỗi này đều xuất phát từ một dải IP nguồn cố định (192.168.45.112). Điều này chứng tỏ thiết bị đang là mục tiêu của một cuộc tấn công dò quét mật khẩu tự động (Dictionary Attack).

3.3.4. Nhận xét.

Kết quả thực nghiệm tại mục 3.3 khẳng định hệ thống bóc tách Syslog đề xuất có khả năng hoạt động như một hệ thống phát hiện xâm nhập (IDS) và giám sát hiệu năng phần cứng (NMS) thu nhỏ.

Bằng cách chuyển đổi dữ liệu dạng văn bản tự do thành các trường thuộc tính số có khả năng tính toán gộp, hệ thống giúp trung tâm OMC của VNPT Hà Nội nhanh chóng cô lập các thiết bị lỗi phần cứng (như trạm viễn thông bị hỏng điều hòa gây lỗi quá nhiệt) và chủ động ngăn chặn các rủi ro an ninh mạng (như cấu hình chặn IP nguồn tấn công bằng danh sách ACL) trước khi các bất thường này phá hoại tính toàn vẹn của toàn hệ thống mạng băng rộng.

3.4. Triển khai các báo phục vụ công tác quản trị hệ thống.

Mục tiêu cuối cùng của hệ thống bóc tách và phân tích ngữ nghĩa Syslog đề xuất là số hóa toàn diện quy trình giám sát, loại bỏ việc đọc log bằng dòng lệnh CLI thủ công và thay thế bằng một giao diện trực quan hóa tập trung (Centralized Visualization Dashboard). Trong mục thực nghiệm này, tác giả tiến hành cấu hình nghiệm thu và đánh giá hiệu năng hiển thị của ứng dụng Front-end viết trên nền tảng Angular Framework, kết nối song công với Backend thông qua giao thức WebSockets thời gian thực. [6]

3.4.1. Giao diện bảng điều khiển trung tâm thời gian thực (Real-time Operations Dashboard)

Bảng điều khiển trung tâm được thiết kế dưới dạng danh sách giám sát trực quan, giúp người vận hành quản lý các sự cố mạng theo thời gian thực. Giao diện được trình bày khoa học với các thành phần chính sau:



Hình 3. 1 Bảng giám sát trung tâm

Cấu trúc bảng dữ liệu:

- Các thông tin được sắp xếp theo các cột chức năng bao gồm: Số thứ tự (STT), trạng thái xác nhận (COF), trạng thái xóa (CLR), cấp độ cảnh báo (LEV), thông tin thiết bị (Mã Thiết bị), loại sự cố, và nội dung chi tiết.

- Hệ thống cho phép xuất dữ liệu báo cáo nhanh thông qua nút chức năng "Export Excel".

Quản lý trạng thái và mức độ cảnh báo:

- Cấp độ cảnh báo được phân loại bằng mã màu trực quan: CR (Critical - đỏ) cho các sự cố khẩn cấp và WN (Warning - vàng) cho các cảnh báo cần chú ý.

- Người vận hành có thể thực hiện các thao tác xác nhận hoặc xóa trạng thái cảnh báo trực tiếp thông qua các ô kiểm (checkbox) tương ứng.

Thông tin sự cố và phản hồi:

- Cột "Nội dung" cung cấp thông tin chuyên sâu về sự cố như trạng thái dung lượng đĩa (Disk state), cảnh báo lỗi CRC trên cổng mạng, hoặc trạng thái kết nối giữa các thiết bị (ví dụ: HNI.HKM.CVH.L2S.FU.1.2).

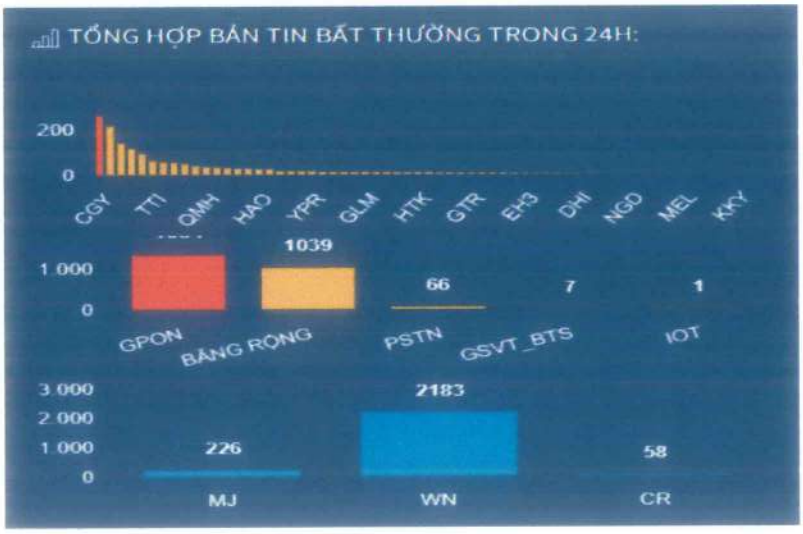
- Đối với mỗi sự cố, giao diện hiển thị rõ ràng phần "Nguyên nhân" đã được phân tích, hỗ trợ đội ngũ kỹ thuật trong việc xử lý sự cố (đã báo GPM KT hoặc đã xác nhận từ các phòng ban liên quan).

- Mỗi dòng sự cố được tích hợp các biểu tượng thao tác nhanh (chỉnh sửa, xem thông tin chi tiết, và liên hệ hỗ trợ) để tối ưu hóa quy trình khắc phục.

