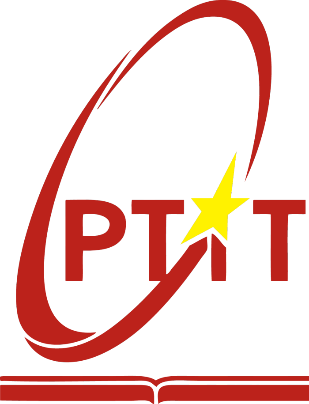
**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**



**KHOA VIỄN THÔNG I**

**----🙣🕮🙡----**

****

**ĐỒ ÁN**

**TỐT NGHIỆP ĐẠI HỌC**

***Đề tài:***

**“****Công nghệ VXLAN và ứng dụng trong mạng N-DCN”**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn :** | **TS. PHẠM ANH THƯ** |
| **Sinh viên thực hiện :** | **ĐẶNG NGUYỄN MINH HƯNG** |
| **Lớp :** | **D16CQVT08-B** |
| **Khóa :** | **2016 - 2021** |
| **Hệ :** | **ĐẠI HỌC CHÍNH QUY** |

**HÀ NỘI – 12/2020**

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

**1. Đánh giá chung**

Đồ án tốt nghiệp đại học của sinh viên Đặng Nguyễn Minh Hưng tìm hiểu các cơ sở lý thuyết về công nghệ VXLAN và mạng N-DCN, đồng thời nêu được ứng dụng của VXLAN trong mạng N-DCN triển khai trạm thu phí không dừng. Qua đó giúp sinh viên làm sâu thêm các kiến thức đã học để có thể áp dụng trong thực tế làm việc.

**2. Đánh giá chi tiết**

Đồ án “Công nghệ VXLAN và ứng dụng trong mạng N-DCN” của sinh viên Đặng Nguyễn Minh Hưng gồm ba chương. Chương 1 của đồ án trình bày về công nghệ VXLAN, trong đó trình bày tổng quan về công nghệ, cấu trúc và cơ chế hoạt động của VXLAN. Chương 2 xem xét đến mạng trung tâm dữ liệu thế hệ mới N-DCN với các nội dung liên quan đến khái niệm, vai trò của mạng N-DCN, những yêu cầu, thách thức và sự phát triển hình thành mạng N-DCN. Chương 3 đi vào ứng dụng của công nghệ VXLAN trong mạng N-DCN được triển khai trong trạm thu phí không dừng. Trong chương này, đồ án đã đưa ra mô hình thiết kế triển khai trạm thu phí dựa trên mạng N-DCN sử dụng công nghệ VXLAN, hoạt động của trạm và các kết quả đạt được.

**3. Nhận xét về tinh thần, thái độ làm việc**

Trong thời gian làm đồ án, sinh viên Đặng Nguyễn Minh Hưng đã thể hiện tinh thần nghiêm túc, có khả năng tìm hiểu và tinh thần học hỏi cao. Từ đó, sinh viên đã nắm vững các kiến thức cơ bản và cả các kiến thức mở rộng trong đồ án.

**4. Kết luận**

Đồ án tốt nghiệp đại học của sinh viên Đặng Nguyễn Minh Hưng đạt yêu cầu của một đồ án tốt nghiệp đại học. Đề nghị hội đồng chấm đồ án thông qua.

**5. Điểm hướng dẫn: 10 (bằng chữ: mười)**

*Hà Nội, ngày 12 tháng 1 năm 2021*

**GIẢNG VIÊN HƯỚNG DẪN**

*(ký, họ tên)*

TS. Phạm Anh Thư

**NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN**

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

…………………………………………………………………………………………..

**Điểm:** …………………….……… (bằng chữ:…..……………….)

*…………, ngày… tháng… năm 20…*

**CÁN BỘ - GIẢNG VIÊN PHẢN BIỆN**

*(ký, họ tên)*

# LỜI CẢM ƠN

Sau một thời gian dài được học tập, nghiên cứu tại trường em cũng đã hoàn thành đồ án tốt nghiệp cuối khóa. Ngoài sự nỗ lực của bản thân, em còn nhận được sự quan tâm giúp đỡ của nhiều tập thể và cá nhân.

Lời đầu tiên, em xin gửi tới toàn thể các thầy, cô giáo trong Khoa Viễn thông I, cùng thầy cô trong Học viện Công nghệ Bưu chính Viễn thông lời cảm ơn chân thành nhất. Đặc biệt, em xin bày tỏ lòng biết ơn sâu sắc tới giảng viên hướng dẫn TS. Phạm Anh Thư, người đã tận tâm hướng dẫn em trong suốt quá trình thực tập và hoàn thiện đề tài.

Em xin gửi lời cảm ơn tới Tổng Công ty Mạng lưới Viettel, nơi đã tạo những điều kiện thuận lợi nhất cho em được tiếp cận môi trường làm việc thực tế, giúp em hình thành khả năng nghiên cứu độc lập và trưởng thành hơn trong suy nghĩ., đặc biệt là người hướng dẫn - anh Hoàng Văn Trung. Qua đây, em xin chân thành cảm ơn ban lãnh đạo và toàn thể cán bộ công nhân viên trong công ty đã tạo mọi điều kiện thuận lợi để em có được những kiến thức thực tế cần thiết.

Em kính chúc quý thầy cô dồi dào sức khỏe và thành công trong sự nghiệp. Đồng kính chúc các cô, chú, anh, chị trong Tổng Công ty Mạng lưới Viettel luôn hoàn thành tốt công việc được giao và đạt được nhiều thành tựu to lớn trong sự nghiệp.

Cuối cùng, em xin được cám ơn Gia đình, bạn bè đã thường xuyên quan tâm, giúp đỡ, chia sẻ kinh nghiệm, trong thời gian học tập, nghiên cứu cũng như trong suốt quá trình thực hiện làm đồ án tốt nghiệp.

Hà Nội, tháng 12 năm 2020

Đặng Nguyễn Minh Hưng

# LỜI MỞ ĐẦU

Tình trạng tắc nghẽn qua các trạm thu phí truyền thống, lực lượng tham gia thu phí rất lớn dẫn đến chi phí bị đội lên rất cao nên việc phải triển khai giải pháp mới là hết sức cấp thiết, vì vậy cần ứng dụng thu phí không dừng. Nhờ ưu điểm thu phí không dừng nên hệ thống ETC góp phần giảm ùn tắc giao thông, giảm thiểu ô nhiễm môi trường và hạn chế tiêu cực trong thu phí. Để triển khai mô hình hệ thống này, mạng trung tâm dữ liệu mới (N-DCN) là một giải pháp phù hợp.

Kiến trúc N-DCN là xu hướng mới cho các tiêu chuẩn của mạng - tạo ra một kiến ​​trúc truyền thông nhanh, xác định được độ trễ, khả năng mở rộng và hiệu quả trong môi trường trung tâm dữ liệu. Ngoài ra, mô hình này kết hợp với VXLAN là một công nghệ mạng ảo được thiết kế để cung cấp các dịch vụ kết nối lớp 2 và lớp 3 thông qua mạng IP. Mạng IP cung cấp khả năng mở rộng, hiệu suất cân bằng và khả năng phục hồi khi có lỗi xảy ra. VXLAN đạt được điều này bằng cách tạo các khung lớp 2 bên trong gói IP. VXLAN chỉ yêu cầu khả năng giao tiếp IP giữa các thiết bị biên trong mô hình VXLAN, được cung cấp bởi các giao thức định tuyến. Trên cơ sở đó, em xin phép nghiên cứu và trình bày về đồ án *“Công nghệ VXLAN và ứng dụng trong mạng N-DCN”.* Bố cục đồ án này bao gồm 3 chương chính:

* Chương I: Công nghệ VXLAN.
* Chương II: Mạng trung tâm dữ liệu thế hệ mới N-DCN.
* Chương III: Ứng dụng công nghệ VXLAN trong mạng N-DCN cho trạm thu phí không dừng ETC.

Một lần nữa, em xin chân thành cảm ơn cô Phạm Anh Thư đã trực tiếp  
hướng dẫn, chỉ bảo, định hướng và giúp đỡ tận tình trong suốt quá trình thực hiện đồ  
án này. Trong quá trình tìm hiểu và nghiên cứu đồ án mặc dù đã có nhiều cố gắng, nhưng do khả năng còn nhiều hạn chế nên đồ án không tránh hỏi những thiếu sót, các sơ suất. Em rất mong nhận được sự phê bình và góp ý của thầy cô trong hội đồng và các bạn để em có thể hoàn thiện, trau dồi thêm kiến thức của mình.

Em xin chân thành cảm ơn!

Sinh viên thực hiện

Đặng Nguyễn Minh Hưng

**MỤC LỤC**

[**LỜI CẢM ƠN** i](#_Toc61368229)

[**LỜI MỞ ĐẦU** ii](#_Toc61368230)

[**DANH MỤC HÌNH VẼ** vi](#_Toc61368231)

[**DANH MỤC BẢNG BIỂU** viii](#_Toc61368232)

[**CÁC THUẬT NGỮ VIẾT TẮT** ix](#_Toc61368233)

[**CHƯƠNG I: CÔNG NGHỆ VXLAN** 1](#_Toc61368234)

[**1.1. Tổng quan VXLAN** 1](#_Toc61368235)

[**1.2. Cấu trúc VXLAN** 7](#_Toc61368236)

[***1.2.1. Điểm đầu cuối đường hầm VXLAN - VTEP*** 8](#_Toc61368237)

[***1.2.2. Định dạng gói tin VXLAN*** 9](#_Toc61368238)

[***1.2.3. Định danh mạng VXLAN - VNI*** 10](#_Toc61368239)

[***1.2.4. Cổng VXLAN*** 11](#_Toc61368240)

[**1.3. Cơ chế hoạt động trong VXLAN** 11](#_Toc61368241)

[***1.3.1. Cơ chế F&L*** 11](#_Toc61368242)

[***1.3.2. Cơ chế BGP EVPN*** 15](#_Toc61368243)

[**1.4. Kết luận chương** 21](#_Toc61368244)

[**CHƯƠNG II: MẠNG TRUNG TÂM DỮ LIỆU THẾ HỆ MỚI N-DCN** 23](#_Toc61368245)

[**2.1. Mạng trung tâm dữ liệu** 23](#_Toc61368246)

[***2.1.1. Khái niệm*** 23](#_Toc61368247)

[***2.1.2. Vai trò của trung tâm dữ liệu*** 23](#_Toc61368248)

[***2.1.3. Thành phần trong trung tâm dữ liệu*** 24](#_Toc61368249)

[***2.1.3. Quá trình phát triển giao thức trong mạng trung tâm dữ liệu*** 25](#_Toc61368250)

[**2.2. Những yêu cầu phát triển mạng trung tâm dữ liệu thế hệ mới** 29](#_Toc61368251)

[**2.3. Mạng trung tâm dữ liệu thế hệ mới** 33](#_Toc61368252)

[***2.3.1. Những đặc tính trong mô hình*** 34](#_Toc61368253)

[***2.3.2. Ưu điểm*** 35](#_Toc61368254)

[***2.3.3. Nhược điểm*** 37](#_Toc61368255)

[**2.4. Tính năng VXLAN BGP EVPN trong mạng N-DCN** 37](#_Toc61368256)

[***2.4.1. Cổng IP anycast phân phối*** 37](#_Toc61368257)

[***2.4.2. Tính di động thiết bị*** 39](#_Toc61368258)

[***2.4.3. Kênh cổng ảo*** 40](#_Toc61368259)

[**2.5. Kết luận chương** 42](#_Toc61368260)

[**CHƯƠNG III: ỨNG DỤNG CÔNG NGHỆ VXLAN TRONG MẠNG N-DCN CHO TRẠM THU PHÍ KHÔNG DỪNG ETC** 43](#_Toc61368261)

[**3.1. Kiến trúc tổng thể hệ thống** 43](#_Toc61368262)

[***3.1.1. Mô hình tổng thể hệ thống*** 43](#_Toc61368263)

[***3.1.2. Hạ tầng mạng vật lý*** 44](#_Toc61368264)

[**3.2. Thiết kế chi tiết hệ thống** 46](#_Toc61368265)

[***3.2.1. Thiết kế chi tiết phân hệ backend*** 46](#_Toc61368266)

[***3.2.2. Thiết kế chi tiết phân hệ Frontend*** 47](#_Toc61368267)

[***3.2.3. Thiết kế định tuyến*** 48](#_Toc61368268)

[**3.3. Mô hình luồng dữ liệu kết nối trong hệ thống** 50](#_Toc61368269)

[***3.3.1. Lưu lượng truy cập xem thông tin tính cước từ internet*** 51](#_Toc61368270)

[***3.3.2. Lưu lượng truy cập giữa Backend và đối tác*** 54](#_Toc61368271)

[***3.3.3. Lưu lượng truy cập Backend từ các điểm bán, cửa hàng đại lý*** 57](#_Toc61368272)

[***3.3.4. Lưu lượng truy cập xem video từ trung tâm giám sát*** 61](#_Toc61368273)

[***3.3.5. Lưu lượng đồng bộ video từ Frontend về Backend*** 65](#_Toc61368274)

[***3.3.6. Lưu lượng tính cước từ Frontend về Backend*** 66](#_Toc61368275)

[***3.3.7. Lưu lượng truy cập từ phân vùng ATTT*** 67](#_Toc61368276)

[**3.4. Kết luận chương** 71](#_Toc61368277)

[**KẾT LUẬN** 72](#_Toc61368278)

[**DANH MỤC TÀI LIỆU THAM KHẢO** 73](#_Toc61368279)

[**PHỤ LỤC: bài báo: “Công nghệ VXLAN và ứng dụng trong mạng DCN”** 74](#_Toc61368280)

**DANH MỤC HÌNH VẼ**

[Hình 1.1: Mô hình mạng Overlay 4](#_Toc61382010)

[Hình 1.2: Mô hình mạng Underlay 4](#_Toc61382011)

[Hình 1.3: Mô hình mạng VXLAN 7](#_Toc61382012)

[Hình 1.4: Thiết bị đầu cuối đường hầm VXLAN - VTEP 8](#_Toc61382013)

[Hình 1.5: Định dạng gói tin VXLAN 9](#_Toc61382014)

[Hình 1.6: Chi tiết định dạng VXLAN 10](#_Toc61382015)

[Hình 1.7: Định danh mạng VXLAN – VNI 11](#_Toc61382016)

[Hình 1.8: Cơ chế VXLAN F&L 12](#_Toc61382017)

[Hình 1.9: Ánh xạ VLAN đến VNI 15](#_Toc61382018)

[Hình 1.10: BGP EVPN tuyến loại 2 18](#_Toc61382019)

[Hình 1.11: BGP EVPN tuyến loại 3 18](#_Toc61382020)

[Hình 1.12: BGP EVPN tuyến loại 5 19](#_Toc61382021)

[Hình 1.13: Quá trình truyền ARP 21](#_Toc61382022)

[Hình 2.1: Giao thức Spanning Tree (STP) 26](#_Toc61382023)

[Hình 2.2: Sự phát triển từ STP sang vPC 28](#_Toc61382024)

[Hình 2.3: Mô hình 3 lớp 30](#_Toc61382025)

[Hình 2.4: Cấu trúc spine-leaf 33](#_Toc61382026)

[Hình 2.5: Vị trí cổng mặc định 38](#_Toc61382027)

[Hình 2.6: VPC với VXLAN BGP EVPN 42](#_Toc61382028)

[Hình 3.1: Mô hình tổng quan hệ thống 43](#_Toc61382029)

[Hình 3.2: Mô hình tổng quan trạm thu phí 45](#_Toc61382030)

[Hình 3.3: Mô hình Backend 46](#_Toc61382031)

[Hình 3.4: Mô hình chạy HA 47](#_Toc61382032)

[Hình 3.5: Kết nối OSPF với GRE phục vụ lưu lượng tính cước 48](#_Toc61382033)

[Hình 3.6: Kết nối OSPF với GRE qua IPsec phục vụ lưu lượng đồng bộ video 49](#_Toc61382034)

[Hình 3.7: Lưu lượng từ Internet đến phân hệ BCCS 51](#_Toc61382035)

[Hình 3.8: Lưu lượng từ phân hệ BCCS đến phân hệ OCS 53](#_Toc61382036)

[Hình 3.9: Lưu lượng từ BCCS Server đến BCCS DB 54](#_Toc61382037)

[Hình 3.10: Lưu lượng từ đối tác đến phân hệ BCCS 55](#_Toc61382038)

[Hình 3.11: Lưu lượng từ đối tác đến phân hệ VMS 56](#_Toc61382039)

[Hình 3.11: Lưu lượng từ đối tác đến phân vùng VMS 57](#_Toc61382040)

[Hình 3.12:Lưu lượng từ Intranet vào phân hệ BCCS 58](#_Toc61382041)

[Hình 3.13: Lưu lượng giữa phân hệ BCCS và OCS 59](#_Toc61382042)

[Hình 3.14: Lưu lượng giữa BCCS Server và BCCS DB 60](#_Toc61382043)

[Hình 3.15: Lưu lượng từ Intranet vào phân hệ BCCS 62](#_Toc61382044)

[Hình 3.16: Flow traffic giữa phân hệ BCCS và VMS 63](#_Toc61382045)

[Hình 3.17: Flow traffic giữa BCCS Server và BCCS DB 64](#_Toc61382046)

[Hình 3.18: Lưu lượng đồng bộ video từ frontend về backend 65](#_Toc61382047)

[Hình 3.19: Lưu lượng tính cước từ frontend về backend 66](#_Toc61382048)

[Hình 3.20: Kết nối giữa ATTT và OCS 67](#_Toc61382049)

[Hình 3.21: Kết nối giữa ATTT và BCCS 69](#_Toc61382050)

[Hình 3.22: Kết nối giữa ATTT và VMS 70](#_Toc61382051)

# DANH MỤC BẢNG BIỂU

[Bảng 1.1: So sánh mạng Overlay và Underlay 6](#_Toc61324860)

**CÁC THUẬT NGỮ VIẾT TẮT**

|  |  |  |
| --- | --- | --- |
| **Từ viết tắt** | **Thuật ngữ tiếng Anh** | **Thuật ngữ tiếng Việt** |
| ARP | Address Resolution Protocol | Giao thức phân giải địa chỉ |
| BD | Bridge domain | Cầu nối |
| BGP | Border Gateway Protocol | Giao thức định tuyến liên miền |
| BUM | Broadcast, unknown unicast, multicast | Quảng bá, đơn hướng không xác định, đa hướng |
| DC | Data Center | Trung tâm dữ liệu |
| ECMP | Equal Cost MultiPath Routing | Định tuyến đa đường giá trị ngang bằng |
| EVPN | Ethernet Virtual Private Network | Mạng riêng ảo Ethernet |
| FHRP | First Hop Redundancy Protocol | Giao thức dự phòng bước đầu |
| FIB | Forwarding information base | Cơ sở thông tin chuyển tiếp |
| FW | Firewall | Tường lửa |
| FCS | Frame check sequence | Chuỗi kiểm tra khung |
| GRE | Generic routing encapsulation | Đóng gói định tuyến chung |
| HA | High Availability | Sẵn sàng cao |
| HSRP | Hot Standby Router Protocol | Giao thức dự phòng nóng |
| IR | Ingress Replication | Sao chép xâm nhập |
| LACP | Link Aggregation Control Protocol | Giao thức nhóm các kết nối vật lý thành kết nối logic |
| LISP | Locator/ID Separation Protocol | Giao thức phân tách vị trí, định danh |
| M-BGP | Multiprotocol BGP | Đa giao thức BGP |
| MTU | Maximum transmission unit | Đơn vị truyền tải tối đa |
| MAC | Media Access Control | Kiểm soát truy cập phương tiện truyền thông |
| MC-LAG | MultiChassis Link Aggregation | Tổng hợp liên kết nhiều thiết bị |
| ND | Neighbor discovery | Khám phá hàng xóm |
| N-DCN | Nextgen – Data center network | Mạng trung tâm dữ liệu thế hệ mới |
| NLRI | Network Layer Reachability Information | Thông tin truy cập lớp mạng |
| NVE | Network Virtualization Edge | Thiết bị mạng biên ảo |
| NVGRE | Network Virtualization using Generic Encapsulation | Ảo hóa mạng sử dụng tính năng đóng gói chung |
| OSI | Open Systems Interconnection | Kết nối hệ thống mở |
| PIM | Protocol Independent Multicast | Giao thức đa hướng độc lập |
| SDN | Software Defined Networking | Mạng định nghĩa phần mềm |
| RFID | Radio Frequency Identification | Nhận dạng tần số vô tuyến |
| STP | Spanning Tree Protocol | Giao thức Spanning Tree |
| SVI | Switch Virtual Interface | Giao diện chuyển mạch ảo |
| TTL | Time to live | Thời gian tồn tại |
| UDP | User Datagram Protocol | Giao thức dữ liệu người dùng |
| VLAN | Virtual Local Area Network | Mạng LAN ảo |
| VM | Virtual Machine | Máy ảo |
| VNI | VXLAN Network Identification | Định danh mạng VXLAN |
| VPC | Virtual Port Channels | Kênh cổng ảo |
| VRF | Virtual Routing and Forwarding | Chuyển tiếp định tuyến ảo |
| VRRP | Virtual Route Rendundancy Protocol | Giao thức dự phòng tuyến đường ảo |
| VSS | Virtual Switching System | Hệ thống chuyển mạch ảo |
| VTEP | VXLAN Tunnel End Point | Thiết bị đầu cuối đường hầm VXLAN |
| VXLAN | Virtual Extensible LAN | Mạng LAN ảo mở rộng |

# CHƯƠNG I: CÔNG NGHỆ VXLAN

## 1.1. Tổng quan VXLAN

Sự ảo hóa máy chủ đã đặt ra nhu cầu ngày càng cao đối với cơ sở hạ tầng mạng vật lý. Một máy chủ vật lý có nhiều máy ảo (VMs), mỗi máy ảo có một địa chỉ MAC riêng. Điều này đòi hỏi các bảng địa chỉ MAC lớn trong mạng Ethernet do có sự trao đổi thông tin giữa hàng trăm nghìn máy ảo.