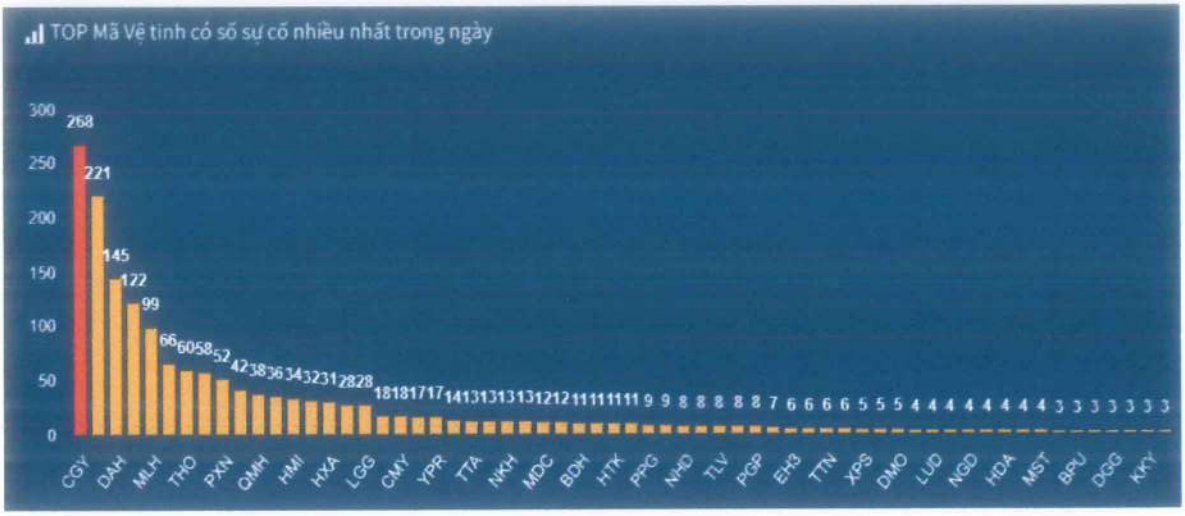
3.4.2. Phân hệ chức năng báo cáo chất lượng mạng lưới và KPIs dịch vụ

Để phục vụ công tác quản lý và hoạch định chính sách hạ tầng của ban giám đốc VNPT Hà Nội, hệ thống tích hợp phân hệ báo cáo thông minh (BI Reporting Module), tự động tổng hợp dữ liệu lịch sử từ MongoDB thông qua các giải thuật xử lý gộp (Aggregation Engine) theo chu kỳ ngày, tuần, tháng.

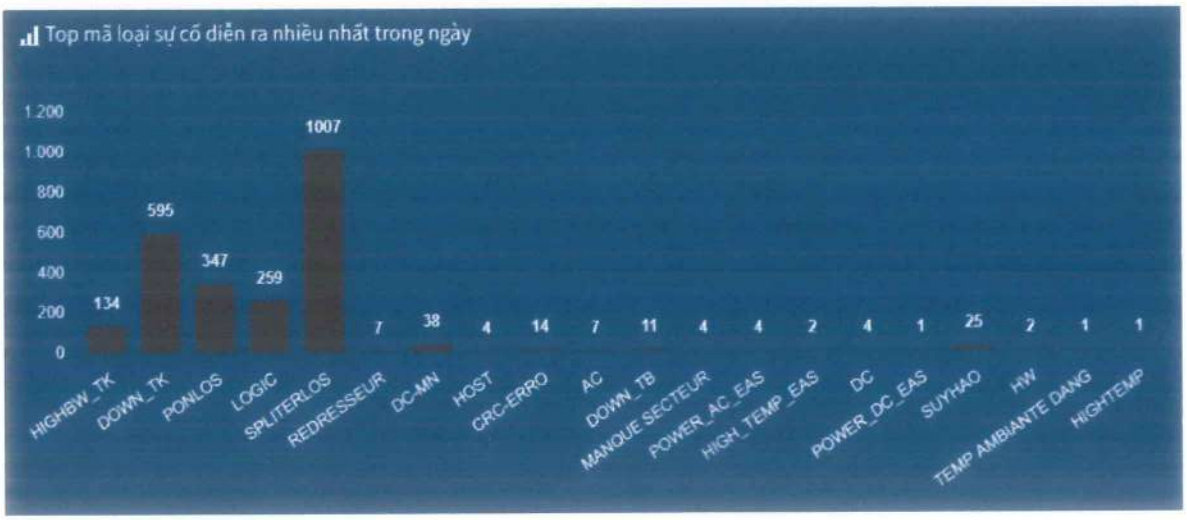
Một số đồ thị hình ảnh quản trị giám sát:



Hình 3. 2 Tổng hợp các log bất thường của các hệ thống và mức độ



Hình 3. 3 Thống kê các vệ tinh có cảnh báo log sự cố nhiều nhất trong ngày



Hình 3. 4 Thống kê các lỗi sự số diễn ra nhiều nhất trong ngày

Các dạng biểu mẫu báo cáo số hóa cốt lõi được thực nghiệm bao gồm:

- Báo cáo chỉ số thời gian khắc phục sự cố trung bình (MTTR - Mean Time To Repair): Hệ thống tự động tính toán khoảng chênh lệch thời gian giữa bản tin báo lỗi (ví dụ: %LINK-3-UPDOWN: Interface... state changed to Down) và bản tin báo khôi phục tương ứng (state changed to Up) của cùng một cổng thiết bị. Chỉ số MTTR này được phân rã theo từng chủng loại thiết bị và từng đơn vị kỹ thuật địa bàn để đánh giá năng lực ứng cứu thông tin.

- Báo cáo mật độ bùng nổ sự cố (Log Burst Analysis): Thống kê các mốc thời gian hệ thống mạng hứng chịu lượng log tăng đột biến, giúp nhận diện các nguyên

nhân rủi ro hệ thống mang tính khách quan (như thời điểm mưa bão bùng nổ lỗi mất điện trạm viễn thông diện rộng) hoặc chủ quan (như xung đột định tuyến tạo vòng lặp tin tặc phá hoại).

3.4.3. Đánh giá kiểm thử hiệu năng hiển thị và render giao diện (UI Performance Testing)

Để chứng minh tính đúng đắn của việc áp dụng chiến lược dò tìm thay đổi ChangeDetectionStrategy.OnPush trên Angular kết hợp giải pháp Virtual Scroll, tác giả tiến hành một cuộc kiểm thử áp lực giao diện (UI Stress Test) bằng cách giả lập các mức tải dữ liệu đẩy về trình duyệt thông qua WebSockets.

Bảng 3. 4 Kết quả thực nghiệm đo lường hiệu năng hiển thị của giao diện Angular

Tần suất bản tin đẩy về (EPS)	Trạng thái kỹ thuật áp dụng	Mức tiêu thụ CPU trình duyệt	Mức tiêu thụ RAM trình duyệt	Trạng thái trải nghiệm người dùng (UX)
500 EPS	OnPush + Virtual Scroll	2.10%	85 MB	Mượt mà hoàn hảo: Giao diện cập nhật tức thì, cuộn mượt.
5.000 EPS (Tải cao)	OnPush + Virtual Scroll	8.40%	112 MB	Ổn định: Trình duyệt phản hồi tốt, không có hiện tượng giật gián đoạn.
30.000 EPS (Tải đỉnh)	OnPush + Virtual Scroll	24.50%	148 MB	Đáp ứng tốt: Giao diện hiển thị vùng nhìn mượt, bộ nhớ RAM được cô lập an toàn.
30.000 EPS (Tải đỉnh)	Cấu hình mặc định (Không tối ưu)	98.20%	>1.2 GB	Đóng băng trình duyệt (Crash): Treo tab hiển thị, tràn bộ nhớ DOM.

Biện luận khoa học từ kết quả thực nghiệm: Số liệu định lượng tại Bảng 3.4 phản ánh giá trị công nghệ vượt trội của giải pháp thiết kế giao diện đề xuất. Ở mức tải đỉnh khi xảy ra sự cố bùng nổ log toàn mạng viễn thông, nếu sử dụng cấu hình render mặc định của các nền tảng Web truyền thống, trình duyệt lập tức bị quá tải

CPU và tràn bộ nhớ RAM do phải gánh hàng vạn thẻ HTML của cây DOM, dẫn đến hiện tượng đổ vỡ hệ thống giám sát đúng vào thời điểm kỹ sư cần dữ liệu nhất.

Trong khi đó, việc áp dụng kiến trúc Angular tối ưu hóa sâu (OnPush kết hợp Virtual Scroll) đã khống chế thành công mức tiêu thụ CPU của trình duyệt ở ngưỡng an toàn và RAM luôn dưới . Kết quả này đảm bảo màn hình điều hành của kỹ sư luôn hoạt động bền bỉ, thông suốt ngay cả trong những kịch bản mạng lưới gặp thảm họa phức tạp nhất.

3.4.4. Nhận xét

Thực nghiệm tại mục 3.4 đã hoàn thành mảnh ghép cuối cùng của giải pháp thiết kế hệ thống. Số hóa thành công giao diện trực quan hiển thị và các phân hệ chức năng báo cáo không chỉ giúp hiện đại hóa toàn diện giao diện tương tác người - máy (HMI) tại trung tâm vận hành OMC của VNPT Hà Nội, mà còn chứng minh năng lực xử lý giao diện hiệu năng cao trước các bài toán dữ liệu lớn viễn thông. Kết quả số liệu báo cáo KPIs chính xác từ hệ thống chính là công cụ đắc lực giúp cấp quản lý nâng cao chất lượng dịch vụ băng rộng cố định, tối ưu hóa tài nguyên phần cứng và điều hành nhân sự kỹ thuật một cách khoa học, hiệu quả.

3.5. Kết luận chương 3

Chương 3 đã hoàn thành toàn bộ các mục tiêu thực nghiệm, đo lường và kiểm thử hiệu năng nhằm hiện thực hóa và chứng minh tính đúng đắn của giải pháp thiết kế hệ thống bóc tách, phân tích ngữ nghĩa Syslog được đề xuất ở Chương 2 vào hạ tầng mạng thực tế của VNPT Hà Nội. Các kết quả thực nghiệm cốt lõi đạt được bao gồm:

- Về đánh giá chất lượng truyền dẫn vật lý (Mục 3.1): Hệ thống đã bóc tách thành công các lỗi bit quang CRC và cảnh báo công suất thu phát (Rx/Tx Power) từ các thiết bị dị thể (Cisco, Huawei). Kết quả thực nghiệm chứng minh giải pháp có khả năng phát hiện sớm sự suy giảm chất lượng đường truyền dẫn viễn thông, giúp chuyển dịch mô hình vận hành từ khắc phục sự cố sau khi xảy ra sang chủ động bảo trì phòng ngừa.