Những hạn chế về khả năng mở rộng của mạng LAN ảo (VLAN) đã tạo ra nhu cầu về các giải pháp thay thế khi ảo hóa ngày càng mở rộng. Tuy nhiên, việc sử dụng khái niệm cầu nối (BD) của lớp 2 mô hình OSI cần được duy trì để liên kết các máy ảo và các thiết bị trong mạng. Để đạt được điều này, mạng LAN ảo mở rộng (VXLAN) đã được phát triển. VXLAN mở rộng số lượng dịnh danh (ID) mạng từ 4094 lên 16 triệu và cung cấp cơ chế lớp phủ MAC dựa trên IP đóng gói UDP giúp mở rộng không gian VLAN hiện tại, do đó số lượng miền quảng bá được mở rộng. Vì vậy, VXLAN đại diện cho mạng Overlay lớp 2 trong lớp 3. Các thiết bị biên chịu trách nhiệm đóng và mở tiêu đề VXLAN được gọi là thiết bị đầu cuối đường hầm VXLAN (VTEP).

Về bản chất, VXLAN kết nối hai hoặc nhiều phân đoạn mạng qua lớp 3 trong khi vẫn cho phép các thiết bị đầu cuối trên các phân mạng khác nhau này chia sẻ cùng một miền quảng bá lớp 2. Do đó, trong khi VLAN luôn chỉ hoạt động trong lớp liên kết dữ liệu Ethernet (lớp 2), VXLAN hoạt động trên lớp 3. Ngoài ra, mạng Ethernet lớp 2 sử dụng giao thức STP (Spanning-tree) để ngăn chặn vòng lặp, do đó có một tuyến đường duy nhất hoạt động để chuyển tiếp lưu lượng VLAN. Tuy nhiên, bằng cách sử dụng lớp 3 với VXLAN, tất cả đường dẫn giữa các VTEP được sử dụng thông qua định tuyến đa đường giá trị ngang bằng (ECMP), điều này cũng làm tăng đáng kể hiệu suất hoạt động của mạng. Ngay cả khi sử dụng ECMP trong mạng Underlay, một đường dẫn duy nhất giữa các VTEP trong mạng Overlay được sử dụng trong khi vẫn tận dụng tất cả các đường ECMP sẵn có. Và bằng cách sử dụng lõi IP như một phần của lớp mạng Underlay, VXLAN đã giúp việc theo dõi giám sát trung tâm dữ liệu trở nên dễ dàng hơn bằng cách tận dụng kiến thức chuyên môn về mạng hiện có.

VXLAN là một miền quảng bá lớp 2 được xác định bởi định danh mạng VXLAN (VNI), đây là một giá trị mở rộng so với VLAN truyền thống dùng để định danh cho đường hầm lưu lượng từ một VTEP này sang một VTEP khác.

Như tên gọi của nó, VXLAN được thiết kế để cung cấp các dịch vụ mạng Ethernet lớp 2 giống như VLAN hiện nay, nhưng với khả năng mở rộng và tính linh hoạt cao hơn. So với VLAN, VXLAN mang lại những lợi ích sau:

* Tính linh hoạt của VLAN trong các phân đoạn cho nhiều đối tượng: Nó cung cấp giải pháp để mở rộng các phân đoạn lớp 2 trên cơ sở hạ tầng mạng bên dưới để khối lượng công việc của đối tượng thuê có thể được đặt trên các nhóm vật lý trong trung tâm dữ liệu.
* Khả năng mở rộng cao hơn để có thể xử lý nhiều phân đoạn lớp 2: Các VLAN sử dụng ID VLAN 12 bit để định danh cho các phân đoạn lớp 2, điều này dẫn đến việc hạn chế khả năng mở rộng của chỉ 4094 VLAN như đã nói ở trên. VXLAN sử dụng 24 bit để định danh mạng VXLAN, cho phép tối đa 16 triệu phân đoạn VXLAN cùng tồn tại trong cùng một miền quản trị.
* Tận dụng tốt hơn các tuyến đường có sẵn trong cơ sở hạ tầng mạng vật lý. VLAN sử dụng giao thức STP (Spanning-tree) để ngăn chặn vòng lặp, dẫn đến việc không sử dụng một nửa số liên kết mạng trong mạng bằng cách chặn các tuyến đường còn lại. Ngược lại, các gói VXLAN được chuyển qua mạng vật lý dựa trên tiêu đề lớp 3 của nó và có thể tận dụng hoàn toàn lợi ích của định tuyến lớp 3, định tuyến đa đường chi phí ngang bằng (ECMP) và các giao thức liên kết tổng hợp để sử dụng tất cả các đường dẫn có sẵn.

Như đã nêu ở trên, VXLAN là một trong những công nghệ đại diện cho mạng Overlay. Khái niệm về Overlay đã xuất hiện trong thế giới mạng khá lâu. Mạng Overlay có thể được định nghĩa như là một mạng logic ảo, là một đường hầm tĩnh hoặc động chạy trên cơ sở hạ tầng mạng vật lý có sẵn. Các đường hầm này được sử dụng cho mục đích bảo mật, để việc tra cứu định tuyến dễ dàng. Tất cả các nút trong mạng Overlay được kết nối với nhau bằng các liên kết logic và mỗi liên kết này tương ứng với một đường dẫn trong mạng Underlay.

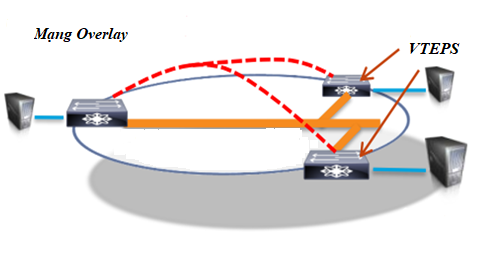
Với mạng Overlay, khung hoặc gói tin ban đầu được đóng gói tại thiết bị biên nguồn có tiêu đề bên ngoài để chuyển đến một thiết bị biên đích thích hợp. Các thiết bị mạng trung gian đều chuyển tiếp gói tin dựa trên tiêu đề bên ngoài và không biết về gói tin hay khung ban đầu. Tại thiết bị biên đích, tiêu đề mạng overlay bị loại bỏ và gói tin được chuyển tiếp dựa trên dữ liệu bên trong.

Một tính năng chính mà tất cả các mạng Overlay này đưa có liên quan đến vị trí và định danh. Định danh xác định một thiết bị đầu cuối có thể là địa chỉ IP, địa chỉ MAC của nó,... Vị trí xác định thiết bị biên của đường hầm chịu trách nhiệm đóng gói và phân tách lưu lượng đường hầm cho thiết bị đầu cuối đó. Tiêu đề bên ngoài của lớp mạng Overlay tham chiếu đến vị trí nguồn và đích còn tiêu đề bên trong tham chiếu đến định danh thiết bị đầu cuối nguồn và đích.

Tính năng đáng chú ý khác liên quan đến dịch vụ lớp mạng Overlay cung cấp. Tính năng này cho biết loại lớp mạng Overlay cũng như định dạng tiêu đề liên quan. Các dịch vụ lớp mạng Overlay thường được cung cấp dưới dạng dịch vụ lớp 2 (bắc cầu) hoặc lớp 3 (định tuyến). Tuy nhiên, nhiều lớp mạng Overlay hiện tại cung cấp cả dịch vụ lớp 2 và lớp 3. Gói ban đầu (lớp 3) hoặc khung (lớp 2) có thể được đóng trong một gói (lớp 3) hoặc một khung khác (lớp 2). Điều này dẫn đến khả năng có bốn sự kết hợp, tùy thuộc vào việc một gói hoặc khung được mang trong một gói hoặc khung khác. Nếu tiêu đề bên ngoài là khung lớp 2, lớp mạng Overlay được gọi là đóng khung. Mặt khác, nếu tiêu đề bên ngoài là thông tin lớp 3, lớp mạng Overlay được gọi là đóng gói . Ví dụ về các lớp mạng Overlay sử dụng đóng gói là LISP, VXLAN.

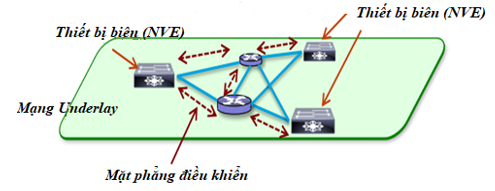
Tiếp theo, lớp mạng Overlay yêu cầu cơ chế để biết thông tin thiết bị đầu cuối đứng sau thiết bị biên. Điều này cho phép các thiết bị biên xây dựng cơ sở dữ liệu ánh xạ đến định danh, vị trí. Thông tin ánh xạ có thể được hỗ trợ thông qua bộ điều khiển mạng định nghĩa phần mềm (SDN) trung tâm (chẳng hạn như OpenDaylight) hoặc thông qua giao thức phân phối thiết bị đầu cuối (BGP EVPN) hoặc cơ chế khám phá dựa trên mặt phẳng dữ liệu (Flood and learn [F&L]).Sử dụng thông tin này để tra cứu thiết bị đầu cuối, thiết bị biên nguồn có thể đẩy một tiêu đề trên lớp mạng Overlay với các trường thích hợp để gói hoặc khung có thể được phân phối đến thiết bị biên đích và sau đó, đến thiết bị đầu cuối đích. Vì các thiết bị đầu cuối thường thuộc về một thực thể nhất định nên tiêu đề lớp Overlay bắt buộc phải có thông tin về thực thể này. Bằng cách truyền thông tin của thực thể thích hợp vào tiêu đề lớp Overlay tại thiết bị nguồn trong quá trình đóng gói, thiết bị biên đích có thể giải đóng gói để tra cứu điểm đến cho mỗi thực thể, do đó đảm bảo tách biệt lưu lượng truy cập của giữa nhiều đối tượng khác nhau. Đây là yêu cầu chính trong môi trường trung tâm dữ liệu đa thực thể.

Mặt phẳng dữ liệu cho lớp Overlay, ngoài việc cung cấp vận chuyển được đóng gói tin đơn hướng của lưu lượng lớp 2 hoặc lớp 3, cũng cần xử lý việc vận chuyển của lưu lượng truy cập đa điểm. Lưu lượng truy cập đa điểm thường được gọi là “BUM”. Việc xử lý các loại lưu lượng này yêu cầu có khả năng đặc biệt để nhận các gói đa điểm được đặt từ lớp Overlay và sau đó nhân bản và vận chuyển các gói tương ứng. Hai phương pháp phổ biến nhất có thể đáp ứng việc này trong lớp Underlay là đa hướng IP và sao chép xâm nhập (IR).



##### Hình 1.1: Mô hình mạng Overlay

Hình 1.1 minh họa mô hình mạng Overlay khi sử dụng VXLAN. Trong các thiết bị biên có chức năng đóng gói và giải mã lưu lượng dữ liệu để xây dựng toàn bộ lớp mạng Overlay. Trong thuật ngữ VXLAN, chức năng này trên thiết bị biên được gọi là VTEP. Với các dịch vụ Overlay được xác định cùng với các gói tin trên mặt phẳng dữ liệu khác nhau, cần phải có một phương thức truyền tải để di chuyển dữ liệu qua mạng vật lý. Phương thức vận chuyển này thường là một mạng truyền tải lớp dưới Underlay. Khi xác định lớp mạng Underlay, chúng ta cần biết về mô hình OSI nơi xảy ra quá trình đóng gói đường hầm. Theo một nghĩa nào đó, loại tiêu đề lớp Overlay quy định loại mạng truyền tải. Ví dụ, với VXLAN, mạng truyền tải bên lớp Underlay sẽ là mạng lớp 3, mạng này sẽ vận chuyển các gói đóng trong VXLAN giữa các thiết bị biên nguồn và đích.



##### Hình 1.2: Mô hình mạng Underlay

Underlay là thuật ngữ chỉ một hạ tầng mạng vật lý cung cấp kết nối trong nền tảng của mạng. Việc định tuyến IP lớp 3 cung cấp kết nối giữa các thiết bị bất kì với nhau. Nhờ đó các thiết bị có thể truyền các giao thức báo hiệu cho mạng Overlay. Các thiết bị biên trong mạng Underlay được gọi là thiết bị mạng biên ảo (NVE), còn tên gọi VTEP được sử dụng trong mạng Overlay. Hình 1.2 trên mô tả mạng Underlay.

Mạng truyền tải, được gọi là lớp Underlay, cần phải tính đến các byte bổ sung do các tiêu đề lớp phủ phát sinh. Thông thường, trong trường hợp VXLAN, mạng truyền tải cần được cung cấp thêm 50 byte trong đơn vị truyền tải tối đa (MTU). Tương tự như vậy, mạng truyền tải cũng phải phù hợp với các yêu cầu về khả năng phục hồi và hội tụ của lớp Overlay. Điều quan trọng là phải xem xét khả năng mở rộng, khả năng phục hồi, hội tụ và dung lượng khi thiết kế mạng trục của trung tâm dữ liệu. Với các loại mạng trục trung tâm dữ liệu dựa trên lớp mạng Overlay, tầm quan trọng của lớp mạng Underlay là không thể phủ nhận. Thông thường, hiệu suất của lớp mạng Overlay chỉ hoạt động tốt khi việc chuyển tại lớp mạng Underlay chạy trơn tru. Ngay cả việc khắc phục sự cố, gỡ lỗi và hội tụ lưu lượng lớp mạng Overlay, Underlay cũng đóng vai trò quan trọng.

Khi thiết kế và xây dựng lớp mạng Underlay cho mạng trục, gán địa chỉ IP là một khía cạnh quan trọng của quá trình này. Việc gán địa chỉ IP là điều kiện tiên quyết cho bất kỳ loại giao thức định tuyến nào cung cấp khả năng truy cập giữa các thiết bị khác nhau trong mạng trục.

Mạng Underlay có các đặc điểm gồm:

* Lớp mạng Underlay là một cơ sở hạ tầng vật lý trên đó mạng Overlay được xây dựng.
* Trong môi trường trung tâm dữ liệu, vai trò của mạng Underlay là cung cấp kết nối IP đơn hướng từ thiết bị vật lý này (máy chủ, thiết bị lưu trữ, bộ định tuyến hoặc bộ chuyển mạch) đến bất kỳ thiết bị vật lý nào khác.
* Mạng Underlay là mạng bên dưới chịu trách nhiệm phân phối các gói qua các mạng.
* Một mạng Underlay lý tưởng cung cấp kết nối băng thông cao có độ trễ thấp, không bị chặn từ một điểm bất kỳ đến bất kỳ điểm nào khác trong mạng.
* Các tính năng của mạng Underlay xác định hiệu quả hoạt động mạng Overlay.
* Các tùy chọn ít khả năng mở rộng hơn của chuyển tiếp đa đường. Trên thực tế, việc sử dụng nhiều đường dẫn có thể có chi phí và độ phức tạp liên quan.

###### Bảng 1.1: So sánh mạng Overlay và Underlay

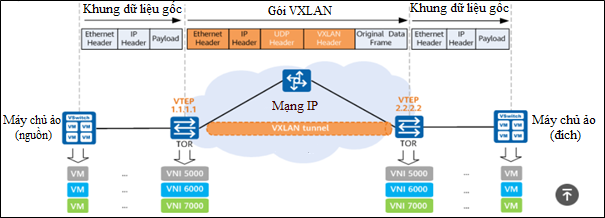
|  |  |  |
| --- | --- | --- |
|  | **Underlay** | **Overlay** |
| **Luồng lưu lượng** | Truyền các gói đi qua các thiết bị mạng như router và switch. | Chỉ truyền các gói dọc theo các liên kết ảo giữa các nút lớp overlay. |
| **Thời gian triển khai** | Ít khả năng mở rộng, tốn thời gian để thiết lập các dịch vụ và chức năng mới. | Thời gian triển khai nhanh. |
| **Điều khiển gói** | Dựa trên phần cứng | Dựa trên phần mềm |
| **Đóng gói và tiêu đề** | Phân phối gói có độ tin cậy ở lớp 3 và lớp 4. | Nhu cầu đóng các gói nguồn và đích, cần bổ sung tiêu đề bên ngoài. |
| **Chuyển tiếp đa đường** | Các tùy chọn của chuyển tiếp đa đường ít khả năng mở rộng. Trên thực tế, việc sử dụng nhiều đường dẫn khá phức tạp. | Hỗ trợ chuyển tiếp đa đường trong mạng ảo. |
| **Khả năng mở rộng** | Do giới hạn về kỹ thuật, ít khả năng mở rộng. | Được thiết kế cung cấp khả năng mở rộng hơn so với mạng Underlay. Ví dụ: VLAN mạng Underlay hỗ trợ 4096 VLAN trong khi VXLAN mạng Overlay cung cấp tới 16 triệu giá trị VNI. |
| **Vận chuyển gói tin** | Chịu trách nhiệm vận chuyển các gói tin. | Giảm tải từ phân phối các gói tin |
| **Giao thức sử dụng** | Các giao thức Underlay bao gồm Chuyển mạch Ethernet, VLAN, Định tuyến, v.v. | Các giao thức mạng Overlay bao gồm mạng LAN ảo mở rộng (VXLAN), ảo hóa mạng sử dụng đóng gói chung (NVGRE), kiểm tra truyền tải không trạng thái (SST), đóng gói định tuyến chung (GRE), đa hướng IP và ảo hóa mạng Overlay 3 (NVO3). |
| **Quản lý đa đối tượng** | Yêu cầu tách biệt dựa trên NAT hoặc VRF có thể gặp thách thức trong môi trường lớn. | Khả năng quản lý các địa chỉ IP chồng chéo giữa nhiều đối tượng. |
|

Mạng Underlay là cơ sở hạ tầng vật lý mà mạng Overlay được xây dựng trên đó. Trong mô hình spine-leaf, thuật ngữ underlay thường được sử dụng để chỉ việc định tuyến IP lớp 3 giữa các thiết bị với nhau.

Mạng Overlay là một mạng ảo, hay một đường hầm tĩnh hoặc động chạy trên cơ sở hạ tầng mạng vật lý (Underlay) có sẵn.

## 1.2. Cấu trúc VXLAN

VXLAN là công nghệ mở rộng VLAN truyền thống, nó chèn thêm IP vào trong gói khung lớp 2, cụ thể là sử dụng cổng UDP 4789 được chỉ định bởi tổ chức cấp phát số hiệu internet. Địa chỉ MAC trong gói tin UDP tạo ra một đường hầm (tunnel) cho phép bạn mở rộng Layer 2 trên bất kỳ mạng Layer 3 nào. Một tiêu đề VXLAN được chèn thêm vào khung lớp 2 sau đó được đóng gói chung với gói tin UDP để chuyến đến bảng định tuyến. VXLAN bao gồm một mã định danh gọi là VNI. Yếu tố này giúp định nghĩa phân đoạn VXLAN. Định danh này được định nghĩa bởi 24 bit nhị phân, lớn hơn so với 12 bit của VLAN truyền thống tương ứng với con số 16 triệu VXLAN so với 4094 VLAN. Các thiết bị thiết lập giữa 2 đầu đường hầm gọi là VTEP. Những thiết bị này có thể là máy vật lý, máy ảo, thiết bị mạng như Router và Switch. Đây là nơi tiến trình đóng gói (encapsulation) và mở gói (de-encapsulation) diễn ra. Một kết nối với mạng IP ở giữa và một kết nối tới phân đoạn mạng nội bộ, bạn có thể chuyển dữ liệu VXLAN sang một mạng khác. Hình 1.3 cho ta thấy mô hình mạng VXLAN được cấu thành từ các thiết bị biên VTEP được cấu hình định danh mạng VXLAN (NVI), kết nối ngay sau là máy chủ ảo hóa gồm nhiều máy ảo (VM) truyền dữ liệu ban đầu thông qua đóng gói VLXAN tại VTEP dựa trên mạng IP lớp 3.



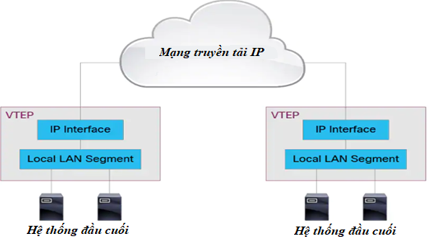
##### Hình 1.3: Mô hình mạng VXLAN

### 1.2.1. Điểm đầu cuối đường hầm VXLAN - VTEP

VXLAN sử dụng VTEP để ánh xạ các thiết bị đầu cuối với các phân đoạn VXLAN và thực hiện đóng gói và giải đóng gói VXLAN. Các phân đoạn VXLAN độc lập với cấu trúc liên kết mạng bên dưới. Ngược lại, mạng IP bên dưới giữa các VTEP độc lập với lớp phủ VXLAN. Nó định tuyến các gói dựa trên tiêu đề địa chỉ IP bên ngoài, tiêu đề này có địa chỉ IP VTEP nguồn và địa chỉ IP VTEP đích. VTEP có hai tính năng tương ứng hai giao diện:

* Giao diện chuyển mạch trên phân đoạn mạng LAN cục bộ để hỗ trợ giao tiếp điểm cuối cục bộ
* Giao diện còn lại là giao diện IP để định tuyến IP mạng truyền tải.

Các thành phần chức năng của VTEP và cấu trúc liên kết logic được tạo ra cho kết nối lớp 2 trên mạng IP truyền tải được thể hiện trong hình 1.5.



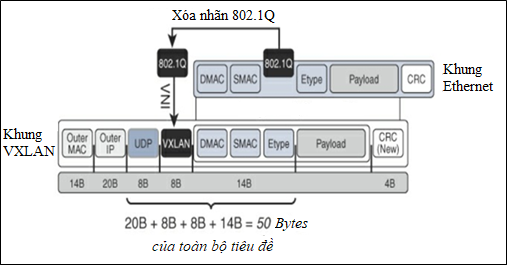
##### Hình 1.4: Thiết bị đầu cuối đường hầm VXLAN - VTEP

Hạ tầng VLAN có một địa chỉ IP duy nhất xác định thiết bị VTEP trên mạng IP truyền tải. Thiết bị VTEP sử dụng địa chỉ IP này để đóng gói các khung Ethernet và truyền các gói này tới mạng truyền tải thông qua giao diện IP.