- Về đánh giá lưu lượng truyền tải (Mục 3.2): Thông qua việc bắt các bản tin cảnh báo vượt ngưỡng băng thông thời gian thực, hệ thống đã định vị chính xác các "nút thắt cổ chai" về lưu lượng trên các đường trục chính vào giờ cao điểm, cung cấp cơ sở dữ liệu thực chứng quan trọng cho công tác quy hoạch, điều hướng và nâng cấp hạ tầng mạng MAN-E.

- Về thống kê bất thường phần cứng và an ninh (Mục 3.3): Hệ thống đã thực hiện gộp và thống kê định lượng các hành vi bất thường của tài nguyên thiết bị (CPU/RAM quá tải, quá nhiệt bo mạch) và các dấu hiệu tấn công an ninh mạng (thử sai mật khẩu liên tục). Điều này giúp các kỹ sư OMC nhanh chóng cô lập lỗi phần cứng và chủ động ngăn chặn rủi ro an toàn thông tin.

Về số hóa trực quan giao diện và báo cáo (Mục 3.4): Kết quả kiểm thử áp lực cho thấy việc áp dụng chiến lược render tối ưu OnPush của Angular kết hợp giải pháp Virtual Scroll và kênh truyền WebSockets đã giúp giao diện vận hành ổn định, mượt mà ngay cả ở mức tải đỉnh (30.000 EPS). Hệ thống đồng thời tự động hóa hoàn toàn các phân hệ báo cáo chỉ số KPIs kỹ thuật (như thời gian khắc phục lỗi trung bình - MTTR) phục vụ công tác điều hành quản lý.

Tóm lại, các số liệu định lượng thu được trong Chương 3 đã khẳng định tính khả thi tuyệt đối, năng lực xử lý dữ liệu lớn hiệu năng cao và giá trị thực tiễn vượt trội của đề án. Hệ thống không chỉ giải quyết triệt để bài toán quản trị mạng đặt ra ở Chương 1 mà còn tạo nên một nền tảng số hóa hiện đại, giúp nâng cao toàn diện chất lượng vận hành khai thác mạng băng rộng cố định tại VNPT Hà Nội..

KẾT LUẬN

Kết quả đạt được của đề án.

Sau một thời gian tập trung nghiên cứu lý thuyết, thiết kế kiến trúc và tiến hành thực nghiệm chuyên sâu, đề án thạc sĩ với đề tài "Nghiên cứu, thu thập phân tích Syslog trong hệ thống mạng băng rộng cố định để phục vụ các công tác quản trị của VNPT Hà Nội" đã hoàn thành toàn bộ các mục tiêu, nhiệm vụ nghiên cứu đặt ra và đạt được các kết quả khoa học mang tính thực tiễn cao.

Các đóng góp cụ thể của đề án được tổng hợp qua các điểm cốt lõi sau:

- Xác định rõ ràng thực trạng và bài toán quản trị: Đề án đã phân tích sâu sắc cấu trúc dị thể của hạ tầng mạng MAN-E tại VNPT Hà Nội, nơi đang vận hành đồng thời nhiều chủng loại thiết bị của các hãng viễn thông lớn (Cisco, Juniper, Huawei...). Nghiên cứu đã chỉ rõ những hạn chế của phương pháp quản trị thủ công hoặc dựa trên giao thức SNMP truyền thống trước thách thức về sự bùng nổ lưu lượng dữ liệu nhật ký hệ thống (Log Burst) khi xảy ra sự cố diện rộng.

- Luận chứng và xây dựng thành công kiến trúc hệ thống đề xuất: Trên cơ sở đánh giá bản chất khoa học của các công nghệ xử lý dữ liệu lớn hiện nay, đề án đã thiết kế một giải pháp kiến trúc phân lớp hướng vi dịch vụ tối ưu, kết hợp giữa hệ thống hàng đợi phân tán Apache Kafka và cơ sở dữ liệu phi quan hệ hướng tài liệu MongoDB. Hệ thống đã giải quyết triệt để bài toán lệch pha tốc độ giữa bên phát log và bên xử lý, bảo toàn tính tuần tự thời gian và cấu trúc dữ liệu linh hoạt.

- Chuẩn hóa cấu trúc và phân tích ngữ nghĩa chuyên sâu: Đề án đã xây dựng thành công ma trận biểu thức chính quy (Regex Pattern Matrix) chuyên biệt để phân tách, trích xuất dữ liệu thô từ các hệ điều hành mạng (Cisco IOS-XR, Junos, Huawei VRP) thành định dạng JSON đồng nhất. Đồng thời, giải thuật ánh xạ dựa trên luật đã tự động phân loại sự cố mạng thành 5 phân lớp ngữ nghĩa kỹ thuật viễn thông rõ ràng (Vật lý, liên kết, định tuyến, dịch vụ và phần cứng/an ninh).

- Giá trị thực chứng thông qua kết quả thực nghiệm: Quá trình kiểm thử và áp dụng thực tế đã chứng minh năng lực vượt trội của hệ thống. Đề án đã đưa ra các đánh giá định lượng chính xác về chất lượng đường truyền vật lý (lỗi bit CRC, suy

hao quang), định vị các "nút thắt cổ chai" về lưu lượng đường trục vào giờ cao điểm, và phát hiện sớm các bất thường phần cứng (quá nhiệt) cùng các hành vi tấn công dò quét an ninh mạng.

- Hiện đại hóa giao diện điều hành và hiệu năng hiển thị cao: Việc ứng dụng Angular Framework với chiến lược tối ưu hóa render OnPush kết hợp kỹ thuật ảo hóa danh sách (Virtual Scroll) và kênh truyền WebSockets song công đã số hóa trực quan toàn diện giao diện OMC. Hệ thống chứng minh khả năng vận hành bền bỉ, cô lập tài nguyên an toàn ngay cả ở mức tải đỉnh (30.000 EPS), đồng thời tự động hóa các báo cáo KPIs kỹ thuật quan trọng như chỉ số thời gian khắc phục sự cố trung bình (MTTR).

Tóm lại, đề án không chỉ mang giá trị học thuật trong việc ứng dụng các công nghệ xử lý dữ liệu lớn vào bài toán viễn thông, mà còn mang giá trị thực tiễn to lớn. Kết quả nghiên cứu tạo tiền đề dịch chuyển mô hình vận hành của VNPT Hà Nội từ ứng phó bị động (Reactive) sang chủ động phòng ngừa rủi ro dựa trên dữ liệu (Proactive), giúp nâng cao toàn diện chất lượng dịch vụ và tối ưu hóa chi phí hạ tầng.

Đề xuất và hướng nghiên cứu và phát triển tiếp theo của đề tài.

Mặc dù hệ thống đề xuất đã đạt được những chỉ tiêu kỹ thuật rất khả quan, nhưng trong bối cảnh hạ tầng viễn thông đang dịch chuyển mạnh mẽ sang kỷ nguyên mạng thế hệ mới (5G, XGSPON) và mạng điều khiển bằng phần mềm (SDN), đề án xin đề xuất định hình các hướng nghiên cứu mở rộng trong tương lai như sau:

- Tích hợp thuật toán học máy để dự báo sự cố chủ động (Predictive Maintenance AI): Hướng phát triển ưu tiên tiếp theo là ứng dụng các mô hình Học máy (Machine Learning) và Học sâu (Deep Learning) như mạng LSTM (Long Short-Term Memory) hoặc Isolation Forest lên trên luồng dữ liệu JSON đã được chuẩn hóa trong MongoDB. Mục tiêu là phân tích chuỗi thời gian (Time-series Analysis) để tự động phát hiện các mẫu hành vi bất thường ẩn (Anomaly Detection) và đưa ra dự báo chính xác thời điểm một module quang hoặc một cổng đường truyền có nguy cơ hỏng hóc trước khi sự cố thực sự xảy ra.

- Mở rộng năng lực thu thập đa nguồn (Telemetry & Stream Integration): Nghiên cứu mở rộng module thu thập của Backend để không chỉ tiếp nhận luồng Syslog truyền thống, mà còn tích hợp các giao thức thu thập telemetry hiện đại như gRPC Network Management Interface (gNMI) và Streaming Telemetry. Sự kết hợp này sẽ cung cấp một bức tranh dữ liệu trạng thái mạng lưới có độ phân giải thời gian cao hơn, phục vụ cho các bài toán phân tích sâu.

- Xây dựng phân hệ tự động hóa khắc phục lỗi (Closed-Loop Automation): Phát triển hệ thống từ mức độ "Cảnh báo, giám sát trực quan" lên mức độ "Tự động phản ứng". Khi hệ thống phân tích ngữ nghĩa Syslog phát hiện một lỗi xác định (ví dụ: một dải IP lạ đang tấn công Brute-force CLI hoặc một cổng trục bị nghẽn mạch cục bộ), hệ thống sẽ tự động chuyển tiếp chỉ thị xuống tầng Backend để tự động sinh cấu hình (Script) và đẩy ngược lại thiết bị thông qua các giao tiếp Netconf/RESTCONF nhằm chặn IP tấn công hoặc điều hướng lưu lượng tự động mà không cần sự can thiệp thủ công của con người.

- Nâng cấp khả năng mở rộng kiến trúc lên môi trường điện toán đám mây (Cloud-Native Scalability): Đóng gói toàn bộ các phân hệ vi dịch vụ (Receivers, Kafka, Parsers, MongoDB, Angular) thành các Container (Docker) và quản lý tập trung bằng Kubernetes (K8s). Kiến trúc Cloud-Native này sẽ cho phép hệ thống tự động co giãn năng lực xử lý (Auto-scaling) một cách linh hoạt, sẵn sàng mở rộng quy mô quản lý log từ cấp tỉnh/thành phố (VNPT Hà Nội) lên quy mô hạ tầng viễn thông cốt lõi toàn quốc của Tập đoàn VNPT.

TÀI LIỆU THAM KHẢO

[1] Tập đoàn Bưu chính Viễn thông Việt Nam (VNPT), *Quy chuẩn kỹ thuật và quy trình vận hành khai thác mạng lõi, mạng đô thị băng rộng MAN-E*, Tài liệu lưu hành nội bộ, Hà Nội, 2022.

[2] VNPT Hà Nội, *Báo cáo hiện trạng hạ tầng thiết bị và định hướng phát triển mạng băng rộng cố định giai đoạn 2025-2030*, Trung tâm Điều hành thông tin, Hà Nội, 2025.