Thiết bị VTEP cũng phát hiện ra các VTEP đầu xa trong phân đoạn VXLAN của nó và học các địa chỉ MAC của VTEP đầu xa thông qua giao diện IP của nó. Các thành phần chức năng của VTEP và cấu trúc liên kết logic được tạo cho kết nối lớp 2 trên mạng IP truyền tải được chỉ ra trong hình 1.4.

### 1.2.2. Định dạng gói tin VXLAN

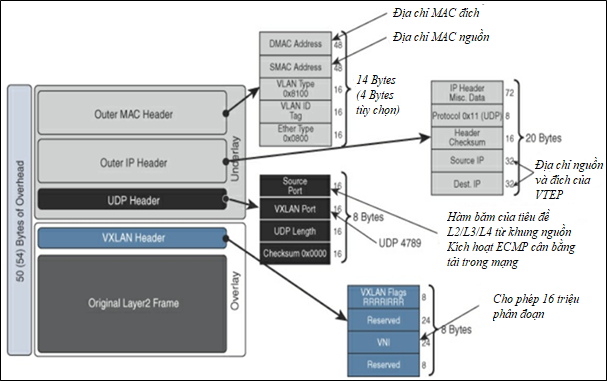
Bên trong tiêu đề VXLAN đề cập đến các thành phần lớp 2, trong khi các tiêu đề bên ngoài là thông tin các thiết bị biên lớp 3. Do đó, tiêu đề VXLAN chứa khung Ethernet lớp 2 ban đầu, bao gồm địa chỉ MAC nguồn cũng như địa chỉ MAC đích bên trong, thông tin dữ liệu Ethernet và chuỗi kiểm tra lỗi (FCS). Ngoài ra, nhãn 802.1Q bên trong khung Ethernet lớp 2 ban đầu được loại bỏ và ánh xạ tới một VNI để hoàn thành tiêu đề VXLAN, như được minh họa ở hình 1.5.



##### Hình 1.5: Định dạng gói tin VXLAN

Tiêu đề VXLAN đứng trước những tiêu đề như UDP, Outer IP và Outer MAC. Cổng đích của tiêu đề UDP bên ngoài được đặt với giá trị 4789. Số cổng nguồn UDP được tạo ra dựa trên các trường của tiêu đề ban đầu. Trong tiêu đề Outer IP thiết lập địa chỉ của VTEP nguồn là địa chỉ IP nguồn, địa chỉ VTEP đích là địa chỉ IP đích. Nguồn gốc của tiêu đề Outer MAC dựa trên tra cứu định tuyến trên tiêu đề Outer IP. Do đó, VXLAN cần thêm 50 byte vào khung Ethernet hiện tại hoặc 54 byte, nếu nhãn IEEE 802.1Q được thêm vào. Hình 1.6 mô tả chi tiết định dạng khung VXLAN

Thông thường, trong mạng lớp 3 với ECMP, cân bằng tải đạt được để chuyển tiếp lưu lượng bằng cách chọn một trong những tuyến đường có giá trị như nhau dựa trên 5 thuộc tính bao gồm địa chỉ IP nguồn, địa chỉ IP đích, giao thức lớp 4 (L4), cổng nguồn L4 và cổng đích L4. Với VXLAN, đối với lưu lượng giữa các thiết bị đầu cuối phía sau cặp VTEP nguồn và đích, tất cả các trường trong tiêu đề bên ngoài, ngoại trừ UDP đều giống hệt nhau. Vì cổng nguồn UDP thay đổi dựa trên nội dung của gói hoặc khung bên trong, nó cung cấp thông tin cần thiết để chọn các tuyến đường khác nhau cho các thiết bị đầu cuối giữa cùng một cặp VTEP nguồn và đích. Do đó, việc tạo giá trị cổng nguồn UDP một cách thích hợp là một tính năng quan trọng mà VTEP cần hỗ trợ, bất kể chúng được triển khai ở phần mềm hay phần cứng. Đây là điểm khác biệt quan trọng của VXLAN so với các lớp mạng overlay khác.

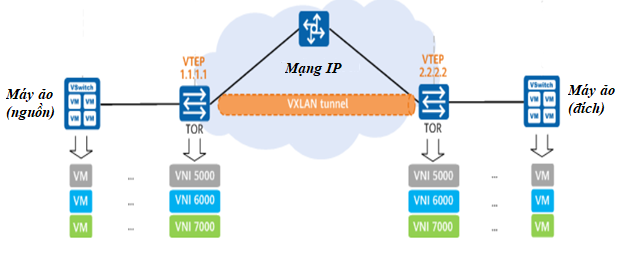


##### Hình 1.6: Chi tiết định dạng VXLAN

### 1.2.3. Định danh mạng VXLAN - VNI

Trong một mạng VLAN truyền thống, theo tiêu chuẩn, tối đa khoảng 4094 VLAN khả dụng. Sau khi ảo hóa máy chủ, một máy chủ vật lý lưu trữ nhiều máy ảo. Mỗi máy ảo có một địa chỉ IP và địa chỉ MAC độc lập. Các đám mây hoặc trung tâm dữ liệu đám mây ảo hóa lớn cần phải chứa hàng chục nghìn thiết bị đầu cuối hoặc thậm chí nhiều hơn. Trong trường hợp này, VLAN không thể đáp ứng.

VXLAN mở rộng không gian địa chỉ VLAN bằng cách thêm ID phân đoạn 24 bit được gọi là định danh mạng VXLAN (VNI) và cho phép 16 triệu phân đoạn mạng VXLAN. VNI trong mỗi khung phân tách các mạng logic riêng lẻ, cho phép hàng triệu phân đoạn VXLAN lớp 2 cùng tồn tại chung trên một mạng lớp 3. Mỗi VLAN được ánh xạ tới một VNI duy nhất để mở rộng phân đoạn VLAN lớp 2 đến một thiết bị đầu xa. Mỗi VNI lớp 2 được ánh xạ tới một miền cầu nối (BD) để chuyển tiếp mạng con nội bộ của các gói VXLAN. Khi các thiết bị trao đổi dữ liệu nằm có địa chỉ thuộc các dải mạng khác nhau thì ngoài giá trị VNI lớp 2, gói tin VXLAN sẽ sử dụng VNI lớp 3. Hình 1.7 chỉ ra giá trị VNI được cấu hình trên VTEP tương ứng với mỗi máy ảo và dải mạng khác nhau.



##### Hình 1.7: Định danh mạng VXLAN – VNI

### 1.2.4. Cổng VXLAN

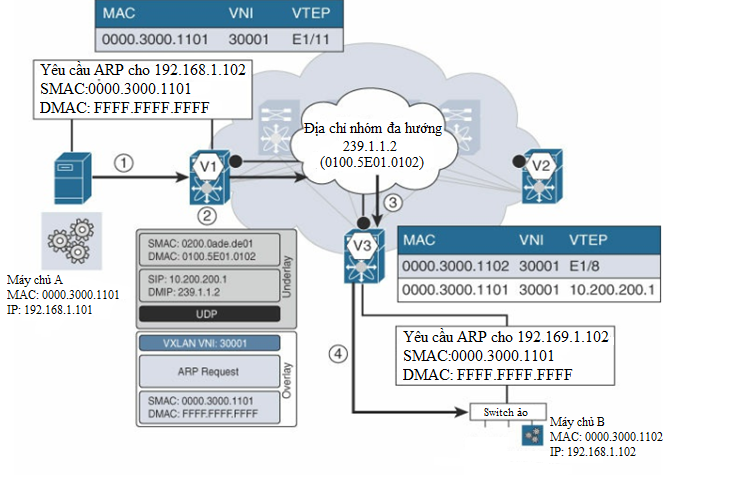
Giống như trong VLAN, các thiết bị đầu cuối trong các VNI khác nhau hay thiết bị trên mạng VXLAN và mạng không phải VXLAN không thể giao tiếp trực tiếp với nhau. Để đáp ứng các yêu cầu giao tiếp này, VXLAN giới thiệu các cổng VXLAN. Có hai loại cổng VXLAN là cổng lớp 2 và cổng lớp 3.

* Cổng VXLAN lớp 2: kết nối các thiết bị đầu cuối với mạng VXLAN và cho phép giao tiếp mạng con nội bộ trên cùng một mạng VXLAN.
* Cổng VXLAN lớp 3: cho phép giao tiếp giữa các mạng con trên mạng VXLAN và truy cập mạng bên ngoài.

## 1.3. Cơ chế hoạt động trong VXLAN

### 1.3.1. Cơ chế F&L

Với lớp phủ IP, bất kỳ giao thức định tuyến thích hợp nào như OSPF, IS-IS, v.v. có thể được sử dụng cho tất cả các VTEP để thiết lập kết nối với nhau. Lớp mạng IP vật lý cần cung cấp một phương tiện để vận chuyển dữ liệu mạng Overlay đến một hoặc nhiều điểm. Lớp mạng Underlay cần cung cấp một phương tiện để vận chuyển lưu lượng truy cập điểm, đa điểm trên lớp mạng Overlay là sử dụng phát đa hướng IP. Trong trường hợp này, lớp mạng Underlay sử dụng giao thức đa hướng độc lập (PIM), là một họ các giao thức định tuyến đa hướng cho mạng IP.



##### Hình 1.8: Cơ chế VXLAN F&L

VXLAN VNI được ánh xạ tới một nhóm IP đa hướng thích hợp, khi VNI được cấu hình tại một VTEP. Nhờ đó, tất cả các VTEP trở thành thành viên của cùng một VNI đều thuộc cùng một nhóm phát đa hướng. Lưu lượng truy cập đa điểm bắt nguồn từ thiết bị đầu cuối là một phần của VNI được chuyển đến tất cả các thiết bị đầu cuối khác trong VNI đó qua VXLAN nơi địa chỉ Outer IP đích được thiết lập là địa chỉ IP đa hướng được liên kết với VNI đó. Hãy xem xét cấu trúc liên kết mạng mẫu được hiển thị trong hình 1.8, minh họa một giao tiếp dựa trên F&L điển hình trong mạng VXLAN.

Máy chủ A (IP 192.168.1.101) đứng sau VTEP V1 (10.200.200.1) và máy chủ B (IP 192.168.1.102) đứng sau VTEP V3 (10.200.200.3) là một phần của cùng một phân đoạn VXLAN với VNI 30001. Mạng con được liên kết với VNI 30001 là 192.168.1.0/24 và địa chỉ IP đa hướng là 239.1.1.2. Khi VNI 30001 được cấu hình ở V1 và V3, cả hai đều tham gia nhóm đa hướng 239.1.1.2. Giả sử máy chủ A muốn giao tiếp với Máy chủ B.

Như thường lệ, máy chủ A trước tiên sẽ truy vấn địa chỉ IP-MAC được liên kết với máy chủ B, sử dụng giao thức phân giải địa chỉ (ARP). Máy chủ A gửi yêu cầu ARP với DMAC = FFFF.FFFF.FFFF và SMAC = 00:00:30:00:11:01. VTEP V1 nhận yêu cầu ARP này và thực hiện tra cứu lớp 2 dựa trên [VNI = 30001, DMAC =FFFF.FFFF.FFFF]. VTEP V1 xác định khung này cần được gửi đến tất cả các thành viên của VNI 30001. Do đó, nó đóng gói tin này bằng tiêu đề VXLAN với VNI = 30001, IP nguồn = 10.200.200.1 và IP đích = 239.1.1.2.

Gói được đóng gói VXLAN được gửi đến lớp IP. Gói tin được chuyển tiếp qua nhóm đa hướng và sau đó đến tất cả các VTEP thành viên (VTEP V3). Gói được đóng tiêu đề VXLAN được giải đóng gói tại VTEP V3 vì nó có [VNI = 30001, Group = 239.1.1.2] được cấu hình. Sau đó V3 học được thông tin [VNI = 30001, MAC = 00: 00: 30: 00: 11: 01-> 10.200.200.1]. Tra cứu Lớp 2 được thực hiện ở V3 dựa trên [VNI = 30001, DMAC = FFFF.FFFF.FFFF]. Bởi vì đây là một gói quảng bá, nó được chuyển tiếp đến tất cả các máy chủ bên dưới V3 thuộc VNI 30001, trong đó có máy chủ B.

Khi nhận được yêu cầu ARP, Máy chủ B sẽ gửi phản hồi ARP với DMAC = 00:00:30: 00:11:01 và SMAC = 00:00:30:00:11:02. V3 nhận gói tin và nó sẽ học luôn thông tin này và tra cứu điểm đến dựa trên [VNI = 30001, DMAC = 00:00:30:00:11: 01], gói được đóng gói VXLAN và gửi về phía V1.

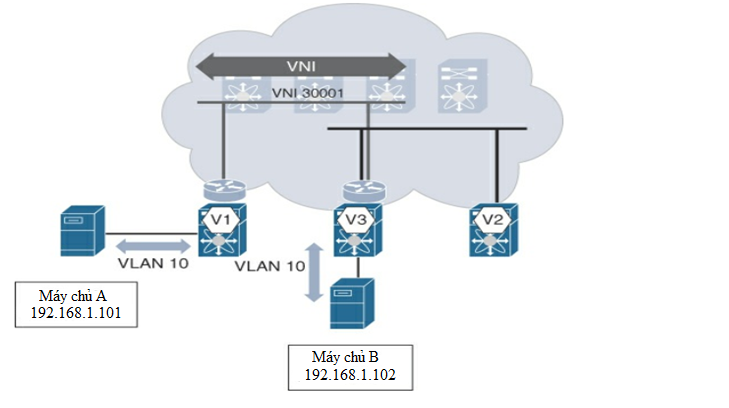
Các trường nổi bật trong gói tin là VNI = 30001, IP nguồn = 10.200.200.3 và IP đích = 10.200.200.1. Gói được đóng gói chuyển đến V1 thông qua định tuyến lớp 3. V1 cũng học thông tin lớp 2 và sau đó tra cứu đích dựa trên [VNI, DMAC] dẫn đến gói phản hồi ARP từ máy chủ B được gửi đến máy chủ A. Bằng cách này, thông tin MAC của máy chủ được lưu lại ở cả VTEP, máy chủ A và B về địa chỉ MAC-IP của nhau. Những gói tin tiếp theo giữa máy chủ A và B sau đó được truyền đơn hướng qua VXLAN giữa V1 và V3, tương ứng.

Thông qua cơ chế này, lưu lượng truy cập đa điểm được đẩy qua VXLAN giữa các VTEP để tìm hiểu về các máy chủ MAC nằm phía sau VTEP, lưu lượng tiếp theo có thể được truyền đơn hướng. Đây được gọi là cơ chế F&L. Do đó, giao tiếp trong VXLAN VNI tuân theo ngữ nghĩa chuyển tiếp lớp 2, giống như trường hợp trong VLAN. Sự khác biệt chính là miền VXLAN VNI trải dài mạng lớp 3 trên lớp mạng IP vật lý.

Nhiều VNI có thể chia sẻ cùng một nhóm phát đa hướng. Điều này là do việc có 16 triệu tùy chọn VNI tiềm năng và có một ánh xạ 1: 1 duy nhất giữa VNI và một nhóm đa hướng đòi hỏi lượng lớn tài nguyên phần mềm và phần cứng, khiến cho việc triển khai không thực tế. Trên thực tế, trong hầu hết các triển khai, tối đa 512 hoặc 1024 nhóm phát đa hướng được hỗ trợ. Điều này có nghĩa là một VTEP có thể nhận được lưu lượng truy cập đa điểm được đóng gói VXLAN cho một VNI không được cấu hình cục bộ. Sự cô lập ở cấp độ mạng vẫn được duy trì, vì mỗi phân đoạn mạng VXLAN có một VNI duy nhất và việc tra cứu luôn là [VNI, DMAC]. Một giải pháp thay thế cho phát IP đa hướng để xử lý lưu lượng truy cập đa điểm đến trong môi trường VXLAN là sử dụng sao chép xâm nhập (IR) hoặc sao chép đầu cuối. Với IR, mọi VTEP phải biết các VTEP khác là thành viên trong một VNI nhất định. VTEP nguồn tạo ra n bản sao cho mỗi khung, mỗi bản sao được gửi tới các VTEP khác là thành viên trong VNI tương ứng. Điều này đặt thêm gánh nặng lên các VTEP, nhưng nó có lợi ích là đơn giản hóa vì không cần chạy đa hướng trong lớp dưới IP.

VXLAN là một trong số ít các lớp mạng Overlay có thể kết hợp giữa mô hình máy ảo và máy thật trao đổi dữ liệu. Với sự phát triển của SDN, VXLAN đã được sử dụng nhiều như một lớp phủ máy chủ, với các máy chủ vật lý lưu trữ phần mềm VTEP thực hiện đóng gói và giải mã VXLAN cho lưu lượng giữa những dữ liệu được ảo hóa. Để thiết lập giao tiếp giữa các thiết bị cũ không hỗ trợ VXLAN, các cổng VXLAN lớp 2 đi đầu, cung cấp khả năng biên dịch giữa các miền dựa trên VXLAN kế thừa và VXLAN được ảo hóa. Các nhà cung cấp mạng bắt đầu hỗ trợ VXLAN trong các thiết bị chuyển mạch phần cứng của họ, do đó cung cấp các tùy chọn để sử dụng VXLAN làm lớp mạng Overlay. Điều này đã mở đường cho một môi trường kết hợp nơi cả VTEP phần mềm và phần cứng có thể được hỗ trợ đồng thời trong cùng một môi trường trung tâm dữ liệu.

VXLAN đã cung cấp một giải pháp đáng chú ý cho các vấn đề về khả năng mở rộng kém, nhu cầu nâng cao tính di động của mạng và khả năng quản lý mạng được bảo toàn. Tuy nhiên, một số nhược điểm với VXLAN đã được xác định. Trong khi VXLAN xác định một mặt phẳng dữ liệu có thể mở rộng cho các mạng ảo, một mặt phẳng điều khiển không được chỉ định về mặt này. Do đó, phương pháp tiếp cận dựa trên F&L gốc mà VXLAN sử dụng là không lý tưởng vì miền quảng bá cho VXLAN hiện trải dài qua ranh giới lớp 3. Nói chung, bất kỳ cơ chế nào dựa trên việc đẩy tràn lưu lượng sẽ có những mối quan tâm nghiêm trọng, đặc biệt là khi quy mô tăng lên. VXLAN F&L gặp phải những nhược điểm tương tự như mạng dựa trên Ethernet F&L. Vì vậy, trong khi VXLAN mang lại những lợi thế đáng kể trong một số khía cạnh liên quan đến nhu cầu ngày càng tăng của trung tâm dữ liệu, nó không đưa ra được một giải pháp toàn diện về tổng thể. Phần còn thiếu là một cơ chế để tìm hiểu về các máy chủ cuối đằng sau một VTEP mà không cần đến lưu lượng truy cập. Điều này được thực hiện bằng cách sử dụng phần mở rộng giao thức điều khiển dựa trên BGP. Hình 1.9 mô tả quá trình ánh xạ VLAN đến VNI khi gói tin được chuyển giữa các VTEP.



##### Hình 1.9: Ánh xạ VLAN đến VNI

### 1.3.2. Cơ chế BGP EVPN

Việc sử dụng VXLAN rõ ràng mang lại nhiều lợi thế cho các trung tâm dữ liệu đa đối tượng thông qua việc nâng cao số lượng định danh mạng và thông qua nền tảng mạng Overlay lớp 3 của nó. Nhưng trong khi giao thức STP được tránh bằng cách kết hợp lõi IP để truyền dữ liệu, các giao thức đa hướng vẫn được yêu cầu cơ chế tràn lan thông tin qua mạng lớp 3 khi tìm kiếm thiết bị đích trước khi định dạng từ máy chủ đến máy chủ cụ thể được học và xác định. Do đó, các VXLAN theo truyền thống đã sử dụng các phương pháp tiếp cận F&L khi vận chuyển dữ liệu qua mạng. Tất cả các thiết bị logic hay vật lý trong mạng đứng phía sau địa chỉ VTEP đích nhận dữ liệu từ thiết bị nguồn và sau khi nhận được địa chỉ đích dự kiến của nó, thông tin địa chỉ MAC cụ thể sẽ được cung cấp trở lại thiết bị nguồn để quá trình lấy thông tin diễn ra. Sau khi biết được, một kết nối trực tiếp từ máy chủ đến máy chủ được thiết lập trong phân đoạn mạng VXLAN.

Với suy nghĩ này, việc sử dụng giao thức BGP và Ethernet VPN (EVPN) giải quyết một số vấn đề liên quan đến việc học thông tin đầu cuối qua việc tràn lưu lượng bằng cách giảm thiểu mức độ mà điều này cần phải xảy ra. BGP là một giao thức được chuẩn hóa lâu đời để trao đổi thông tin khả năng tiếp cận tầng mạng (NLRI) để tạo điều kiện thuận lợi cho khả năng tiếp cận giữa các máy chủ trong mạng. BGP đã chứng minh hiệu quả rằng nó có thể mở rộng ở cấp độ Internet. Là một giao thức trạng thái cứng, một giao thức chỉ bắt đầu truyền tin khi có sự thay đổi nào đó trong thông tin về khả năng truy cập mạng mà nó đang quảng cáo, bản tin BGP đã được mở rộng theo nhiều cách khác nhau. Trong bối cảnh ảo hóa mạng, BGP cung cấp một thư mục dữ liệu của tất cả các địa chỉ MAC và IP của các thiết bị đằng sau một VTEP trong một phân đoạn mạng VXLAN. Do đó, BGP cho phép một cấu trúc rõ ràng về việc đánh số và đặt tên các điểm đến VXLAN được chuẩn hóa, điều này giúp loại bỏ việc phải dồn dập để tìm thông tin địa chỉ. Tương tự, đa giao thức BGP (M-BGP) đã được thiết lập tốt trong việc hoạt động trong các mạng IP với hỗ trợ đa mạng, và việc sử dụng nó kết hợp với EVPN cung cấp các khía cạnh mặt phẳng điều khiển cần thiết.