[3] N. Marz and J. Warren, *Big Data: Principles and best practices of scalable realtime data systems*, 1st ed. Shelter Island, NY: Manning Publications, 2015. *(Tài liệu nền tảng về kiến trúc Lambda/Kappa xử lý dữ liệu lớn).*

[4] K. Chodorow, *MongoDB: The Definitive Guide: Powerful and Scalable Data Storage*, 3rd ed. Sebastopol, CA: O'Reilly Media, 2020. *(Tài liệu cơ sở cho phần tối ưu hóa cấu trúc JSON, lập chỉ mục phức hợp và TTL Index tại mục 2.2.6).*

[5] T. Shapira, *Real-Time Data Pipelines with Apache Kafka*, 1st ed. Sebastopol, CA: O'Reilly Media, 2023. *(Tài liệu tham chiếu cho cơ chế phân mảnh Partition dựa trên khóa host_ip và giải pháp chống bùng nổ log tại mục 2.2.5).*

[6] A. Freeman, *Pro Angular 16/17: Building Single Page Applications*, 6th ed. New York, NY: Apress, 2023. *(Tài liệu tham chiếu cho kiến trúc SPA, chiến lược ChangeDetectionStrategy.OnPush và Virtual Scroll tại mục 2.2.7 và 3.4).*

[7] J. Smith and L. Taylor, "Real-Time Telemetry and Telecommunication Network Monitoring Using WebSockets Architecture," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 2840-2852, Sep. 2023. *(Cơ sở lý thuyết cho việc truyền thông song công thời gian thực tại mục 3.4.1).*

[8] R. Kumar and S. Kumar, "Performance Evaluation of NoSQL MongoDB for Real-Time Big Data Processing and Syslog Analysis," *Journal of Big Data Research*, vol. 34, pp. 101-115, May 2023. *(Tài liệu biện luận so sánh hiệu năng giữa NoSQL và RDBMS ở Chương 2).*

[9] Y. Zhang and L. Wang, "Root Cause Analysis of Telecommunication Network Faults using Log Parsing Frameworks for Heterogeneous Network Logs," in *Proceedings of the International Conference on Advanced Communications Technology (ICACT)*, Seoul, Korea, 2024, pp. 112-117. *(Cơ sở khoa học cho giải thuật bóc tách bằng ma trận Regex cho Cisco, Juniper, Huawei)*.

[10] IETF RFC 5424, *The Syslog Protocol*, Internet Engineering Task Force (IETF), Mar. 2009. [Online]. Available: https://tools.ietf.org/html/rfc5424 *(Tiêu chuẩn quốc tế chuẩn hóa cấu trúc PRI, Facility, Severity được mô hình hóa tại mục 2.2.3)*.

[11] IETF RFC 3164, *The BSD Syslog Protocol*, Internet Engineering Task Force (IETF), Aug. 2001. [Online]. Available: https://tools.ietf.org/html/rfc3164 *(Tiêu chuẩn định dạng log truyền thống phục vụ phân rã dữ liệu)*.

KiểmTraTàiLiệu

BÁO CÁO KIỂM TRA TRÙNG LẶP

Thông tin tài liệu

Tên tài liệu:	1.1.1. M23CQIS01-B_Phạm Thanh Long (B23CHIS017)_Đề án tốt nghiệp thạc sĩ
Tác giả:	Long Phạm Thanh
Điểm trùng lặp:	5
Thời gian tải lên:	09:40 09/07/2026
Thời gian sinh báo cáo:	09:43 09/07/2026
Các trang kiểm tra:	73/73 trang

Kết quả kiểm tra trùng lặp



Có **5%** nội dung trùng lặp



Có **95%** nội dung không trùng lặp



Có **0%** nội dung người dùng loại trừ



Có **0%** nội dung hệ thống bỏ qua

Học Viên

Phạm Thanh Long

Người hướng dẫn khoa học

Lê Nhật Thăng

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

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

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

Họ và tên học viên: Phạm Thanh Long

Chuyên ngành: Hệ thống thông tin

Khóa: 2023 đợt 1

Tên đề tài: Nghiên cứu, thu thập phân tích Syslog trong hệ thống mạng băng rộng cố định để phục vụ các công tác quản trị của VNPT Hà Nội.

Người hướng dẫn khoa học: PGS.TS Lê Nhật Thăng.

Ngày bảo vệ: 08/07/2026

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

TT	Ý kiến hội đồng	Sửa chữa của học viên
1	Chỉnh sửa format	Học viên đã rà soát, chỉnh sửa các lỗi soạn thảo, các lỗi ngữ pháp

Hà Nội, ngày 10 tháng 07 năm 2026

Ký xác nhận của

CHỦ TỊCH HỘI ĐỒNG CHẤM ĐỀ ÁN	THƯ KÝ HỘI ĐỒNG	NGƯỜI HƯỚNG DẪN KHOA HỌC	HỌC VIÊN
PGS. TSKH Hoàng Đăng Hải	TS. Đỗ Tiến Dũng	PGS.TS Lê Nhật Thăng	Phạm Thanh Long

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

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

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

Căn cứ quyết định số Quyết định số 1895/QĐ-HV ngày 29 tháng 06 năm 2026 của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ. Hội đồng đã họp vào hồi...10....giờ..40..phút, ngày 08 tháng 07 năm 2026 tại Học viện Công nghệ Bưu chính Viễn thông để chấm đề án tốt nghiệp thạc sĩ cho:

Học viên: **Phạm Thanh Long**

Tên đề án tốt nghiệp: Nghiên cứu, thu thập phân tích SYSLOG trong hệ thống mạng băng rộng cố định để phục vụ các công tác quản trị của VNPT Hà Nội

Chuyên ngành: Hệ thống thông tin Mã số: **8.48.01.04**

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

TT	HỌ VÀ TÊN	TRÁCH NHIỆM TRONG HĐ	GHI CHÚ
1	**PGS. TSKH. Hoàng Đăng Hải**	Chủ tịch	
2	**TS. Đỗ Tiến Dũng**	Thư ký	
3	**TS. Nguyễn Xuân Thắng**	Phản biện 1	
4	**TS. Bùi Hải Phong**	Phản biện 2	
5	**TS. Đặng Hoàng Long**	Uỷ viên	

Các nội dung thực hiện:

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

Đặng Hoàng tìm hiểu thêm về syslog
Kafka và Mongo DB, mô tả??
Sử dụng loại để phân tích log có ưu nhược gì?
Cần lưu trữ phân tán không?

7. Trả lời của học viên:

Đề chính xác cao và Redis, TS Trần Thiệp
So sánh với mã lỗi để xây dựng sản

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

9. Hội đồng họp riêng:

- Bầu Ban kiểm phiếu:
  1. Trưởng Ban kiểm phiếu: PGS.TSKH Hoàng Đăng Hải
  2. Uỷ viên Ban kiểm phiếu: TS Đặng Thanh Long
  3. Uỷ viên Ban kiểm phiếu: TS Đỗ Tiến Dũng
- Hội đồng chấm đề án tốt nghiệp bằng bỏ phiếu kín.
- Ban kiểm phiếu làm việc:
- Trưởng Ban kiểm phiếu báo cáo kết quả kiểm phiếu *(có Biên bản họp Ban kiểm phiếu kèm theo)*
- **Điểm trung bình của đề án tốt nghiệp:** 7,8

Kết luận:

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

Chỉnh sửa format

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

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

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

Chủ tịch	**Thư ký**
PGS.TSKH. Hoàng Đăng Hải	Đỗ Tiến Dũng

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

Tên đề tài đề án tốt nghiệp: Nghiên cứu, thu thập phân tích Syslog trong hệ thống mạng băng rộng cố định để phục vụ các công tác quản trị của VNPT Hà Nội

Chuyên ngành: Hệ thống thông tin

Mã chuyên ngành: 8.48.01.04

Họ và tên học viên: Phạm Thanh Long

Họ và tên người nhận xét: Nguyễn Xuân Thắng

Học hàm, học vị: Tiến sĩ

Chuyên ngành: Công nghệ thông tin

Cơ quan công tác: Trường Đại học Hà Nội

Số điện thoại: 0912543560 E-mail: nxthang@hanu.edu.vn

NỘI DUNG NHẬN XÉT

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

Đề tài có cơ sở thực tiễn rõ ràng trong bối cảnh hệ thống mạng băng rộng cố định của VNPT Hà Nội phục vụ nhiều dịch vụ quan trọng như Internet, MyTV, thoại, truyền tải di động và kênh thuê riêng. Hạ tầng mạng MAN-E có quy mô lớn, nhiều tầng thiết bị và nhiều chủng loại thiết bị khác nhau, dẫn tới công tác giám sát, cảnh báo, thống kê và báo cáo hiện còn phân tán ở nhiều hệ thống.

Về mặt khoa học, đề án tập trung vào bài toán thu thập, chuẩn hóa và phân tích dữ liệu Syslog. Đây là nguồn dữ liệu dạng chuỗi thời gian, phát sinh liên tục, có cấu trúc bán định dạng hoặc phi cấu trúc, chịu ảnh hưởng của sự dị thể cú pháp giữa Cisco, Juniper, Huawei và các dòng thiết bị khác. Việc biến đổi các bản tin log thô thành dữ liệu có cấu trúc để phục vụ cảnh báo, truy vấn và báo cáo là bài toán phù hợp với chuyên ngành Hệ thống thông tin.

Tính cấp thiết của đề tài thể hiện ở nhu cầu phát hiện sớm sự cố truyền dẫn, nghẽn lưu lượng, bất thường phần cứng và rủi ro an ninh mạng. Khi xảy ra sự cố diện rộng, lưu lượng bản tin Syslog có thể bùng nổ, đòi hỏi kiến trúc thu thập và xử lý có khả năng đệm, mở rộng và hiển thị gần thời gian thực. Vì vậy, hướng nghiên cứu và triển khai hệ thống phân tích Syslog phục vụ quản trị mạng của VNPT Hà Nội có ý nghĩa ứng dụng cao.

II/ Nội dung của đề án tốt nghiệp, các kết quả đã đạt được:

Nội dung đề án được tổ chức tương đối logic, gồm khảo sát công tác quản trị mạng băng rộng cố định và dữ liệu Syslog tại VNPT Hà Nội, phân tích các công nghệ thu thập - lưu trữ - xử lý log, thiết kế kiến trúc hệ thống và triển khai các kịch bản khai thác dữ liệu Syslog cho quản trị vận hành. Học viên đã nêu được các hạn chế của phương pháp giám sát truyền thống, đặc biệt là sự phân tán công cụ, tính định kỳ của báo cáo và khó khăn khi xử lý log đa chủng loại thiết bị.