EVPN là một phần của M-BGP bổ sung đủ thông tin bên trong để trao đổi dữ liệu có thể diễn ra hiệu quả giữa các thiết bị đầu cuối. EVPN vận chuyển thông tin giữa các VTEP nguồn và đích trong VXLAN, sử dụng định dạng BGP/MP-BGP để xác định khả năng truy cập của các đích địa chỉ IP và MAC cụ thể đằng sau VTEP. Thông qua định tuyến và chuyển mạch tích hợp, EVPN tạo điều kiện vận chuyển cho cả lớp 2 và lớp 3, sử dụng địa chỉ các thiết bị đã biết có trong mạng VXLAN. Do đó, cơ chế tràn thông tin không cần thiết có thể giảm đáng kể, và trong một số trường hợp, nó có thể được loại bỏ hoàn toàn.

Tóm lại, họ địa chỉ EVPN cho phép thông tin MAC, IP, mạng, VRF và VTEP của máy chủ được truyền qua MP-BGP. Theo cách này, miễn là VTEP biết về máy chủ lưu trữ đằng sau nó (thông qua ARP / ND / DHCP, v.v.), BGP EVPN sẽ phân phối và cung cấp thông tin này cho tất cả các VTEP khác trong mạng. Miễn là VTEP nguồn tiếp tục biết máy chủ đằng sau nó, thì thông báo cập nhật EVPN sẽ không được gửi đi. Do đó, các VTEP khác không cần phải loại bỏ bất kỳ thông tin về khả năng truy cập máy chủ từ xa nào. Điều này ngăn chặn, gián đoạn đến lão hóa có thể xảy ra trong các tình huống F&L điển hình, do đó tăng cường đáng kể khả năng mở rộng mạng.

Để mở rộng khái niệm này xa hơn một chút, đôi khi người ta cho rằng BGP-EVPN loại bỏ hoàn toàn sự cần thiết của việc tràn thông tin, nhưng điều này thực sự không hợp lệ. VXLAN hoạt động như mặt phẳng dữ liệu của mạng và vận chuyển dữ liệu. Do đó, để xử lý lưu lượng BUM qua VXLAN, nguồn VTEP vẫn được yêu cầu để gửi lưu lượng truy cập đa điểm đến nhiều VTEP, bằng cách sử dụng IP đa hướng hoặc sao chép xâm nhập.

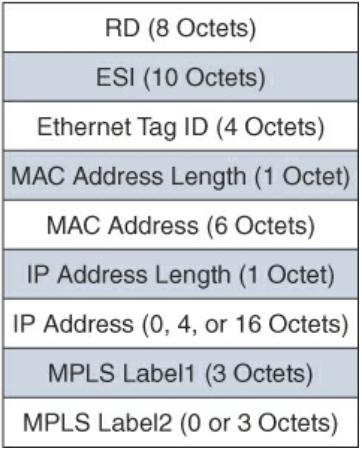
BGP EVPN khác với F&L theo một số cách. Cả việc tìm kiếm thiết bị ngang hàng và học máy chủ lưu trữ đầu cuối từ xa đều được xử lý theo cách hoàn toàn khác vì BGP EVPN phân phối các mã định danh điểm cuối (MAC và địa chỉ IP) cũng như thông tin tiền tố IP liên kết với điểm cuối đằng sau VTEP. Tương tự như vậy, xác thực VTEP ngang hàng thì khác, với mặt phẳng điều khiển chứa thông tin được yêu cầu thay vì chờ đợi các phản hồi để sau đó truyền đơn hướng. Ngoài ra, BGP EVPN thu hẹp phân phối tuyến máy chủ thông qua việc sử dụng mặt phẳng điều khiển này. Do đó, các yêu cầu ARP bị chặn trong mạng VXLAN, điều này giúp tăng cường đáng kể hiệu quả của việc truyền dữ liệu. Với những đặc điểm này, việc sử dụng BGP EVPN với VXLAN cung cấp một phương tiện cải tiến mà nhờ đó, việc vận chuyển dữ liệu được tạo điều kiện và kiểm soát để số lượng các máy chủ cụ thể được xác định trong một mạng có thể tăng lên theo cấp số nhân.

**Loại tuyến đường BGP EVPN**

Có 5 loại tuyến đường trong đó loại 1 và loại 4 hầu như không được sử dụng, còn lại đều là những loại tuyến đường quan trọng.

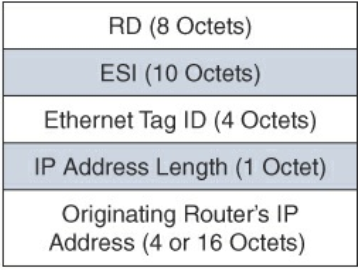
* Tuyến đường loại 1: Ethernet auto-discovery (AD) route
* Tuyến đường loại 2: MAC/IP advertisement route.
* Tuyến đường loại 3: Inclusive multicast Ethernet tag route.
* Tuyến đường loại 4: Ethernet segment route.
* Tuyến đường loại 5: IP prefix route.

Tuyến đường loại 2: xác định tuyến quảng cáo MAC/IP và chịu trách nhiệm phân phối thông tin khả năng truy xuất địa chỉ MAC và địa chỉ IP trong BGP EVPN (Hình 1.11). Tuyến đường loại 2 có nhiều trường khác nhau cung cấp các ngữ cảnh để giải quyết thông tin. Tuyến đường loại 2 có các trường như độ dài địa chỉ MAC và địa chỉ MAC là bắt buộc. Tương tự như vậy, một nhãn phải được điền để xác định VNI lớp 2 cho mặt phẳng dữ liệu VXLAN (nhãn MPLS 1). NLRI này cũng cho phép các trường tùy chọn, địa chỉ IP và độ dài địa chỉ IP. Thông thường trường độ dài địa chỉ IP có thể thay đổi, độ dài là 32 cho IPv4 và 128 cho IPv6. Khi sử dụng tuyến loại 2 cho thông tin bắc cầu, các thuộc tính bổ sung được đưa ra, chẳng hạn như loại đóng gói (Encap 8: VXLAN), mục tiêu tuyến đường (RT) hoặc chuỗi MAC di động để di chuyển điểm cuối hoặc phát hiện máy chủ trùng lặp. Các thuộc tính bổ sung cũng có trong trường hợp định tuyến, trong đó next hop của MAC Router sẽ được kết hợp như VNI thứ hai (VNI lớp 3) và RT, cho hoạt động lớp 3. Lưu ý rằng với tuyến loại 2, các giá trị RT và VNI riêng biệt được chỉ ra trong mỗi thông tin lớp 2 và lớp 3. Trong trường hợp không có thông tin IP trong thông báo tuyến loại 2, các trường tùy chọn địa chỉ IP và độ dài địa chỉ IP được xóa sạch và VNI lớp 3 (nhãn MPLS 2) bị bỏ qua.



##### Hình 1.10: BGP EVPN tuyến loại 2

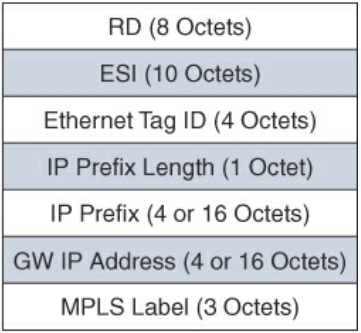
Tuyến đường loại 3: thường được sử dụng để tạo danh sách phân phối cho bản sao xâm nhập (IR). Điều này cung cấp một cách để tái tạo lưu lượng truy cập nhiều người qua việc truyền đơn hướng. Tuyến loại 3 ngay lập tức được tạo và gửi tới tất cả các VTEP ngay sau khi VNI được cấu hình tại VTEP và đang hoạt động. Điều này khác với tuyến loại 2, chỉ được gửi với thông tin MAC và IP, sau khi các máy chủ cuối đã được học. Bằng cách này, mọi VTEP đều biết tất cả các VTEP từ xa khác cần được gửi một bản sao của gói BUM trong một VNI nhất định. Hình 1.11 mô tả các trường trong tuyến loại 3.



##### Hình 1.11: BGP EVPN tuyến loại 3

Lưu ý rằng sao chép xâm nhập là một chức năng không chỉ phải được hỗ trợ bởi mặt phẳng điều khiển BGP EVPN mà còn phải được hỗ trợ bởi mặt phẳng dữ liệu. Ngoài ra, mặt phẳng dữ liệu phải có khả năng tạo n bản sao cho mỗi gói tin đa điểm, trong đó n là VTEP từ xa thuộc thành viên trong VNI liên quan. Vì khả năng của bảng dữ liệu này phụ thuộc vào ASIC phần cứng cụ thể lưu trữ VTEP, trong trường hợp không hỗ trợ sao chép xâm nhập, tùy chọn duy nhất để chuyển tiếp lưu lượng BUM là phát IP đa hướng. Do đó, lầm tưởng rằng EVPN loại bỏ yêu cầu phát đa hướng trong lớp dưới đã bị vô hiệu hóa.

Tuyến đường loại 5: Trong khi EVPN chủ yếu được xây dựng để vận chuyển thông tin lớp 2 và mang lại hiệu quả tốt hơn trong các hoạt động bắc cầu, định tuyến thể hiện nhu cầu chung trong các mạng Overlay về giao tiếp giữa các mạng con hoặc giao tiếp qua hai VNI Lớp 2. Tuyến loại 5 cung cấp khả năng vận chuyển thông tin tiền tố IP trong EVPN bằng cách cho phép vận chuyển tiền tố IPv4 và IPv6 với độ dài tiền tố IP thay đổi (0 đến 32 cho IPv4 và 0 đến 128 cho IPv6). Các tuyến tiền tố IP trên tuyến loại 5 không chứa thông tin MAC của lớp 2 trong NLRI, và do đó tuyến loại 5 chỉ kết hợp VNI lớp 3 cần thiết cho việc định tuyến và tích hợp đa mạng. Hơn nữa, các thuộc tính mở rộng của tuyến loại 5 mang mục tiêu tuyến đường (RT), kiểu đóng gói và router MAC của next hop VTEP trong mạng Overlay. Với tuyến loại 2, địa chỉ MAC được sử dụng để nhận dạng tuyến trong MAC VRF, trong khi đối với loại 5, tiền tố IP được sử dụng để nhận dạng tuyến trong IP VRF. Điều này cho phép phân tách rõ ràng cho các router BGP EVPN, do đó ngăn chặn mọi xử lý liên quan đến địa chỉ MAC đối với các tuyến tiền tố IP được EVPN quảng cáo (tuyến loại 5). Hình 1.12 mô tả các trường thông tin trong tuyến đường loại 5.



##### Hình 1.12: BGP EVPN tuyến loại 5

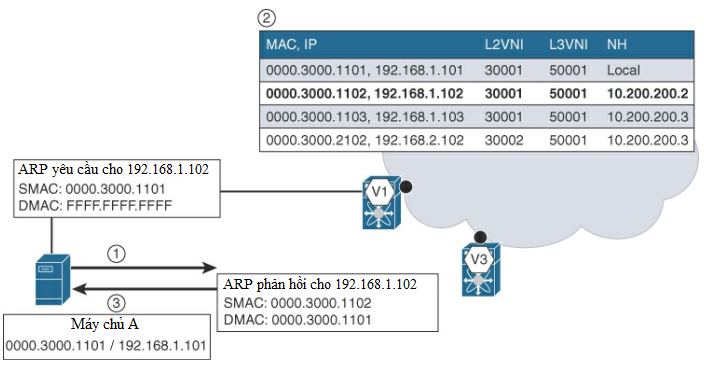
**Sự loại bỏ bản tin ARP**

Giao thức phân giải địa chỉ (ARP) chịu trách nhiệm về ánh xạ địa chỉ IPv4 đến địa chỉ MAC trong mạng IP. ARP tạo điều kiện thuận lợi cho việc lấy thông tin địa chỉ MAC của điểm cuối, tận dụng thông tin địa chỉ IP được gửi đi trong một yêu cầu quảng bá ARP. Yêu cầu quảng bá được coi là lưu lượng truy cập đa đích thường được đóng gói VXLAN và gửi đến mọi VTEP hoặc thiết bị biên là thành viên của VNI lớp 2 tương ứng. Phản hồi đối với yêu cầu ARP thường được gửi dưới dạng một gói tin đơn hướng tới người yêu cầu. Lưu lượng ARP nằm trong phạm vi giới hạn của miền quảng bá được đại diện bởi VNI lớp 2. Tương ứng, mạng IPv6 dựa vào giao thức khám phá lân cận (ND) cho độ phân giải địa chỉ IPv6 thành địa chỉ MAC. Với IPv6, độ phân giải này được kích hoạt thông qua lời kêu gọi hàng xóm ban đầu (NS) phát đa hướng thông qua miền quảng bá lớp 2. Quảng bá hàng xóm (NA) được gửi phản hồi từ đích sẽ hoàn thành giải pháp thiết lập hàng xóm.

Khi một điểm cuối cần giải quyết cổng mặc định, là điểm thoát của mạng con IP cục bộ, nó sẽ gửi một yêu cầu ARP đến cổng mặc định đã được cấu hình. Hoạt động ARP cho phép thiết bị biên cục bộ (VTEP) đưa vào ánh xạ IP/MAC của các điểm cuối được gắn cục bộ. Ngoài bảng MAC-IP được điền trên thiết bị biên, tất cả thông tin MAC được ghi vào giao thức mặt phẳng điều khiển BGP EVPN. Ngoài ra, đối với lớp 2, L2VNI, bộ phân biệt tuyến đường (RD), mục tiêu tuyến đường (RT), lưu trữ VTEP và thông tin IP liên kết được đưa vào, đối với lớp 3, L3VNI, RD, RT, lưu trữ VTEP và thông tin RMAC được điền. Cụ thể, thông tin này được đưa vào quảng bá tuyến đường loại 2 và được phân phối tới tất cả các VTEP từ xa đã biết về điểm cuối đó. Cụ thể, tất cả các yêu cầu quảng bá ARP được gửi đi từ một điểm cuối đều phải tuân theo việc sao chép và chuyển tiếp đến các điểm cuối khác. Thiết bị biên tìm thông tin về một điểm cuối miễn là điểm cuối gửi yêu cầu ARP — không nhất thiết phải là yêu cầu ARP để phân giải cổng mặc định được lưu trữ trên thiết bị biên cục bộ.

Thông thường, khi một điểm cuối muốn nói chuyện với một điểm cuối khác trong cùng một mạng con, nó sẽ gửi yêu cầu ARP để xác định địa chỉ IP với MAC của điểm cuối đích. Yêu cầu ARP được đưa vào tất cả các điểm cuối là một phần của VNI lớp 2 đó. ARP theo dõi cùng với thông tin mặt phẳng điều khiển BGP EVPN có thể giúp tránh việc tràn lưu lượng cho các điểm cuối đã biết. Bằng cách sử dụng ARP theo dõi, tất cả các yêu cầu ARP từ một điểm cuối được chuyển hướng đến thiết bị biên gắn cục bộ với nó. Sau đó, thiết bị biên sẽ trích xuất địa chỉ IP đích trong bản tin ARP và xác định xem nó có phải là một điểm cuối đã biết hay không. Cụ thể, một truy vấn được thực hiện dựa trên thông tin điểm cuối đã biết từ mặt phẳng điều khiển BGP EVPN.

Nếu đích đến được biết, thông tin liên kết địa chỉ IP và MAC sẽ được phản hồi. Sau đó, thiết bị biên cục bộ thực hiện ủy quyền ARP thay mặt cho điểm cuối đích. Nói cách khác, nó sẽ gửi một phản hồi ARP đơn hướng tới người yêu cầu với địa chỉ MAC được phân giải của điểm cuối đích đã biết. Theo cách này, tất cả các yêu cầu ARP tới các điểm cuối đã biết sẽ bị kết thúc sớm nhất có thể, đó là thiết bị biên được gắn cục bộ hoặc VTEP hoặc switch leaf. Điều này được gọi là sự loại bỏ ARP. Có thể ngăn chặn ARP vì tất cả thông tin đều được biết đến trên tất cả các VTEP.



##### Hình 1.13: Quá trình truyền ARP

Lưu ý rằng giảm thiểu ARP khác với ủy quyền ARP, trong đó thiết bị biên hoặc bộ định tuyến có thể đóng vai trò như một thiết bị ủy quyền thay mặt cho điểm cuối đích, sử dụng Router MAC của chính nó. Việc hạn chế bản tin ARP giúp giảm lưu lượng phát ARP bằng cách tận dụng BGP. Hình 1.13 minh họa một kịch bản trong đó bản tin ARP tại VTEP V1 kết thúc yêu cầu ARP sớm đối với máy chủ có địa chỉ 192.168.1.102 từ máy chủ có địa chỉ 192.168.1.101 trong lớp 2 VNI 30001. Điều quan trọng cần lưu ý là tính năng loại bỏ ARP hoạt động dựa trên thiết bị được bật trong VNI lớp 2, bất kể cổng mặc định có được cấu hình trên các VTEP hay không.

## 1.4. Kết luận chương

Máy chủ ảo hóa đang ngày càng áp dụng VXLAN. Tuy nhiên, hiếm khi có một môi trường hoàn toàn ảo hóa trong một trung tâm dữ liệu. Bình thường, trung tâm dữ liệu có các máy ảo, máy chủ, các thiết bị vật lý cùng tồn tại. Máy ảo cần truy cập các dịch vụ trên các máy chủ và thiết bị vật lý, điều này tạo ra nhu cầu về cổng cho các máy ảo trong phân đoạn VXLAN để giao tiếp với các thiết bị trong phân đoạn VLAN.

VXLAN cung cấp giải pháp mở rộng mạng lớp 2 trên cơ sở hạ tầng lớp 3 bằng cách đóng gói và tạo đường hầm MAC-trong-UDP. VXLAN cho phép vị trí các thiết bị linh hoạt thông qua phần mở rộng Lớp 2. Đây cũng là một cách tiếp cận để xây dựng một trung tâm dữ liệu đa đối tượng bằng cách tách các phân đoạn lớp 2 của các thiết bị khỏi mạng truyền tải được chia sẻ.

# CHƯƠNG II: MẠNG TRUNG TÂM DỮ LIỆU THẾ HỆ MỚI N-DCN

## 2.1. Mạng trung tâm dữ liệu

Các trung tâm dữ liệu đã phát triển đáng kể trong vài năm qua. Sự phát triển này đã diễn ra nhanh chóng và trong một khoảng thời gian tương đối ngắn, mang lại các công nghệ phổ biến như ảo hóa, đám mây (riêng tư, công khai và kết hợp), mạng được xác định bằng phần mềm (SDN) và dữ liệu lớn. Đối với thời đại ưu tiên di động và đám mây, quy mô, sự nhanh nhạy, bảo mật, hợp nhất và tích hợp với bộ điều phối máy tính lưu trữ là các yêu cầu chung của trung tâm dữ liệu. Ngoài ra, khả năng hiển thị, tự động hóa, dễ quản lý, khả năng hoạt động, khắc phục sự cố và phân tích nâng cao cũng được mong đợi là một phần của các giải pháp trung tâm dữ liệu ngày nay.

Sự thay đổi từ quản lý từng thiết bị sang một hệ thống tập trung đã diễn ra. Giao diện lập trình ứng dụng mở và giao thức dựa trên tiêu chuẩn ngăn chặn việc khóa nhà cung cấp duy nhất là tiêu chí hàng đầu trong hầu hết các yêu cầu đề xuất của khách hàng. Mạng LAN ảo mở rộng (VXLAN) với kết cấu dựa trên BGP Ethernet VPN (BGP EVPN) trình bày một giải pháp trung tâm dữ liệu thống nhất.

### 2.1.1. Khái niệm

Trung tâm dữ liệu là cơ sở tập trung các hoạt động và thiết bị công nghệ thông tin chung của một tổ chức nhằm mục đích lưu trữ, xử lý và phổ biến dữ liệu và ứng dụng. Bởi vì chúng là nơi chứa các tài sản độc quyền và quan trọng nhất của một tổ chức, các trung tâm dữ liệu rất quan trọng đối với việc các hoạt động hàng ngày diễn ra liên tục. Do đó, tính bảo mật, độ tin cậy của các trung tâm dữ liệu và thông tin của chúng là một trong những ưu tiên hàng đầu của bất kỳ tổ chức nào.

Trước đây, các trung tâm dữ liệu là cơ sở hạ tầng vật lý được kiểm soát cao, nhưng đám mây công cộng sau này đã thay đổi mô hình đó. Ngoại trừ trường hợp các hạn chế quy định yêu cầu trung tâm dữ liệu tại những nơi không có kết nối internet, hầu hết cơ sở hạ tầng trung tâm dữ liệu hiện đại đã phát triển từ các máy chủ vật lý sang cơ sở hạ tầng ảo hóa hỗ trợ các ứng dụng và khối lượng công việc trên các môi trường đa đám mây.

### 2.1.2. Vai trò của trung tâm dữ liệu

Trung tâm dữ liệu là một phần không thể thiếu của doanh nghiệp, được thiết kế để hỗ trợ các ứng dụng kinh doanh và cung cấp các dịch vụ như:

* Lưu trữ, quản lý, sao lưu và phục hồi dữ liệu
* Các ứng dụng năng suất, chẳng hạn như email
* Giao dịch thương mại điện tử khối lượng lớn
* Hỗ trợ cộng đồng trò chơi trực tuyến
* Big Data, machine learning và trí tuệ nhân tạo

Ngày nay, có hơn 7 triệu trung tâm dữ liệu trên toàn thế giới. Trên thực tế, mọi doanh nghiệp và tổ chức chính phủ đều xây dựng và duy trì trung tâm dữ liệu của riêng mình hoặc có quyền truy cập vào của người khác. Nhiều lựa chọn có sẵn ngày nay, chẳng hạn như thuê máy chủ tại cơ sở địa điểm, sử dụng dịch vụ trung tâm dữ liệu do bên thứ ba quản lý hoặc sử dụng các dịch vụ dựa trên đám mây công cộng từ các máy chủ như Amazon, Microsoft, Sony và Google.

### 2.1.3. Thành phần trong trung tâm dữ liệu.

Các kiến trúc và yêu cầu của trung tâm dữ liệu có thể khác nhau đáng kể. Ví dụ: một trung tâm dữ liệu được xây dựng cho một nhà cung cấp dịch vụ đám mây như Amazon đáp ứng các yêu cầu về cơ sở vật chất, cơ sở hạ tầng và bảo mật khác biệt đáng kể so với một trung tâm dữ liệu hoàn toàn riêng tư, chẳng hạn như một trung tâm được xây dựng cho một cơ sở chính phủ chuyên bảo mật dữ liệu đã phân loại.