Về giải pháp kỹ thuật, đề án đề xuất kiến trúc microservices cho hệ thống quản lý Syslog tập trung. Hệ thống sử dụng lớp tiếp nhận Syslog, lớp đệm phân tán Apache Kafka để chống bùng nổ dữ liệu, lớp bóc tách cú pháp và chuẩn hóa dữ liệu bằng các mẫu Regex, lớp cơ sở dữ liệu MongoDB để lưu trữ linh hoạt dạng tài liệu JSON, và giao diện Angular/WebSockets để hiển thị, giám sát và lập báo cáo. Cách lựa chọn Kafka kết hợp MongoDB là hợp lý đối với dữ liệu log có khối lượng lớn, cấu trúc biến đổi và yêu cầu truy vấn theo thời gian.

Về kết quả thực nghiệm, đề án đã kiểm chứng hệ thống trên một số nhóm bài toán quản trị cụ thể. Hệ thống bóc tách được lỗi truyền dẫn vật lý như CRC và suy hao công suất quang qua đó hỗ trợ cảnh báo suy giảm chất lượng trước khi xảy ra đứt kết nối hoàn toàn. Đối với lưu lượng truyền tải, hệ thống định vị được nút thắt tại tuyến kết nối giữa thiết bị thu gom và thiết bị biên, ghi nhận 185 lần vượt ngưỡng và 340 phút nghẽn mạch nghiêm trọng. Đề án cũng đã khai thác dữ liệu Syslog để thống kê bất thường phần cứng và an ninh, trong đó ghi nhận 314 bản tin về vượt ngưỡng CPU/RAM, 12 bản tin vượt ngưỡng nhiệt độ và 1.425 bản tin dấu hiệu quét cổng trong chu kỳ thực nghiệm. Các kết quả này cho thấy hệ thống có thể đóng vai trò hỗ trợ giám sát hiệu năng thiết bị và phát hiện sớm rủi ro an toàn thông tin.

Nhìn chung, đề án đã hoàn thành được mục tiêu xây dựng mô hình thu thập, lưu trữ, phân tích và trực quan hóa dữ liệu Syslog để phục vụ công tác quản trị của VNPT Hà Nội. Các hướng phát triển như tích hợp học máy để dự báo sự cố, mở rộng thu thập telemetry, xây dựng cơ chế tự động khắc phục lỗi và nâng cấp kiến trúc cloud-native là phù hợp, tuy nhiên cần được kiểm chứng thêm bằng các tiêu chí định lượng và dữ liệu vận hành dài hạn.

III/ Những vấn đề cần giải thích thêm:

1. Syslog thường chứa những thông tin cơ bản nào trong một bản tin cảnh báo, và học viên đã sử dụng các thông tin đó như thế nào để phục vụ công tác giám sát mạng của VNPT Hà Nội?

2. Trong kiến trúc đề xuất, Apache Kafka và MongoDB lần lượt giải quyết những vấn đề nào của bài toán thu thập, phân tích Syslog đa chủng loại thiết bị? Nếu lưu lượng log tăng đột biến vượt 30.000 EPS trong thời gian dài thì học viên sẽ điều chỉnh kiến trúc như thế nào để bảo đảm không mất dữ liệu và vẫn đáp ứng yêu cầu gần thời gian thực?

IV/ Kết luận:

Đồng ý cho học viên bảo vệ đề án tốt nghiệp

Ngày 04 tháng 06 năm 2026

NGƯỜI NHẬN XÉT



Nguyễn Xuân Thắng

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

(Dùng cho người phản biện)

Tên đề tài đề án tốt nghiệp: Nghiên cứu, thu thập phân tích syslog trong hệ thống mạng băng thông rộng cố định để phục vụ các công tác quản trị của VNPT Hà Nội

Chuyên ngành: Hệ thống thông tin

Mã chuyên ngành: 8.48.01.04

Họ và tên học viên: Phạm Thanh Long

Họ và tên người nhận xét: Bùi Hải Phong

Học hàm, học vị: Tiến sĩ

Chuyên ngành: Khoa học máy tính

Cơ quan công tác: Trường ĐH Kiến Trúc Hà Nội

Số điện thoại:.0915594033. E-mail:..phongbh@hau.edu.vn

NỘI DUNG NHẬN XÉT

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

Đề tài áp dụng các kỹ thuật phân tích Syslog để tìm lỗi, hỗ trợ cảnh báo đây là công việc có tính thực tiễn cao. Đề tài thu thập nhiều dữ liệu đa dạng trong thực tiễn.

II/ Nội dung của đề án tốt nghiệp, các kết quả đã đạt được:

Đề tài gồm 61 trang, 4 chương. Đề tài thu thập log của hệ thống mạng VNPT Hà Nội. Đề tài đã triển khai thành công các kỹ thuật phân tích log, cảnh báo lỗi, thống kê sự cố. Đề tài còn có

một số lỗi chính tả trong trang 22. Một số
mã log cần trình bày lại cho chính xác
(trang 28, 29).
Làm rõ hơn các thuật sử dụng để phân tích lỗi log.

III/ Những vấn đề cần giải thích thêm:

Đề tài cần làm rõ các mục sau:
Bổ sung các phương pháp phân tích log sử dụng
các kỹ thuật học máy, học sâu hiện đại.
Gom nhóm các mã lỗi cho rõ ràng hơn.
So sánh, đánh giá phương pháp phân tích log sử dụng
luật so với các phương pháp học máy.
Các luật đưa ra có phân tích hết các lỗi trong log?

IV/ Kết luận:

Đồng ý cho phép học viên bảo vệ đề án tốt nghiệp.

Đồng ý (hoặc không đồng ý) cho phép học viên bảo vệ đề án tốt nghiệp.

Ngày.....tháng.....năm 2026

NGƯỜI NHẬN XÉT

Bùi Hải Phong