Bất kể phân loại nào, hoạt động trung tâm dữ liệu hiệu quả đạt được thông qua đầu tư cân bằng vào cơ sở vật chất và trang thiết bị mà trung tâm này có. Ngoài ra, vì các trung tâm dữ liệu thường chứa dữ liệu và ứng dụng quan trọng trong kinh doanh của một tổ chức, nên điều cần thiết là cả cơ sở và thiết bị phải được bảo mật trước những kẻ xâm nhập và tấn công mạng. Các phần tử chính của trung tâm dữ liệu gồm:

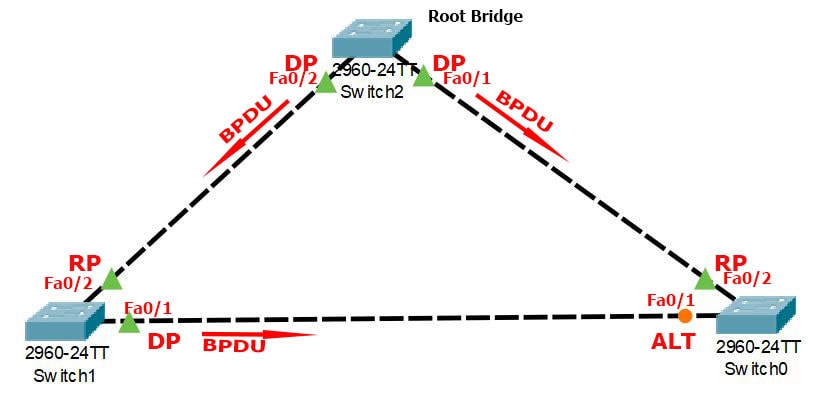
* Cơ sở vật chất - không gian có thể sử dụng được dành cho thiết bị công nghệ thông tin. Cung cấp quyền truy cập thông tin liên tục khiến các trung tâm dữ liệu trở thành một trong những cơ sở tiêu tốn nhiều năng lượng nhất trên thế giới. Thiết kế để tối ưu hóa không gian và kiểm soát môi trường để giữ thiết bị trong phạm vi nhiệt độ / độ ẩm cụ thể đều được nhấn mạnh.
* Các thành phần cốt lõi - thiết bị và phần mềm cho hoạt động công nghệ thông tin, lưu trữ dữ liệu và ứng dụng. Chúng có thể bao gồm hệ thống lưu trữ, máy chủ, cơ sở hạ tầng mạng, chẳng hạn như thiết bị chuyển mạch và bộ định tuyến, các yếu tố bảo mật thông tin khác nhau, chẳng hạn như tường lửa.
* Cơ sở hạ tầng hỗ trợ - thiết bị góp phần duy trì an toàn tính khả dụng cao nhất có thể. Viện Thời gian hoạt động đã xác định bốn cấp trung tâm dữ liệu, với mức độ khả dụng nằm trong khoảng từ 99,671% đến 99,995%. Một số thành phần để hỗ trợ cơ sở hạ tầng bao gồm:
* Nguồn điện liên tục (UPS) - pin dự phòng, máy phát điện và nguồn điện dự phòng.
* Kiểm soát môi trường - máy điều hòa không khí phòng máy tính; hệ thống sưởi, thông gió và điều hòa không khí; và hệ thống xả.
* Hệ thống an ninh vật lý - sinh trắc học và hệ thống giám sát video.
* Nhân viên vận hành - nhân sự sẵn sàng giám sát hoạt động và bảo trì thiết bị CNTT và cơ sở hạ tầng suốt ngày đêm.

### 2.1.3. Quá trình phát triển giao thức trong mạng trung tâm dữ liệu

Các mạng dựa trên giao thức Spanning Tree (STP) đã phục vụ các yêu cầu mạng trong vài năm. Kênh cổng ảo (vPC) được giới thiệu để giải quyết một số nhược điểm của mạng STP đồng thời cung cấp khả năng hoạt động kép. Sau này đã phát triển thành lớp mạng ảo Overlay MAC trong IP với sự phát minh của VXLAN.

Trong khi các mạng lớp 2 đã phát triển vượt ra ngoài các cấu trúc liên kết không có vòng lặp với STP, các chức năng cổng mặc định cho lớp 3 cũng trở nên phức tạp hơn. Các cổng tập trung truyền thống được lưu trữ tại các lớp phân phối hoặc tổng hợp đã chuyển sang triển khai cổng phân tán. Điều này đã cho phép mở rộng quy mô và loại bỏ các điểm tắc nghẽn. Phần này cung cấp một cái nhìn tổng quan về lý do và cách thức tiến hóa này diễn ra.

Xây dựng mạng lớp 2 không vòng lặp cho phép các thiết bị trong cùng một miền quảng bá giao tiếp với nhau, là bước phát triển ban đầu khiến mạng cục bộ (LAN) trở nên cực kỳ phổ biến. Trong một miền quảng bá, chuyển tiếp dựa trên tra cứu địa chỉ MAC. Giao thức STP được tiêu chuẩn hóa như IEEE 802.1D quy định một phương pháp để tạo cấu trúc liên kết cây, theo định nghĩa, không có vòng lặp. Bằng cách biến chức năng này thành một chức năng “cắm và chạy”, các mạng dựa trên STP có hỗ trợ IEEE 802.1Q đã được chấp nhận nhanh chóng và vẫn đang được sử dụng cho đến ngày nay. Hình 2.1 mô tả tổng quan về giao thức STP, trong cùng một miền lớp 2, dựa trên thuộc tính sẵn có, cấu hình thiết bị để bình bầu vai trò giữa các switch cũng như các cổng giao diện trên switch để xác định các liên kết sử dụng truyền dữ liệu và liên kết bị chặn nhằm dự phòng cho các liên kết đang hoạt động.

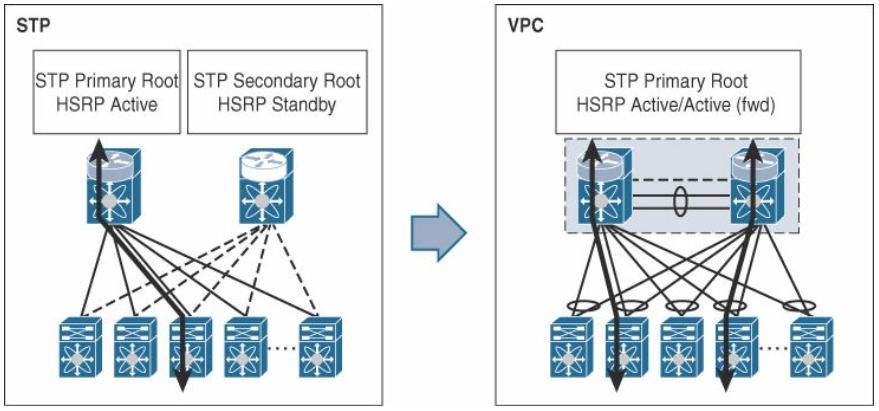


##### Hình 2.1: Giao thức Spanning Tree (STP)

Với nhiều công nghệ, những thiếu sót trở nên rõ ràng khi một công nghệ hoặc giao thức được thiết kế cho một môi trường cụ thể nhưng sau đó được áp dụng cho một môi trường khác có một bộ yêu cầu hoàn toàn khác. Điều này đã xảy ra với các mạng dựa trên STP. Nhiều vấn đề khác nhau làm cho việc áp dụng STP cho một mạng trung tâm dữ liệu lớn trở nên khó khăn. Một số vấn đề chính bao gồm sau đây:

* **Các vấn đề về hội tụ**: Khi lỗi liên kết hoặc lỗi chuyển mạch trong các mạng dựa trên STP, liên kết cần được tính toán lại. Điều này có thể ảnh hưởng đáng kể đến thời gian hội tụ lưu lượng vì các thông báo thay đổi cấu trúc liên kết dẫn đến việc xóa các bảng MAC trên các thiết bị chuyển mạch. Do đó, việc phân cấp lại là bắt buộc. Vấn đề này trở nên khuếch đại khi nút gốc gặp sự cố và một nút mới cần được bầu chọn. Do đó, thời gian hội tụ cao là một trong những nhược điểm chính của mạng dựa trên STP.
* **Các liên kết không được sử dụng**: Như đã đề cập trước đó, STP xây dựng một liên kết để đảm bảo rằng cấu trúc mạng không vòng lặp. Tiềm ẩn của việc này là nhiều liên kết giữa những bộ chuyển mạch bị chặn. Do đó, tài nguyên mạng không được sử dụng tối ưu. Hiện tại, có khả năng tận dụng tất cả các liên kết và băng thông tương ứng của chúng để đạt được việc sử dụng tối ưu các tài nguyên hiện có là một mục tiêu mong muốn.
* **Chuyển tiếp dưới mức tối ưu**: Vì lưu lượng được bắt nguồn từ một switch cụ thể với STP, tất cả lưu lượng từ switch đó sang bất kỳ switch nào khác trong cấu trúc liên kết sẽ được chuyển tiếp dọc theo một tuyến đường duy nhất. Bởi vì lưu lượng truy cập luôn được chuyển tiếp dọc theo một hướng nhất định, hướng còn lại sẽ không được sử dụng. Do đó, lưu lượng giữa các thiết bị switch đó sẽ được chuyển tiếp dưới mức tối ưu và đây chắc chắn là một trường hợp không hề lý tưởng.
* **Thiếu định tuyến đa đường giá trị ngang bằng (**ECMP): Vì chỉ có một đường dẫn hoạt động giữa bộ chuyển mạch nguồn và đích trong mạng lớp 2 dựa trên STP truyền thống, các tùy chọn ECMP không có. Mặt khác, mạng lớp 3 cung cấp khả năng tận dụng nhiều hơn một đường dẫn giá trị ngang bằng giữa một cặp bộ định tuyến. Đây là điều mong muốn và là một trong những lý do chính khiến mạng lớp 3 trở nên phổ biến.
* **Vấn đề về bão lưu lượng**: Với cấu trúc liên kết trong giao thức STP, lưu lượng truy cập bắt đầu từ một bộ chuyển mạch sẽ không được gửi trở lại chính nó. Tuy nhiên, lưu lượng truy cập có thể vẫn lặp lại liên tục trong mạng ở một số trạng thái lỗi. Một sự cố như vậy có thể làm hỏng toàn bộ mạng. Kịch bản này đã được gọi là một cơn bão phát sóng. Bất kỳ loại bão mạng nào cũng có thể làm tiêu hao băng thông không cần thiết trong mạng và cần phải tránh bằng mọi giá. Do không có trường thời gian tồn tại (TTL) trong tiêu đề lớp 2, một khi cơn bão xuất hiện, lưu lượng truy cập có thể chảy vô tận. Mạng lớp 3 có cơ chế an toàn do sự hiện diện của trường TTL và sự giảm liên quan tại mỗi bước nhảy được định tuyến. Khi TTL trở thành 0, gói tin sẽ bị loại bỏ. Việc thiếu trường như vậy trong mạng lớp 2 hạn chế rất nhiều khả năng mở rộng, đặc biệt là khi quy mô của mạng ngày càng lớn.
* **Thiếu hỗ trợ dual-homing**: STP vốn dĩ không cho phép gắn một thiết bị hoặc máy chủ vào nhiều hơn một switch. Khi điều đó được cố gắng, một vòng lặp được hình thành và STP về cơ bản chặn một trong các liên kết. Do đó, từ quan điểm dự phòng hoặc khả năng chịu lỗi, nếu switch đó gặp sự cố, lưu lượng truy cập đến và đi từ thiết bị hoặc máy chủ bên dưới sẽ bị mất đi cho đến khi cây được tính toán lại.
* **Quy mô mạng**: Trong thời đại điện toán đám mây, có thể gặp phải những hạn chế đáng kể nếu chỉ có khoảng 4 nghìn VLAN. Ngay cả việc triển khai trung tâm dữ liệu quy mô trung bình cũng có thể lưu trữ một số đối tượng thuê, mỗi đối tượng có thể có một tập hợp các mạng, tổng hợp lại với nhau vượt quá giới hạn này. Thật không may, do số nhận dạng 12 bit đại diện cho một VLAN hoặc miền quảng bá trong tiêu đề IEEE 802.1Q hoặc dot1q, 4 nghìn khá hạn chế. Những tiến bộ đáng kể trong công nghệ mạng dẫn đến con số này bị vượt quá trong một khoảng thời gian tương đối ngắn.

Các công nghệ như PortChannel ảo (vPC), Multichassis EtherChannel (MCEC) và hệ thống chuyển mạch ảo (VSS) cho phép một thiết bị đầu cuối (máy chủ hoặc bộ chuyển mạch) gắn vào một cặp thiết bị chuyển mạch. Tất cả những công nghệ này đều nằm dưới sự hỗ trợ của tổng hợp kiên kết nhiều thiết bị (MC-LAG). Với vPC, một cặp switch (được gọi là vPC peers) được cấu hình sao cho phần còn lại của mạng coi cặp này là một switch logic duy nhất.

Hình 2.2 cho thấy cách một mạng STP điển hình có thể được thực hiện hiệu quả hơn với vPC. Thiết bị đầu cuối được gắn vào cả hai switch sử dụng vPC bằng cách sử dụng cấu hình PortChannel hoặc EtherChannel. Cả hai liên kết đều ở trạng thái active, do đó cho phép chuyển tiếp lưu lượng một cách chủ động. lưu lượng đến thiết bị đầu cuối có thể chảy qua một trong hai thiết bị ngang hàng. Lưu lượng truy cập đa điểm đến thiết bị đầu cuối chỉ được gửi qua một trong các thiết bị ngang hàng, do đó tránh được sự trùng lặp. 

##### Hình 2.2: Sự phát triển từ STP sang vPC

Mỗi thiết bị ngang hàng vPC có một vai trò chính và một vai trò phụ. Thông tin giữa các thiết bị vPC được đồng bộ hóa qua một kênh điều khiển được gọi là liên kết ngang hàng vPC. Thông tin này bao gồm thông tin MAC, ARP và khám phá hàng xóm (ND) cũng như các thông số kiểm tra cấu hình và tính nhất quán khác nhau. vPC không chỉ cung cấp “dual-homing” từ quan điểm lớp 2, mà việc chuyển tiếp active-active của nó cũng mở rộng sang các giao thức dự phòng bước đầu tiên (FHRP) như giao thức bộ định tuyến chờ nóng (HSRP) và giao thức dự phòng bộ định tuyến ảo (VRRP). Điều này cho thấy một sự cải thiện đáng kể so với triển khai FHRP truyền thống với hoạt động ở chế độ chờ khi chỉ có thiết bị active là chuyển tiếp lưu lượng dữ liệu. Với sự xuất hiện của vPC, cả hai vPC ngang hàng chuyển tiếp lưu lượng dữ liệu đồng thời, trong khi FHRP đang hoạt động chỉ giải quyết các yêu cầu ARP / ND từ góc độ mặt phẳng điều khiển.

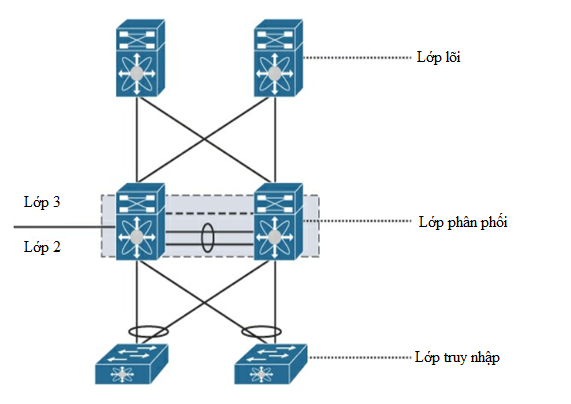
Trong khi vPC giải quyết một số hạn chế của STP, nó vẫn bị giới hạn ở một cặp switch. Một giải pháp đa đường tốt hơn được mong muốn, đặc biệt là khi phát triển các yêu cầu xây dựng miền lớp 2 có thể mở rộng quy mô lớn. Khi số lượng điểm cuối trong mạng lớp 2 mở rộng, việc tìm hiểu tất cả các địa chỉ MAC của điểm cuối trở nên khó khan đối với tất cả các thiết bị chuyển mạch trong mạng. Ngoài ra, với việc IP là tiêu chuẩn thực tế trong ngành công nghiệp mạng, việc thúc đẩy đóng gói lớp phủ dựa trên IP đã xảy ra. Do đó, VXLAN đã được giới thiệu.

VXLAN, cách đóng gói MAC trong IP/UDP, hiện là cách đóng gói lớp phủ phổ biến nhất được sử dụng. Là một tiêu chuẩn mở, nó đã nhận được sự chấp nhận rộng rãi từ các nhà cung cấp mạng. VXLAN giải quyết tất cả các giới hạn STP được mô tả trước đó. Tuy nhiên, với VXLAN, một số 24 bit xác định phân đoạn mạng ảo, do đó cho phép hỗ trợ lên đến 16 triệu miền quảng bá trái ngược với giới hạn khoảng 4 nghìn truyền thống được áp đặt bởi VLAN. Vì VXLAN chạy qua mạng IP nên tính năng ECMP của mạng lớp 3 sẵn có để sử dụng. Nói chung, một lớp mạng Overlay chẳng hạn như VXLAN chạy trên lớp 3 có thể sử dụng địa chỉ phân cấp với IP và bất kỳ phương tiện truyền tải nào.

Các thiết bị switch biên trong mạng VXLAN được gọi là thiết bị biên hay điểm cuối đường hầm VXLAN (VTEP). Các thiết bị biên chịu trách nhiệm đóng gói và giải mã tiêu đề VXLAN. Các switch trong lõi kết nối các VTEP với nhau đóng vai trò như các bộ định tuyến IP thông thường. Đáng chú ý, chúng không cần phải có bất kỳ chức năng phần cứng hoặc phần mềm chuyên biệt nào. Ngoài ra, các thiết bị chuyển mạch trong mạng VXLAN tìm hiểu về nhau bằng cách sử dụng các giao thức định tuyến thông thường như OSPF, IS-IS lớp 3,…

## 2.2. Những yêu cầu phát triển mạng trung tâm dữ liệu thế hệ mới

Cấu trúc liên kết ba tầng cổ điển đã phục vụ ngành công nghiệp mạng trong một khoảng thời gian khá dài (xem hình 2.3). Thông thường, các điểm cuối hoặc máy chủ được gắn vào lớp truy cập, lớp này chỉ hỗ trợ lớp 2 với cấu hình cổng chuyển mạch và VLAN thích hợp. Đối với giao tiếp giữa các mạng con, lưu lượng truy cập từ các điểm cuối được định tuyến thông qua chuyển mạch lớp phân phối.



##### Hình 2.3: Mô hình 3 lớp

Cổng mặc định cho các điểm cuối được lưu trữ tại lớp phân phối ở dạng giao diện chuyển mạch ảo (SVI) lớp 3. Để dự phòng, các thiết bị chuyển mạch lớp tổng hợp được cấu hình với FHRP như HSRP hoặc VRRP. Các switch lớp phân phối lần lượt được kết nối với nhau thông qua các switch lớp lõi. Các giao thức lớp 3 thích hợp cho phép trao đổi thông tin tiền tố giữa chuyển mạch lớp phân phối và lớp lõi.

Với việc triển khai trung tâm dữ liệu quy mô lớn, cấu trúc liên kết ba tầng đã trở thành điểm nghẽn quy mô. Khi số lượng điểm cuối trong mạng mở rộng, số lượng liên kết IP-MAC được lưu trữ trong các switch lớp phân phối cũng tăng lên, do đó yêu cầu hỗ trợ cho một bảng cơ sở thông tin chuyển tiếp (FIB) lớn. Ngoài ra, mặt phẳng điều khiển về việc làm mới ARP và ND cần được xử lý bởi các switch này cũng mở rộng đáng kể. Vấn đề này, cùng với số lượng giao diện lớp 3 và đối tượng VRF tương ứng có thể được lưu trữ tại các switch lớp phân phối, gây ra các vấn đề lớn.

Ngoài ra, tốc độ triển khai là yếu tố cần thiết và điều quan trọng là có thể mở rộng quy mô khi yêu cầu khối lượng công việc mở rộng. Các yêu cầu về khả năng truy cập có tính khả dụng cao vào các ứng dụng và dữ liệu liên quan của chúng ngày càng trở nên nghiêm ngặt. Trong khi nhu cầu về tính sẵn sàng được thúc đẩy bởi sự toàn cầu hóa rộng rãi của các doanh nghiệp trên khắp các châu lục, các ứng dụng cũng ngày càng trở nên phức tạp.

Trong bối cảnh mới này, nơi tất cả khối lượng công việc và ứng dụng đang được chuyển sang trung tâm dữ liệu, các thiết kế truyền thống cho trung tâm dữ liệu không còn đủ để đáp ứng tất cả các yêu cầu. Đây là trường hợp bất kể trung tâm dữ liệu tồn tại kết hợp với đám mây riêng, đám mây công cộng hay đám mây lai. Với hiện tại môi trường, một số nhu cầu chính của việc triển khai trung tâm dữ liệu ngày nay như sau:

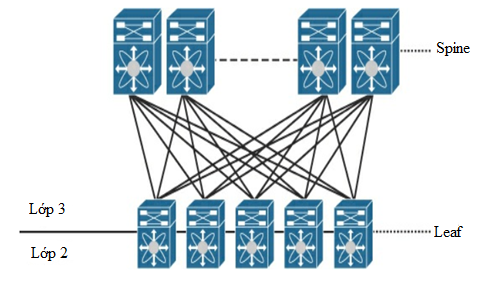
* **Khả năng triển khai nhanh**: Sự linh hoạt xác định thời gian thực hiện một yêu cầu ứng dụng. Trung tâm dữ liệu nhanh là một trung tâm có thể giảm thời gian này xuống mức tối thiểu.
* **Khả năng mở rộng**: Khả năng mở rộng là điều quan trọng, đặc biệt là trong các trung tâm dữ liệu dựa trên đám mây. Một trung tâm dữ liệu phải có thể chứa hàng nghìn và vài nghìn mạng lưới người thuê. Giới hạn mạng 4094 (hoặc 4K) do trường VLAN 12 bit áp đặt là không đủ để hỗ trợ các trung tâm dữ liệu nhiều đối tượng lớn.
* **Tính co giãn:** Một trung tâm dữ liệu phải có khả năng thích ứng với các nhu cầu và yêu cầu thay đổi. Điều này có thể liên quan đến việc bổ sung khối lượng công việc tính toán, bộ nhớ tăng cường, băng thông mạng cao, v.v. Một trung tâm dữ liệu phải có khả năng bổ sung dung lượng để giải quyết sự gia tăng nhu cầu mà không ảnh hưởng đến khối lượng công việc ứng dụng hiện có.
* **Tính khả dụng**: Một trung tâm dữ liệu phải có khả năng hoạt động liên tục (24/7, 365 ngày một năm). Ngoài ra, quyền truy cập vào các ứng dụng cần có sẵn trên tất cả các loại thiết bị (chẳng hạn như máy tính bảng, điện thoại thông minh và đồng hồ thông minh), điều này nâng cao mô hình thiết bị của riêng bạn lên một cấp độ hoàn toàn mới. Để có tính khả dụng cao, các yêu cầu về khôi phục sau lỗi (DR) thường đưa ra các tùy chọn dự phòng tiềm năng có sẵn có thể tiếp quản bất cứ khi nào xảy ra lỗi tại trung tâm dữ liệu chính.
* **Chi phí thấp**: Tổng chi phí sở hữu (TCO) cho một trung tâm dữ liệu bao gồm cả chi phí vốn (CAPEX) và chi phí hoạt động (OPEX). Trong khi phần CAPEX được phân bổ theo thời gian, phần OPEX là một khoản chi liên tục. Do đó, OPEX chịu sự giám sát chặt chẽ của hầu hết các CIO/CFO. Do đó, giảm OPEX cho các trung tâm dữ liệu thường là một mục ưu tiên cao.
* **Tính mở:** Để ngăn chặn mối ràng buộc với một nhà cung cấp duy nhất, việc thúc đẩy xây dựng trung tâm dữ liệu với các tùy chọn dựa trên tiêu chuẩn tồn tại từ cả quan điểm phần cứng và phần mềm. Hầu hết các trung tâm dữ liệu quy mô lớn đã chuyển sang mô hình hộp trắng có thương hiệu (White-box). Doanh nghiệp và nhà cung cấp dịch vụ cũng đang yêu cầu triển khai trung tâm dữ liệu dựa trên tiêu chuẩn mở.
* **Bảo mật**: Đặc biệt trong triển khai trung tâm dữ liệu nhiều đối tượng, yêu cầu hàng đầu là áp đặt các chính sách bảo mật hiệu quả để đảm bảo rằng lưu lượng truy cập từ một đối tượng thuê hoàn toàn bị cô lập với đối tượng thuê khác. Một số yêu cầu khác liên quan đến bảo mật bao gồm việc thực thi các chính sách ứng dụng, ngăn chặn truy cập trái phép, phát hiện các mối đe dọa, tách biệt các thiết bị bị tấn công, phân phối các bản vá an toàn cho các thiết bị ảnh hưởng và các ứng dụng chính sách nhất quán giữa đám mây riêng và đám mây công cộng.
* **Định hướng giải pháp**: Những ngày triển khai từng hộp trong một trung tâm dữ liệu đã trôi qua. Các trung tâm dữ liệu ngày nay yêu cầu một giải pháp thống nhất với nhiều phần khác nhau theo quan điểm mạng và chúng cũng yêu cầu tích hợp chặt chẽ với bộ điều phối máy tính và lưu trữ cũng như các thiết bị dịch vụ (vật lý và ảo). Ngoài ra, cần có sự tự động hóa tinh vi kết hợp với bộ điều khiển SDN trong không gian cạnh tranh cao này.
* **Dễ sử dụng**: Ngay cả với cách tiếp cận theo định hướng giải pháp, việc quản lý, giám sát liên tục và khả năng hiển thị vào các hoạt động hàng ngày của trung tâm dữ liệu là rất mong muốn. Tính dễ sử dụng có tác động trực tiếp đến việc giảm OPEX.
* **Hỗ trợ triển khai kết hợp**: Cả doanh nghiệp và nhà cung cấp dịch vụ đều đã áp dụng mô hình đám mây ở một mức độ nhất đinh. Do đó, một trong những yêu cầu quan trọng của trung tâm dữ liệu là hỗ trợ triển khai đám mây lai trong đó tài nguyên từ đám mây công cộng có thể được mở rộng sang trung tâm dữ liệu doanh nghiệp tư nhân một cách linh hoạt. Các hoạt động như vậy phải liền mạch theo quan điểm của các ứng dụng. Nói cách khác, các ứng dụng phải hoàn toàn không biết liệu chúng được lưu trữ tại cơ sở hay ngoài cơ sở.
* **Hiệu quả sử dụng năng lượng:** Một phần lớn chi phí hoạt động của trung tâm dữ liệu là do các yêu cầu về năng lượng điện của nó. Ngành công nghiệp mạng tại các công ty lớn và các nhà cung cấp trung tâm dữ liệu nói riêng nhận thức được yêu cầu này, và động cơ để xây dựng các trung tâm dữ liệu xanh chắc chắn tồn tại.

## 2.3. Mạng trung tâm dữ liệu thế hệ mới

Với sự ra đời của VXLAN và sự mở rộng cho họ địa chỉ BGP EVPN, giờ đây có thể phân phối các liên kết điểm cuối (IP, MAC) tới VTEP. Theo cách này, tại thời điểm một điểm cuối được học ngay tại một VTEP kết nối trực tiếp với nó, sử dụng BGP EVPN, thông tin về khả năng tiếp cận này có thể được phân phối cho tất cả các VTEP quan tâm. Mọi lưu lượng truy cập đến điểm cuối này sau đó có thể được chuyển tiếp tối ưu từ bất kỳ VTEP nào khác. Trong khi việc loại bỏ việc tràn lưu lượng lớp 2 chắc chắn giúp khả năng mở rộng ở một mức độ lớn trong mạng VXLAN với hỗ trợ BGP EVPN, lưu lượng lớp 3 cũng có thể được chuyển tiếp tối ưu trong các mạng.

Đối với khoảng 4 nghìn VLAN đơn thuần, các switch lớp phân phối trong cấu trúc mạng 3 lớp có thể gặp phải những thách thức lớn về khả năng mở rộng. Do đó, với VXLAN trong hình và sự hỗ trợ của nó cho 16 triệu mạng ảo, cấu trúc liên kết ba tầng truyền thống không thể tồn tại gánh nặng như vậy. Nhìn chung, ngành công nghiệp đã nhận ra hạn chế này và một sự thúc đẩy mạnh mẽ đối với việc chuyển lớp 3 sang lớp truy cập và thoát khỏi việc tiếp cận cổng mặc định lớp phân phối đã diễn ra.

Với cấu trúc mạng trung tâm dữ liệu thế hệ mới (N-DCN), cổng mặc định được di chuyển đến lớp truy cập sẽ giảm bớt các miền lỗi vì switch lớp truy cập chỉ cần hỗ trợ giao thức “bước đầu” cho các thiết bị cuối bên dưới nó. Điều này chắc chắn đại diện cho một mô hình có khả năng mở rộng hơn, so với trường hợp một hoặc một cặp switch lớp phân phối phải hỗ trợ tất cả các điểm cuối bên dưới các switch lớp truy cập. Điều này đã sinh ra cấu trúc spine-leaf.



##### Hình 2.4: Cấu trúc spine-leaf

Như trong hình 2.4, kiến trúc spine-leaf là mô hình 2 lớp với switch lớp truy cập hỗ trợ lớp 3 thường được gọi là leaf (switch leaf). Các switch cung cấp liên kết giữa các switch leaf khác nhau được gọi là spine (switch spine). Tất cả các switch leaf kết nối với các swicth spine trong cấu trúc này. Tuyến đường được chọn ngẫu nhiên để tải lưu lượng được phân bổ đồng đều giữa các thiết bị switch lớp trên. Nếu một trong những switch lớp trên bị lỗi, nó sẽ chỉ làm giảm nhẹ hiệu suất trong toàn bộ trung tâm dữ liệu.

Ngoài ra, việc phân phối thông tin khả năng tiếp cận với BGP EVPN cho phép thực hiện cổng anycast IP phân tán trên cấu trúc mạng này. Cùng một cổng mặc định có thể được cấu hình đồng thời tại bất kỳ thiết bị switch leaf nào khi cần thiết. Bằng cách này, khi các thiết bị cuối di chuyển giữa các switch leaf , nó luôn tìm thấy cổng mặc định được gắn trực tiếp vào nó. Mọi thiết bị cuối có thể được đặt bên dưới bất kỳ switch leaf nào. Với cổng anycast IP phân tán, mọi lưu lượng được định tuyến sẽ được chuyển tiếp tối ưu từ leaf nguồn đến leaf đích. Các thiết bị spine cung cấp khả năng dự phòng cần thiết, do đó cung cấp cấu trúc liên kết mạng đơn giản với đủ dự phòng.

### 2.3.1. Những đặc tính trong mô hình

Kết cấu có một số đặc tính mong muốn đặc biệt hấp dẫn đối với các mạng trung tâm dữ liệu. Cấu trúc liên kết tương đối đơn giản này cung cấp khả năng mở rộng, khả năng phục hồi và hiệu quả được nâng cao. Không có sự quá tải lưu lượng và với băng thông đường lên và đường xuống bằng nhau ở mọi thiết bị leaf, kết cấu cung cấp băng thông phân cao đặc biệt hấp dẫn cho các ứng dụng trung tâm dữ liệu.

Với mạng trục hai lớp, mỗi switch leaf có thể tiếp cận mọi switch khác thông qua một bước nhảy qua một thiết bị spine đã chọn. Điều này giúp xác định độ trễ cho tất cả lưu lượng truy cập được chuyển tiếp qua kết cấu. Việc thêm một switch spine làm cho số tuyến đường có giá trị bằng nhau này tăng lên một. ECMP áp dụng như nhau cho cả lưu lượng đơn hướng và đa hướng cũng như cho cả lưu lượng lớp 2 (bắc cầu) và lớp 3 (định tuyến) với cổng IP bất kỳ phân tán ở lớp dưới (leaf).

Mạng trục này có tính đàn hồi, linh động trong việc xử lý lỗi liên kết và các nút, do đó nó cung cấp mức độ dự phòng cao. Việc mất một spine làm giảm băng thông mạng trục có sẵn, nhưng lưu lượng truy cập tiếp tục được chuyển tiếp theo cách tối ưu mặc dù tốc độ giảm. Tương tự, khi lỗi liên kết, lưu lượng truy cập được chuyển hướng đến các tuyến đường có sẵn và thời gian phục hồi bị chi phối bởi tốc độ hội tụ của các giao thức lớp mạng Overlay và Underlay.

Trong một kết cấu, tất cả các mạng con có thể được cấu hình đồng thời trên tất cả các thiết bị leaf , cung cấp khả năng kiểm soát mọi thiết bị cuối ở bất kỳ đâu trong mạng. Nếu việc quá tải lưu lượng của một liên kết xảy ra (nghĩa là, nếu lượng truy cập được tạo ra nhiều hơn mức mà cấu trúc liên kết đang hoạt động có thể cung cấp), thì quá trình mở rộng dung lượng rất đơn giản. Một switch spine được thêm vào, và các liên kết lên có thể được mở rộng cho mọi switch leaf, dẫn đến băng thông tăng lên giữa các lớp và giảm quá tải lưu lượng. Nếu dung lượng cổng của thiết bị trở thành mối quan tâm, một switch leaf mới có thể được thêm vào bằng cách kết nối nó với các switch spine và thêm cấu hình mạng cho các switch này. Việc dễ dàng mở rộng sẽ tối ưu hóa quy trình mở rộng mạng của bộ phận công nghệ thông tin.

Loại mạng này cũng có thể được thiết kế để có thể hỗ trợ hàng chục đến hàng trăm đến hàng nghìn cổng máy chủ 10G. Với cấu hình mong muốn, cấu trúc mềm dẻo để hỗ trợ nhiều môi trường công việc khác nhau với các công suất khác nhau.

### 2.3.2. Ưu điểm

Thiết kế mạng spine - leaf ban đầu được thực hiện trong các trung tâm dữ liệu như một cách để cải thiện hiệu suất khi xử lý lưu lượng. Nó làm được điều đó phần lớn bằng cách giảm số lượng “bước nhảy” giữa hai thiết bị bất kỳ trong mạng xuống chỉ còn một, bởi vì mọi thiết bị switch leaf trong mạng đều có kết nối trực tiếp với tất cả các switch spine. Ngược lại, kiến trúc mạng ba lớp truyền thống có nhiều bước nhảy, tùy thuộc vào nơi lưu lượng truy cập cần lưu thông. Nhưng đây thực sự chỉ là một trong những ưu điểm của kiến trúc spine-leaf so với kiến trúc truyền thống. Sau đây là một vài ưu điểm khác.

* Cải thiện khả năng dự phòng: Trong kiến trúc spine - leaf, tất cả switch leaf được kết nối với nhiều switch spine. Điều này cung cấp mức dự phòng cao hơn so với mô hình ba lớp thường được triển khai bằng cách sử dụng giao thức STP, để ngăn chặn các vòng lặp mạng. STP cho phép các đường dẫn dự phòng giữa hai điểm, với chỉ một trong số chúng hoạt động tại bất kỳ thời điểm nào.
* Cải thiện hiệu suất: Khả năng sử dụng nhiều đường dẫn mạng cùng lúc cũng cải thiện hiệu suất. Với STP, nếu con đường khả dụng duy nhất bị tắc nghẽn, hiệu suất sẽ bị ảnh hưởng. Hơn nữa, chỉ có một bước nhảy giữa hai điểm bất kỳ cũng tạo ra một đường mạng rõ ràng hơn, điều này cũng có thể cải thiện hiệu suất.
* Cải thiện khả năng mở rộng: Các cấu trúc liên kết spine-leaf cũng có khả năng mở rộng. Cung cấp rất nhiều đường dẫn giữa hai điểm mạng bất kỳ, tất cả chúng đều có sẵn để truyền lưu lượng, làm giảm khả năng tắc nghẽn ngay cả trong một mạng lớn. Việc thêm thiết bị switch vào mô hình spine-leaf cung cấp thêm các tuyến đường, do đó tăng khả năng mở rộng.
* Hỗ trợ các thiết bị switch cấu hình cố định, ít chi phí: Các thiết bị switch cấu hình cố định ít tốn kém hơn so với các thiết bị switch được yêu cầu trong mạng ba lớp để cung cấp mật độ cổng cần thiết cho phép số kết nối thích hợp giữa các thiết bị chuyển mạch ở các lớp khác nhau. Kiến trúc spine-leaf cho phép tất cả các cổng trên switch spine hỗ trợ kết nối với switch leaf, thay vì với các thiết bị spine khác. Ngoài ra, nó cho phép các kết nối được lan truyền giữa một số lượng lớn các switch spine.
* Thích ứng với doanh nghiệp: Mặc dù đúng là kiến ​​trúc spine-leaf ban đầu được thiết kế cho các mạng trung tâm dữ liệu, để giải quyết hướng đi lưu lượng giữa các máy chủ và hệ thống lưu trữ, kiến ​​trúc cũng có thể được mở rộng bên ngoài trung tâm dữ liệu cho doanh nghiệp mạng nói chung - mang lại nhiều lợi ích giống nhau và hơn thế nữa.
* Như được cấu hình cho các trung tâm dữ liệu, về cơ bản, kiến ​​trúc spine-leaf thu gọn lớp lõi và lớp phân phối thành lớp spine, trong khi leaf gần giống với lớp truy cập trong mô hình truyền thống.
* Trong doanh nghiệp, cách tiếp cận là khác nhau vì thách thức là khác nhau. Kiến trúc truyền thống đang chịu áp lực chủ yếu do sự gia tăng đáng kể số lượng thiết bị kết nối ở tầng truy cập. Đây là kết quả của hiện tượng Internet vạn vật (IoT) với tiềm năng hàng nghìn cảm biến và thiết bị kết nối ở biên mạng. Xu hướng mang thiết bị của riêng mình là một trình điều khiển khác, với việc mỗi người dùng kết nối hai, ba hoặc nhiều thiết bị với mạng, mỗi thiết bị có khả năng tạo ra lượng lớn lưu lượng truy cập từ các ứng dụng bao gồm cả video. Việc sử dụng nhiều các dịch vụ đám mây cũng có nghĩa là nhiều lưu lượng truy cập vào mạng lưới trên đường tới đám mây. Để giải quyết với tất cả lưu lượng đến ở biên mạng, cần thực hiện các bước ở lớp truy cập, trái ngược với lớp lõi và lớp phân phối như với mạng trung tâm dữ liệu truyền thống.
* Với spine-leaf, có thể loại bỏ giao thức STP giữa switch lớp truy cập và switch spine, thay vào đó sử dụng MLAG hay vPC. Với STP cho phép mỗi swicth lớp truy cập có một cặp kết nối với nhau để dự phòng. Nhưng với vPC, cả hai liên kết đều có thể hoạt động mọi lúc mà không bị mất khả năng dự phòng. vPC đồng bộ hóa trạng thái chuyển tiếp giữa chúng, vì vậy nếu switch spine hay switch leaf bị lỗi, lưu lượng sẽ tự động được định tuyến lại để có thời gian hoạt động liên tục.

### 2.3.3. Nhược điểm

Spine-leaf không phải là không có khuyết điểm. Một hạn chế là số switch lớn để đạt được quy mô yêu cầu. Các cấu trúc liên kết spine-leaf trong trung tâm dữ liệu cần mở rộng quy mô đến mức mà chúng có thể hỗ trợ các máy chủ vật lý kết nối với chúng. Số lượng switch leaf cần thiết để liên kết tất cả các vật chủ vật lý càng lớn, thì switch spine cần phải rộng hơn.

Lớp spine chỉ có thể kéo dài đến một điểm nhất định trước khi các thiết bị switch spine hết cổng và không thể kết nối thêm với các switch leaf, hoặc tỷ lệ mức quá tải giữa hai lớp spine và leaf là không thể chấp nhận được. Nói chung, tỷ lệ mô tả quá mức 3: 1 giữa lớp spine và leaf được coi là chấp nhận được. Ví dụ: 48 máy chủ kết nối với lớp leaf ở tốc độ 10Gbps sử dụng tối đa tiềm năng là 480Gbps. Nếu lớp leaf kết nối với lớp spine bằng 4 liên kết lên 40Gbps, thì băng thông kết nối liên kết là 160Gbps, tỷ lệ mức quá tải là 3: 1.

Mạng lưới spine-leaf cũng có các yêu cầu về hệ thống cáp đáng kể. Số lượng cáp cần thiết giữa các thiết bị spine-leaf tăng lên khi có thêm switch spine. Spine càng rộng thì càng cần nhiều liên kết với nhau. Thách thức đối với các nhà quản lý trung tâm dữ liệu là cấu trúc các nhà máy cáp để có đủ sợi cáp quang để kết nối các lớp với nhau. Ngoài ra, việc kết nối các công tắc cách xa nhau hàng chục mét đòi hỏi các mô-đun quang học đắt tiền, làm tăng thêm chi phí tổng thể của việc triển khai mô hình spine-leaf. Mặc dù có những mô-đun đồng giá rẻ hữu ích cho khoảng cách ngắn, nhưng mô-đun quang học là cần thiết và là một chi phí đáng kể trong các trung tâm dữ liệu hiện đại.

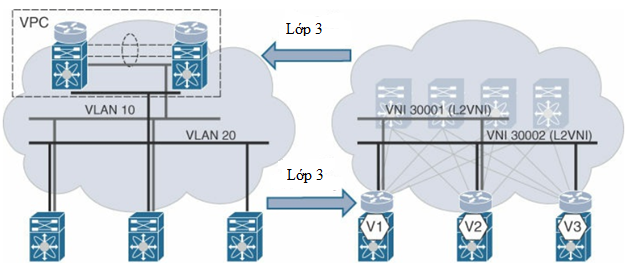
## 2.4. Tính năng VXLAN BGP EVPN trong mạng N-DCN

### 2.4.1. Cổng IP anycast phân phối

Để hai điểm cuối trong các mạng con IP khác nhau giao tiếp với nhau, cần phải có cổng mặc định. Theo truyền thống, cổng mặc định đã được triển khai tập trung tại lớp phân phối trung tâm dữ liệu, trong một cấu hình dự phòng. Để một điểm cuối tiếp cận cổng mặc định tập trung, trước tiên nó phải đi qua mạng lớp 2. Các tùy chọn mạng lớp 2 bao gồm Ethernet, vPC, và thậm chí cả VXLAN F&L. Giao tiếp trong cùng một mạng con IP thường được bắc cầu mà không có bất kỳ giao tiếp nào với cổng mặc định tập trung. Để định tuyến giao tiếp giữa các mạng IP hay các mạng con khác nhau, cổng mặc định tập trung có thể truy cập được qua cùng một đường dẫn mạng lớp 2.

Thông thường, mạng được thiết kế để dự phòng với khả năng sử dụng cao. Tương tự như vậy, cổng tập trung có sẵn. Các giao thức dự phòng bước đầu tiên (FHRP) được thiết kế để hỗ trợ cổng mặc định tập trung có dự phòng. Các giao thức như HSRP, VRRP là những ví dụ về FHRP. HSRP và VRRP có một thiết bị duy nhất chịu trách nhiệm phản hồi các yêu cầu ARP cũng như định tuyến lưu lượng truy cập đến mạng IP, mạng con khác. Khi thiết bị chính bị lỗi, FHRP sẽ thay đổi trạng thái hoạt động trên nút dự phòng thành chủ. Cần một khoảng thời gian nhất định để chuyển từ nút này sang nút khác.

FHRPs trở nên linh hoạt hơn khi được triển khai với các kênh cổng ảo (vPC), cho phép cả hai nút chuyển tiếp lưu lượng định tuyến và chỉ cho phép nút chính phản hồi các yêu cầu ARP. Kết hợp vPC và FHRP giúp tăng cường đáng kể khả năng phục hồi cũng như thời gian hội tụ trong trường hợp hỏng hóc. Bật đồng bộ hóa ARP cho FHRP trong môi trường vPC cho phép đồng bộ hóa ARP giữa vPC chính và vPC phụ. FabricPath với anycast HSRP tăng số lượng cổng hoạt động từ hai lên bốn nút. Trao đổi giao thức FHRP và các thay đổi trạng thái hoạt động vẫn hiện diện với anycast HSRP. Với FHRPs, cổng mặc định, được triển khai ở lớp tổng hợp trung tâm dữ liệu, cung cấp cổng mặc định tập trung cũng như ranh giới mạng lớp 2 hay lớp 3.



##### Hình 2.5: Vị trí cổng mặc định

Tầm quan trọng ngày càng tăng của các hoạt động lớp 2 và lớp 3, đặc biệt là trong kết cấu trung tâm dữ liệu lớn, đã đòi hỏi khả năng phục hồi lớn hơn so với những gì đã được cung cấp bởi các giao thức FHRP truyền thống. Việc di chuyển ranh giới mạng lớp 2 và lớp 3 đến switch leaf (lớp truy cập) giúp giảm thiểu lỗi. Cách tiếp cận mở rộng quy mô của cổng anycast IP phân tán làm giảm đáng kể trạng thái mạng và giao thức. Việc triển khai cổng anycast IP phân tán tại mỗi switch leaf không còn yêu cầu mỗi điểm cuối phải đi qua một miền lớp 2 lớn để đến cổng mặc định. Hình 2.5 mô tả sự chuyển dịch cổng mặc định ở lớp phân phối cũ sang vị trí mới tại lớp truy cập.

Cổng anycast IP phân tán áp dụng khái niệm mạng anycast “một cho liên kết gần nhất”. Anycast là một phương pháp định tuyến và địa chỉ mạng trong đó lưu lượng dữ liệu từ một điểm cuối được định tuyến theo cấu trúc liên kết đến nút gần nhất trong một nhóm các cổng đều được xác định bởi cùng một địa chỉ IP đích. Với cổng anycast IP phân tán, cổng mặc định được di chuyển đến gần điểm cuối - cụ thể là đến switch leaf nơi mỗi điểm cuối được gắn vào vật lý. Cổng anycast đang hoạt động trên mỗi thiết bị biên VTEP trên toàn bộ kết cấu mạng, loại bỏ yêu cầu phải có các gói giao thức “hello” truyền thống trên kết cấu mạng. Do đó, cùng một cổng cho một mạng con có thể tồn tại đồng thời ở nhiều switch leaf, nếu cần, mà không cần bất kỳ giao thức nào giống FHRP.

Mạng VXLAN BGP EVPN cung cấp các dịch vụ lớp 2 và lớp 3 và sự kết hợp cổng mặc định tồn tại giữa thiết bị biên cục bộ và điểm cuối. Khi điểm cuối cố gắng tìm đến cổng mặc định, thiết bị biên VTEP được kết nối trực tiếp giải quyết yêu cầu ARP đó. Bằng cách này, mọi thiết bị biên VTEP đều có trách nhiệm thực hiện chức năng cổng mặc định cho các điểm cuối gắn trực tiếp vào nó. Thiết bị biên cũng đảm nhận việc theo dõi trạng thái hoạt động của các điểm cuối sau nó bằng cách gửi bản tin ARP định kỳ.

### 2.4.2. Tính di động thiết bị

BGP EVPN bổ sung một cơ chế để cung cấp tính di động của điểm cuối trong mạng. Khi một điểm cuối di chuyển, tiền tố mà máy chủ liên kết được quảng bá trong mặt phẳng điều khiển thông qua một số thứ tự được cập nhật. Số thứ tự tránh được việc phải rút lại và học lại một tiền tố nhất định trong quá trình di chuyển điểm cuối. Thay vào đó, một bản cập nhật trong mặt phẳng điều khiển của vị trí mới được thực hiện. Vì việc chuyển tiếp trong kết cấu luôn được quyết định bởi những gì có trong mặt phẳng điều khiển BGP EVPN, lưu lượng truy cập nhanh chóng được chuyển hướng đến vị trí cập nhật của điểm cuối đã di chuyển, do đó cung cấp sự hội tụ lưu lượng thông suốt. Khi một điểm cuối di chuyển, hai tiền tố máy chủ cho cùng một điểm cuối sẽ xuất hiện trong mặt phẳng điều khiển BGP EVPN. Tiền tố thứ nhất được xác định bởi vị trí VTEP ban đầu và sau khi di chuyển, tiền tố được xác định bởi vị trí VTEP mới.

Địa chỉ MAC của điểm cuối trước khi di chuyển được học dựa trên tuyến đường EVPN loại 2 và chuỗi di động MAC – thuộc tính mở rộng BGP được đặt thành 0. Giá trị 0 cho biết rằng địa chỉ MAC không diễn ra sự di rời nào và điểm cuối vẫn ở vị trí ban đầu. Nếu một sự kiện di rời địa chỉ MAC diễn ra, một tuyến đường loại 2 mới (quảng bá MAC/IP) sẽ được thêm vào mặt phẳng điều khiển BGP EVPN bởi VTEP mới bên dưới điểm cuối di chuyển (vị trí mới của nó). Mặt phẳng điều khiển sau đó đặt số thứ tự di động MAC thành 1 trong thuộc tính BGP mở rộng cũng như ở vị trí mới. Hiện có hai thông tin quảng bá MAC/IP giống hệt nhau trong mặt phẳng điều khiển BGP EVPN. Nhưng thông tin quảng bá có số thứ tự di động MAC đang thiết lập giá trị 1 sẽ được tất cả các VTEP ưu tiên lưu trữ trong nó, do đó gửi lưu lượng truy cập đến điểm cuối tại vị trí mới của nó. Một điểm cuối có thể di chuyển nhiều lần trong suốt thời gian tồn tại của nó trong mạng. Mỗi khi điểm cuối di chuyển, VTEP phát hiện vị trí mới của nó sẽ tăng số thứ tự lên 1 và sau đó quảng bá tiền tố máy chủ cho điểm cuối đó vào mặt phẳng điều khiển BGP EVPN. Vì thông tin BGP EVPN được đồng bộ hóa trên tất cả các VTEP, nên mọi VTEP đều biết liệu một điểm cuối sắp xuất hiện lần đầu tiên hay đó là sự kiện di chuyển điểm cuối dựa trên thông tin về khả năng truy cập điểm cuối trước đó. Ngoài ra, một điểm cuối có thể bị tắt hoặc không thể truy cập được vì nhiều lý do. Nếu điểm cuối này không còn được lưu trong bảng ARP, MAC và BGP, thì chuỗi di động MAC cộng đồng mở rộng cũng bằng 0. Nếu cùng một điểm cuối xuất hiện lại (theo quan điểm của mạng và BGP EVPN), lần xuất hiện lại này được coi là một sự kiện tìm hiểu mới. Nói cách khác, mặt phẳng điều khiển BGP EVPN nhận biết được các điểm cuối hoạt động hiện tại và vị trí tương ứng của chúng, nhưng mặt phẳng điều khiển không lưu trữ bất kỳ lịch sử nào về các điểm cuối trước đó.

Bằng cách sử dụng số thứ tự di động MAC được mang trong tuyến đường Loại 2 (quảng bá MAC/IP), mặt phẳng điều khiển BGP EVPN có thể xác định khi nào một điểm cuối có thể xảy ra thay đổi vị trí. Khi một tiền tố nhất định được xác định là vẫn đang hoạt động, mặt phẳng điều khiển sẽ chủ động xác minh xem một điểm cuối cụ thể đã thực sự di chuyển hay chưa. Mặt phẳng điều khiển cung cấp một lượng giá trị lớn không chỉ trong việc xác minh và làm mới các sự kiện di động mà còn phát hiện các bản sao trong mạng VXLAN BGP EVPN

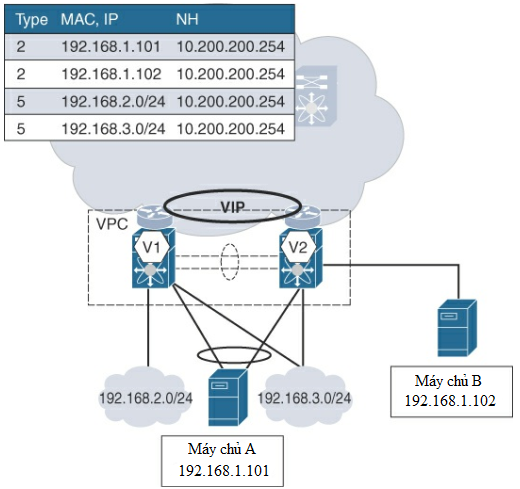
### 2.4.3. Kênh cổng ảo

Việc gộp nhiều giao diện vật lý vào một giao diện logic duy nhất giữa hai thiết bị được gọi là nhóm tổng hợp liên kết. Kênh cổng ảo (vPC) là công nghệ cung cấp khả năng dự phòng lớp 2 trên hai hoặc nhiều thiết bị vật lý. Cụ thể, một thiết bị bất kì được kết nối với hai thiết bị khác cấu hình như một cặp vPC, thuật ngữ cho điều này gọi là nhóm tổng hợp liên kết đa khung (MC-LAG). Nhớ lại rằng trong khi mạng truyền tải cơ bản cho VXLAN được xây dựng với giao thức định tuyến lớp 3 tận dụng ECMP, các thiết bị đầu cuối vẫn kết nối với các thiết bị switch leaf thông qua môi trường Ethernet truyền thống. Cần lưu ý rằng mạng VXLAN BGP EVPN không bắt buộc các điểm cuối phải có các kết nối dự phòng. Tuy nhiên, trong hầu hết các triển khai thực tế, tính sẵn sàng cao thường là một yêu cầu và vì mục đích đó, các điểm cuối được kết nối với các thiết bị biên qua các cổng giao diện.

Một số giao thức tồn tại để hình thành các kênh cổng ảo (vPC), các giao thức như giao thức tổng hợp cổng (PAgP) và các giao thức tiêu chuẩn trong mạng như giao thức điều khiển tổng hợp liên kết (LACP). Các kênh cổng ảo có thể được cấu hình giữa một điểm cuối duy nhất và một bộ chuyển mạch mạng. Kênh cổng ảo cho phép các giao diện của một điểm cuối kết nối vật lý với hai thiết bị chuyển mạch mạng khác nhau. Từ góc độ điểm cuối, họ thấy một bộ chuyển mạch duy nhất được kết nối qua một kênh duy nhất với nhiều liên kết. Điểm cuối được kết nối với miền vPC có thể là bộ chuyển mạch, máy chủ hoặc bất kỳ thiết bị mạng nào khác hỗ trợ tiêu chuẩn IEEE 802.3. vPC cho phép tạo các kênh cổng giao diện lớp 2 sử dụng trên hai switch . vPC bao gồm hai switch thành viên vPC được kết nối bằng một liên kết ngang hàng, với một thiết bị chính và một thiết bị dự phòng. Hệ thống được tạo bởi các thiết bị chuyển mạch mạng được gọi là miền vPC.

Miền vPC được thể hiện trong hình 2.6. VTEP V1 có địa chỉ IP 10.200.200.1/32 và VTEP V2 có địa chỉ IP 10.200.200.2/32. Đây là các địa chỉ IP vật lý riêng lẻ trên giao diện VTEP. Địa chỉ IP phụ được thêm vào hai VTEP, đại diện cho địa chỉ IP ảo (VIP) (cụ thể là 10.200.200.254/32). Địa chỉ ảo mô tả vị trí của bất kỳ điểm cuối nào được gắn vào cặp vPC. Địa chỉ này là bước tiếp theo được quảng bá trong mặt phẳng điều khiển BGP EVPN, đại diện cho vị trí của tất cả các điểm cuối cục bộ bên dưới một cặp switch vPC nhất định.

VIP được quảng bá từ cả hai thiết bị chuyển mạch thành viên vPC để cả hai thành viên vPC có thể nhận lưu lượng truy cập trực tiếp từ bất kỳ điểm cuối nào được gắn cục bộ. Các VTEP đầu xa có thể tiếp cận địa chỉ VIP được quảng bá bởi cả thiết bị chuyển mạch thành viên vPC qua ECMP trên mạng định tuyến IP vật lý. Do đó, việc trao đổi dữ liêu giữa các điểm cuối có thể đạt được bằng ít nhất một tuyến đường bất kỳ thông qua một switch thành viên vPC.



##### Hình 2.6: VPC với VXLAN BGP EVPN

Trong miền vPC, điểm cuối có 2 kết nối hay điểm cuối có một kết nối đến thiết bị biên đều được quảng bá cho phép truy cập thông qua địa chỉ VIP được liên kết với VTEP. Do đó, đối với các điểm cuối mồ côi bên dưới các switch thành viên vPC, một đường dẫn dự phòng được thiết lập thông qua liên kết ngang hàng vPC, nếu khả năng truy cập tới các switch spine bị mất do lỗi đường lên. Vì vậy, bạn nên sử dụng một định tuyến vật lý cận kề qua liên kết ngang hàng vPC để giải quyết các tình huống lỗi.

## 2.5. Kết luận chương

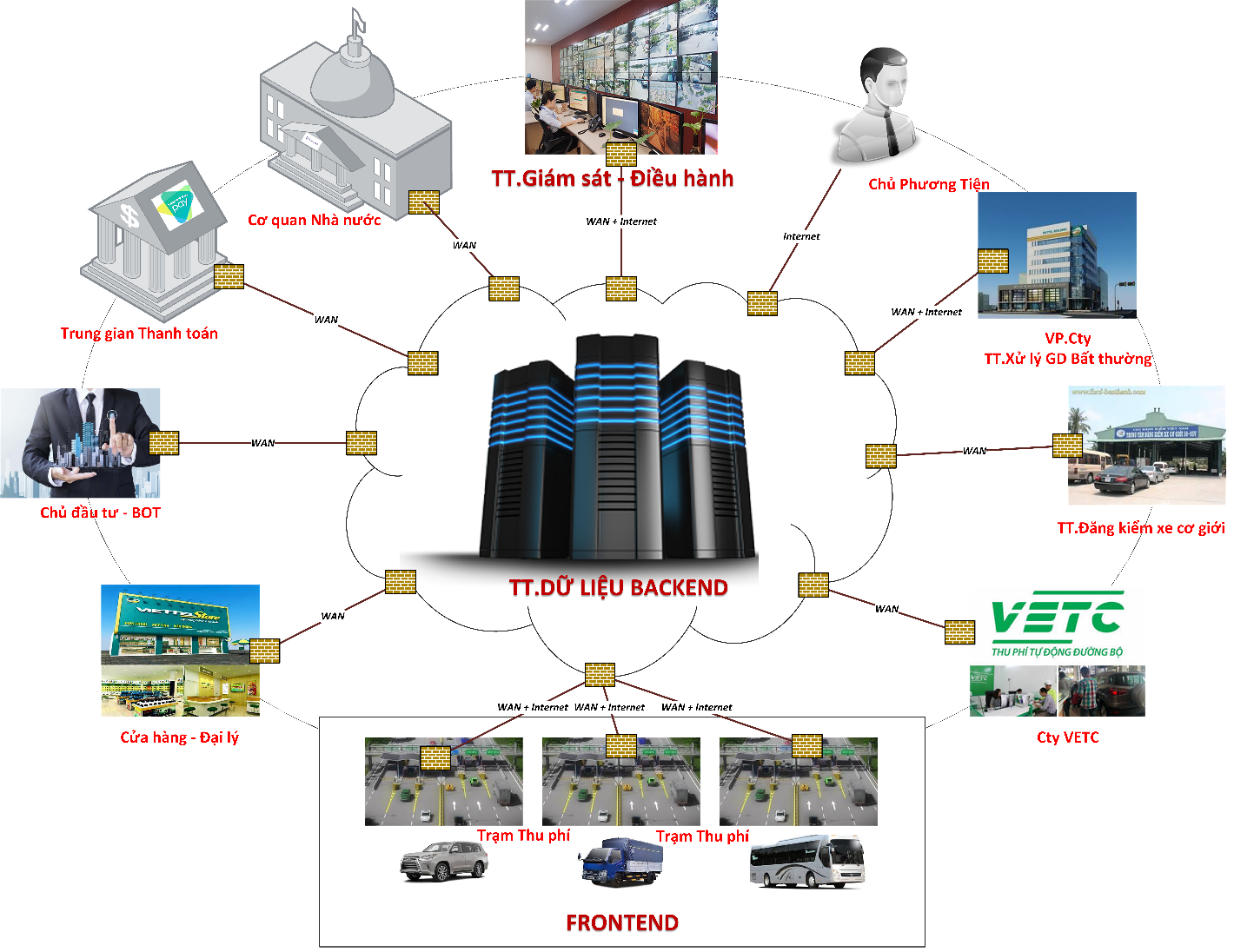
Mạng Leaf-Spine cung cấp nhiều lợi ích độc đáo so với mô hình 3 lớp truyền thống. Việc sử dụng định tuyến lớp 3 với ECMP cải thiện tổng băng thông bằng cách sử dụng tất cả các liên kết có sẵn. Với cấu hình và thiết kế dễ điều chỉnh, Leaf-Spine đã cải thiện khả năng quản lý và khả năng mở rộng thiết kế. Loại bỏ Spanning Tree (STP) đã dẫn đến sự ổn định mạng được cải thiện đáng kể. Sử dụng các công cụ mới và khả năng khắc phục các hạn chế vốn có với các giải pháp khác như SDN, môi trường Leaf-Spine cho phép các bộ phận CNTT và trung tâm dữ liệu phát triển mạnh trong khi hoàn thành mọi nhu cầu và mong muốn của doanh nghiệp.

# CHƯƠNG III: ỨNG DỤNG CÔNG NGHỆ VXLAN TRONG MẠNG N-DCN CHO TRẠM THU PHÍ KHÔNG DỪNG ETC

## 3.1. Kiến trúc tổng thể hệ thống

### 3.1.1. Mô hình tổng thể hệ thống

Hệ thống thu phí giao thông tự động không dừng (ETC - Electronic toll collector) là dịch vụ thu phí tự động tại các trạm thu phí thông qua thẻ định danh dán trực tiếp trên xe. Dịch vụ thu phí tự động áp dụng công nghệ nhận dạng tần số vô tuyến (RFID) với việc sử dụng sóng radio để nhận diện tự động các phương tiện xe cơ giới. Các thiết bị truyền thông như bộ dò vòng từ phát hiện xe cũng góp phần tích cực cho việc duy trì hệ thống**.** Thu phí không dừng trên các tuyến đường góp phần tiết kiệm thời gian, tiết kiệm chi phí, giảm thiểu ùn tắc giao thông và đặc biệt có ảnh hưởng tốt đến việc giảm thiểu ô nhiễm môi trường.

****

##### Hình 3.1: Mô hình tổng quan hệ thống

Hình 3.1 mô tả tổng thể hệ thống thu phí dịch vụ sử dụng đường bộ tự động không dừng ETC được xây dựng bao gồm 3 thành phần chính như sau:

* Hệ thống ETC tại trạm thu phí (ETC Frontend): Gồm các thiết bị lắp đặt ngoài trạm thu phí bao gồm máy ghi hình, đầu đọc thẻ, ăng ten, vòng từ, cảm biến laser v.v… nhằm phát hiện, định danh phương tiện để xử lý giao dịch khi có phương tiện đi qua, đồng thời thực hiện một số nghiệp vụ khác theo yêu cầu của các cơ quan chức năng.
* Hệ thống trung tâm dữ liệu (ETC Backend): Hệ thống Trung tâm dữ liệu đặt tại trung tâm dữ liệu tiêu chuẩn quốc tế để thực hiện kết nối và xử lý thông tin với các trạm thu phí, gồm các thiết bị máy chủ xử lý, hạ tầng công nghệ thông tin, hệ thống phần mềm… thực hiện các chức năng quản lý khách hàng và thẻ eTag, quản lý và xử lý giao dịch, tính cước, thanh toán, đối soát, báo cáo, chăm sóc khách hàng, kết nối đến các hệ thống của Bộ, ban ngành, BOO1 (VETC) v.v…
* Hệ thống trung tâm điều hành (ETC Operation): Hệ thống trung tâm điều hành được xây dựng hỗ trợ công tác quản lý thông tin, theo dõi, giám sát toàn bộ thiết bị, hạ tầng và các sự kiện diễn ra trong hệ thống thu phí tự động không dừng.

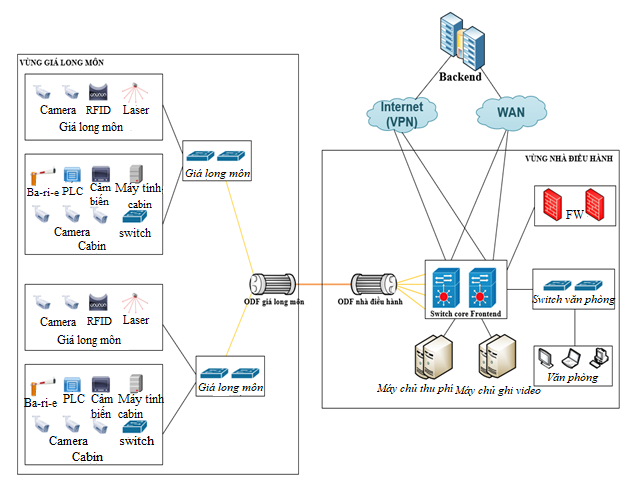
### 3.1.2. Hạ tầng mạng vật lý

Hạ tầng mạng vật lý bao gồm 2 hạ tầng chính là trạm thu phí (frontend) và trung tâm dữ liệu (backend).

**Hạ tầng mạng lớp trạm thu phí (frontend)**

Hệ thống được chia làm 2 khu vực: Khu vực ngoài trạm thu phí gồm các thiết bị phục vụ việc giám sát phương tiện giao thông đi qua và khu vực nhà điều hành gồm các máy chủ ghi hình, máy chủ thu phí và máy tính cá nhân phục vụ quản lý trực tiếp ngoài trạm thu phí (Hình 3.2 mô tả tổng quan trạm thu phí).

* Khu vực ngoài trạm thu phí gồm các thiết bị phục vụ phát hiện, định dang phương tiện để xử lý giao dịch cũng như các thiết bị theo dõi, hậu kiểm khác. Các thiết bị trên giá long môn được nối với switch trên giá long môn, các thiết bị trong cabin thu phí được nối với switch cabin và switch cabin được đấu nối lên switch giá long môn.
* Khu vực trong nhà điều hành trạm thu phí gồm các thiết bị máy chủ, switch core, firewall v.v… nhằm mục đích giao tiếp với backend để xử lý giao dịch, đồng bộ video quan sát trạm thu phí và các nghiệp vụ giám sát khác.
* Giữa khu vực ngoài trạm thu phí và nhà điều hành được kết nối bằng cáp quang thông qua ODF để giảm suy hao tín hiệu với khoảng cách lớn, cụ thể là 1 cặp switch gom tại giá long môn đấu nối đường liên kết lên switch core trong nhà điều hành.



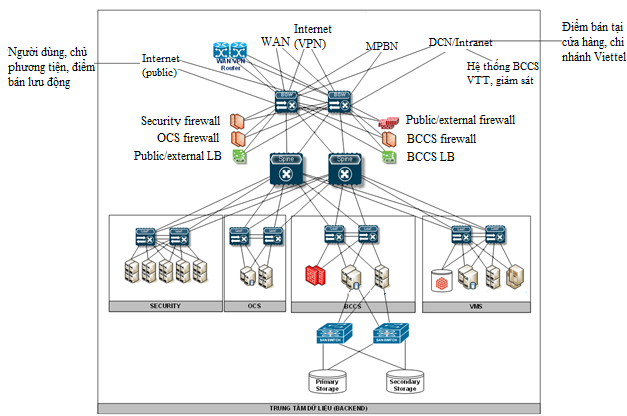
##### Hình 3.2: Mô hình tổng quan trạm thu phí

**Hạ tầng mạng lớp trung tâm dữ liệu (backend)**

Hạ tầng được triển theo mô hình spine-leaf với kiến trúc mạng lớp 3 sử dụng các công nghệ và giao thức mới: VXLAN, BGP EVPN, tách biệt giữa lớp điều khiển và lớp dịch vụ. Hệ thống được chia làm 4 phân hệ: BCCS, OCS, VMS và ATTT. Mỗi phân hệ được kết nối đến các cặp switch leaf đấu dự phòng cho nhau, trong đó:

* Phân hệ BCCS sẽ có 4 link vật lý đấu về 4 switch leaf.
* Phân hệ OCS, ATTT, VMS (máy chủ) có 2 link vật lý đấu về 2 switch leaf.
* Phân hệ thiết bị lưu trữ NAS của VMS có 4 link vật lý, trong đó 2 link đấu về 2 switch leaf kết nối vào mạng lưới và 2 link đấu về 2 switch nội bộ thiết lập mạng NAS của VMS.

Các switch leaf của hệ thống được kết nối đến switch spine. Các switch border gateway kết nối với switch spine và kết nối đến các vùng mạng tương ứng mà hệ thống cần giao tiếp như mạng tổng đài để gửi tin nhắn tới khách hàng hay mạng giám sát theo dõi các thiết bị trong hệ thống trung tâm dữ liệu,… được mô tả trong hình 3.3



##### Hình 3.3: Mô hình Backend

## 3.2. Thiết kế chi tiết hệ thống

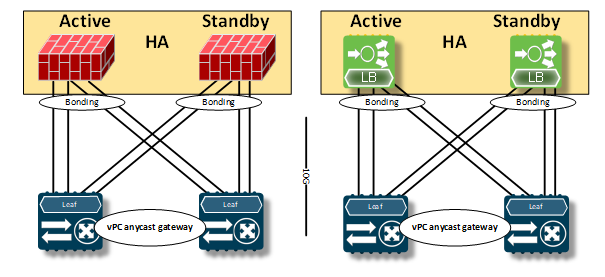
### 3.2.1. Thiết kế chi tiết phân hệ backend

Phân hệ Backend thiết kế tuân theo mô hình Spine-Leaf

* Các thiết bị Leaf Boder Gateway (BGW) đóng vai trò cung cấp kết nối đến trục, firewall, loadbalancer và kết nối ngoại mạng. Các thiết bị Leaf Boder Gateway kết nối đến spine, router backbone.
* Các thiết bị spine làm nhiệm vụ trung chuyển lưu lượng giữa các thiết bị leaf, các thiết bị spine không đấu nối bắt cặp với nhau.
* Các thiết thiết bị Leaf dịch vụ đóng vai trò cung cấp kết nối đến máy chủ, bộ lưu trữ.
* Các thiết bị Leaf chạy bắt cặp với nhau và cho phép liên kết vật lý trên 2 thiết bị gom lại thành 1 cổng logic (channel), giúp tăng tính dự phòng dịch vụ.
* Các thiết bị máy chủ kết nối tối thiểu 2 đường lên cặp thiết bị leaf để đảm bảo tính dự phòng.

Loadbalancer, Firewall kết nối sử dụng cơ chế sẵn sàng cao (HA)

* Các thiết bị firewall, loadbalancer (LB) được đấu nối theo từng cặp, kết nối bắt cặp qua 2 cổng điện, sử dụng cơ chế HA, một thiết bị đóng vai trò chính, một thiết bị đóng vai trò dự phòng (Hình 3.4).
* Thiết bị firewall riêng và firewall VAS được kết nối tới thiết bị leaf dịch vụ, firewall công cộng kết nối tới thiết bị switch border gateway.
* Thiết bị loadbalancer (LB) riêng được kết nối tới thiết bị Leaf dịch vụ và loadbalancer công cộng kết nối tới thiết bị switch border gateway.



##### Hình 3.4: Mô hình chạy HA

### 3.2.2. Thiết kế chi tiết phân hệ Frontend

Các thiết bị mạng được đặt trong nhà điều hành, cặp switch lõi gom kết nối tập trung toàn bộ thiết bị (bao gồm cả phân hệ giá long môn và vùng nhà điều hành). Thiết bị switch lõi cung cấp đấu nối liên kết với Firewall, các kết nối Internet, WAN. Đảm bảo dự phòng 1+1 với các đường internet, WAN, mỗi kết nối được cắm trên 1 thiết bị switch lõi.

**Thiết kế Switch core**

Thiết kế kênh cổng ảo (vPC)

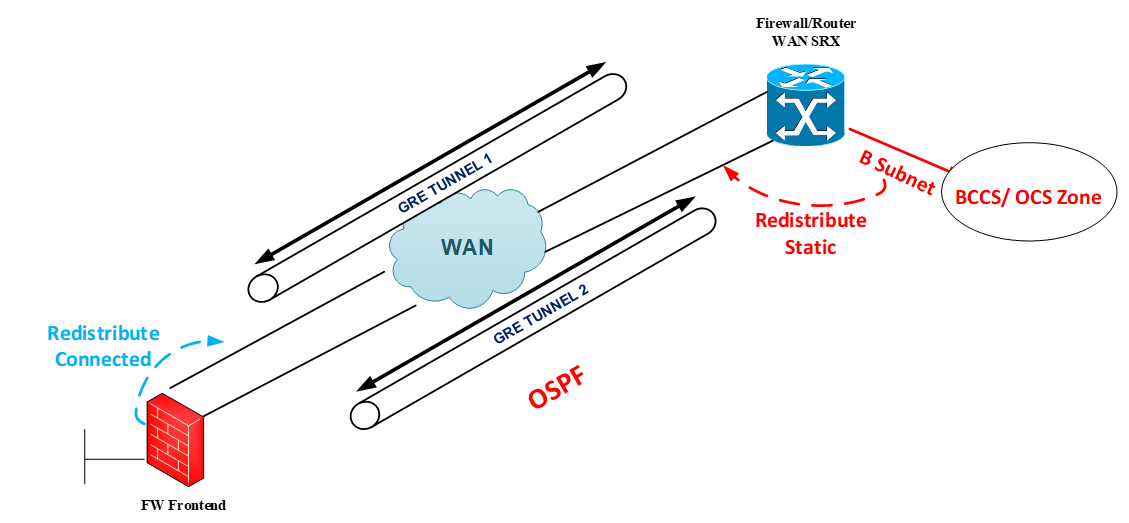
Công nghệ kênh giúp đơn giản hóa quá trình cấu hình và vận hành hạ tầng mạng đảm bảo hệ thống sẽ không bị vòng lặp ở sơ đồ lớp 2 bằng cách gom 2 switch thành 1 switch ảo duy nhất. Switch ảo này giúp giảm thiểu số lượng định tuyến lớp 3 thông qua kết nối lớp 2 giữa các switch lớp truy cập và switch lớp phân phối. Một vPC là một cặp thiết bị switch hoạt động như 1 thiết bị mạng duy nhất với khả năng dự phòng (redundancy) và khả năng cân bằng tải (load balancing). Một switch sẽ đóng vai trò chính và switch còn lại sẽ đóng vai trò dự phòng. Các thành viên vPC được kết nối và giao tiếp với nhau bằng các cổng vPC.

Thiết kế giao thức STP (Spanning-tree)

Sử dụng giao thức STP để phòng chống vòng lặp lớp 2, đảm bảo toàn bộ các VLAN có Core Switch đóng vai trò chính trong vận chuyển dữ liệu. Trong môi trường lớp 2 không sử dụng giao thức định tuyến và cũng không cho phép các tuyến đường dự phòng. Giao thức STP cung cấp liên kết dự phòng để mạng chuyển mạch lớp 2 có thể khôi phục khi lỗi xảy ra tại tuyến đường đang hoạt động mà không cần có sự can thiệp nào khác.

### 3.2.3. Thiết kế định tuyến

Thực hiện chia thực thể chuyển tiếp ảo (VRF) để tách biệt về mặt định tuyến giữa phân vùng WAN (lưu lượng tính cước) và phân vùng Internet (lưu lượng đồng bộ video), chạy OSPF trên toàn bộ các chân sử dụng đóng gói định tuyến chung (GRE) qua IPsec cũng như chân GRE thông thường để tăng băng thông và tối ưu dự phòng. Với thiết kế này, toàn bộ lưu lượng sẽ được cân bằng tải giữa các đường IPsec/ WAN. Thiết bị Router Backend sẽ thiết lập chính sách an ninh để giới hạn quản lý lưu lượng từ Frontend gửi lên.



##### Hình 3.5: Kết nối OSPF với GRE phục vụ lưu lượng tính cước

* Thiết kế kết nối đường WAN (mô tả trong hình 3.5)
* Các trạm sẽ có 2 kết nối WAN về điểm trung tâm, đảm bảo đường dự phòng cho nhau. Đường WAN sẽ sử dụng truyền tải dữ liệu từ các trạm về trung tâm OCS như: trả phí lưu thông, kiểm tra tài khoản thẻ, kiểm tra thông tin phương tiện,…
* Sử dụng giao thức định tuyến OSPF để cân bằng lưu lượng qua 2 đường WAN và chạy đường hầm GRE để thiết lập ngang hàng giữa firewall Frontend và Router WAN SRX.
* Cơ chế hoạt động: Firewall Frontend thực hiện phân phối dải mạng đấu nối tới máy chủ FrontEnd tại nhà điều hành các trạm vào miền OSPF qua đường WAN. Firewall/Router WAN SRX tạo các VRF đường WAN và đường về BGW trỏ định tuyến tĩnh xuống máy chủ OCS có next-hop là mạng con tại OCS Firewall và thực hiện phân phối định tuyến tĩnh Static vào miền OSPF. Trung tâm OCS sẽ có 1 thiết bị OCS Firewall, dữ liệu đi vào và đi ra OCS đều qua OCS Firewall. BGW tạo 1 VRF OCS và trỏ tới mạng con tại OCS Firewall, đồng thời OCS Firewall trỏ định tuyến tĩnh về máy chủ OCS có nexthop là mạng con tại BGW.



##### Hình 3.6: Kết nối OSPF với GRE qua IPsec phục vụ lưu lượng đồng bộ video

* Thiết kế kết nối đường internet (mô tả trong hình 3.6)
* GRE là giao thức không bảo mật, việc kết hợp với IPSec sẽ giúp tăng cường tính năng bảo mật cho kênh truyền. Khi IPSec kết hợp với GRE sẽ cung cấp khả năng định tuyến động trên kênh truyền, do đó tạo ra khả năng mở rộng hệ thống mạng rất lớn.
* GRE over IPSec là sự kết hợp giữa GRE và IPSec. Lúc này các gói tin GRE sẽ được truyền thông qua kênh truyền bảo mật do IPSec thiết lập. Điều này được thực hiện thông qua việc IPSec sẽ đóng gói gói tin GRE bởi các thông tin bảo mật của mình.
* Mỗi trạm sử dụng 2 đường internet, sử dụng truyền và đồng bộ video giám sát của các phương tiện khi đi qua trạm. Sử dụng giao thức định tuyến OSPF để cân bằng lưu lượng qua 2 đường internet.
* Cơ chế hoạt động: FW Frontend thực hiện phân phối dải mạng đấu nối tới máy chủ ghi hình tại nhà điều hành các trạm vào miền OSPF qua đường internet. Router SRX tạo các VRF đường internet và đường về BGW định tuyến tĩnh xuống máy chủ VMS có nexthop là mạng con tại BGW và thực hiện phân phối định tuyến tĩnh vào miền OSPF. BGW tạo 1 VRF VMS và trỏ next-hop tới mạng con tại Router SRX.

## 3.3. Mô hình luồng dữ liệu kết nối trong hệ thống

Hệ thống ETC tại trung tâm dữ liệu (backend) được chia làm 4 phân hệ: BCCS, ATTT, OCS, VMS. Trong đó:

* Phân hệ BCCS bao gồm máy chủ ứng dụng ảo hóa và máy chủ cơ sở dữ liệu có kết nối với thiết bị lưu trữ SAN. Phân hệ BCCS giao tiếp cả trong mạng nội bộ (giao tiếp với các phân hệ khác, kết nối với chi nhánh, điểm bán của Viettel và các văn phòng của VDTC), WAN (kết nối với trạm thu phí, VETC, Tổng cục đường bộ v.v…) và mạng internet (phục vụ cho khách hàng truy cập tra cứu hoặc điểm bán lưu động). Dữ liệu trong DB phân hệ BCCS đặc biệt quan trọng và nhạy cảm, vì là nơi lưu trữ toàn bộ thông tin phương tiện giao thông bao gồm cả số điện thoại, tên, ảnh của chủ sở hữu xe cũng như lịch sử đấu nối, kích hoạt, hủy v.v… Phân hệ ứng dụng của BCCS triển khai giải pháp vCloud, nên cần có đường giao tiếp phục vụ giám sát các máy chủ Cloud.
* Phân hệ OCS giao tiếp với các kênh thanh toán (ViettelPay, v.v…) để trừ cước cho chủ phương tiện, và giao tiếp với các tổng đài thông qua kết nối với vùng mạng MPBN. Phân hệ này không cần kết nối ra internet.
* Phân hệ ATTT bao gồm các máy chủ thực hiện giám sát và phát hiện hành vi bất thường của hệ thống.

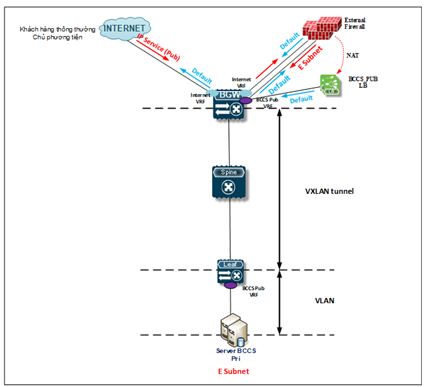
Từ trạm thu phí (frontend) về hạ tầng trung tâm dữ liệu (backend) có 2 luồng nghiệp vụ chính như sau:

* Luồng xử lý trừ cước khi xe đi qua trạm: Các thiết bị cảm biến laser, đầu đọc thẻ, camera nhận dạng biển số v.v… thực hiện phát hiện và chụp ảnh phương tiện khi đi qua làn thu phí.
* Luồng này cần đảm bảo yêu cầu về đường truyền ổn định, độ trễ thấp: Định tuyến luồng xử lý cước đi qua giao diện WAN để kết nối với trung tâm dữ liệu.
* Luồng đồng bộ video giám sát: Máy chủ ghi video thực hiện ghi lại luồng video từ các camera quan sát làn, camera cabin, camera toàn cảnh v.v… để đồng bộ về trung tâm dữ liệu phục vụ đối soát, hậu kiểm.
* Luồng này tuy có yêu cầu băng thông lớn nhưng không liên tục và không đòi hỏi về độ ổn định, độ trễ: Định tuyến luồng đồng bộ video đi qua đường truyền FTTH, có thiết lập VPN với trung tâm dữ liệu để giảm chi phí đường truyền nhưng vẫn đảm bảo về an toàn thông tin.

### 3.3.1. Lưu lượng truy cập xem thông tin tính cước từ internet

Khách hàng truy cập hệ thống thông qua trình duyệt, ứng dụng di động để xem thông tin khách hàng, thông tin thẻ, cước, v.v… Chi tiết cấu trúc mạng cho dịch vụ này được tách ra làm 3 luồng riêng biệt như sau:

**Lưu lượng từ internet vào phân vùng BCCS (hình 3.7)**



##### Hình 3.7: Lưu lượng từ Internet đến phân hệ BCCS

Lưu lượng yêu cầu từ internet đến BCCS:

* Người dùng internet truy cập vào địa chỉ IP công cộng (được đặt trên Firewall External).
* Firewall External thực hiện NAT vào địa chỉ riêng đặt trên trên LB công cộng.
* LB Public được thiết kế theo mô hình one-arm thực hiện yêu cầu cân bằng tải đi vào các máy chủ backend.
* Lưu lượng đến Boder Leaf được đóng gói VXLAN gửi đến Leaf BCCS, Leaf BCCS tách tiêu đề VXLAN sau đó chuyển tiếp theo L2 tới máy chủ BCCS.

Lưu lượng phản hồi BCCS về internet:

* Máy chủ BCCS gửi phản hồi địa chỉ trên LB – Gói lúc này vẫn là gói tin truyền thống.
* Leaf BCCS nhận được gói tin từ máy chủ, chèn thêm tiêu đề VXLAN và gửi tới Border Leaf.
* Border Leaf tách tiêu đề VXLAN, gửi gói tin trả lại LB công cộng.
* LB công cộng trả gói tin ra cho người dùng internet tới Boder Leaf, sau đó Boder Leaf chuyển tiếp gói tin tới Firewall và trả lại Internet..

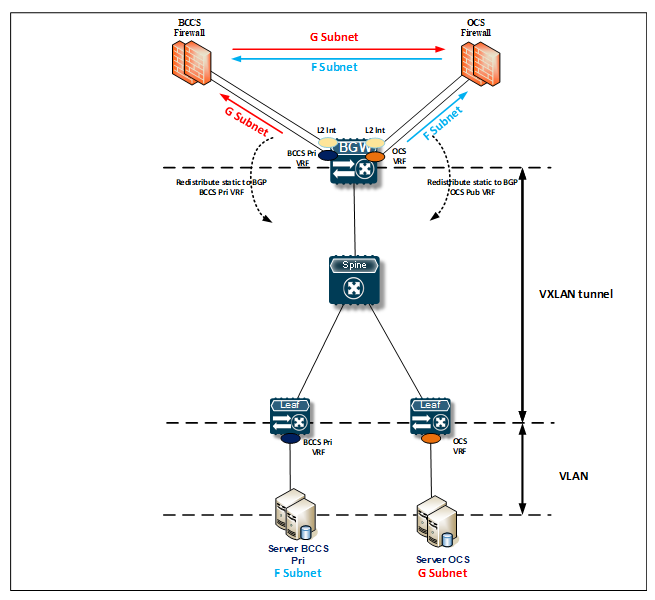
**Lưu lượng giữa phân vùng BCCS và OCS (hình 3.8)**

Lưu lượng yêu cầu từ BCCS đến OCS:

* Máy chủ BCCS khởi tạo dữ liệu tới máy chủ OCS (cổng mặc định đặt trên Leaf BCCS).
* Leaf BCCS học được tuyến đường mặc định từ Boder Leaf qua tuyến đường EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF BCCS được gateway tới Firewall BCCS để ra các vùng mạng người, đóng gói truyền thống để gửi tới Firewall BCCS.
* Firewall BCCS có next-hop sang phân vùng OCS qua Firewall OCS, chuyển tiếp gói tin tới Firewall OCS qua chân đấu nối giữa 2 Firewall.
* Firewall OCS có next-hop toàn bộ mạng con phân vùng OCS là Border Leaf – VRF OCS, chuyển tiếp gói tin đến VRF OCS trên Border Leaf.
* Boder Leaf – VRF OCS có tuyến đường EVPN loại 5 tới địa chỉ server OCS, đóng gói VXLAN gửi tới Leaf OCS.
* Leaf OCS tách tiêu đề VXLAN, chuyển tiếp gói tin tới server OCS.

Lưu lượng phản hồi từ OCS về BCCS:

* Máy chủ OCS khởi tạo lưu lượng tới máy chủ BCCS (cổng mặc định đặt trên Leaf OCS).
* Leaf OCS học được Default Route từ Boder Leaf qua tuyến đường EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF OCS được gateway tới Firewall OCS để ra các vùng mạng người, đóng gói truyền thống để gửi tới Firewall OCS.
* Firewall OCS có next-hop sang phân vùng BCCS qua Firewall BCCS, chuyển tiếp gói tin tới Firewall BCCS qua chân đấu nối giữa 2 Firewall
* Firewall BCCS có next-hop toàn bộ mạng con phân vùng BCCS là Border Leaf – VRF BCCS, chuyển tiếp gói tin đến VRF BCCS trên Border Leaf.
* Boder Leaf – VRF BCCS có tuyến đường EVPN loại 5 tới địa chỉ server BCCS, đóng gói VXLAN gửi tới Leaf BCCS.
* Leaf BCCS tách tiêu đề VXLAN, chuyển tiếp gói tin tới máy chủ BCCS.



##### Hình 3.8: Lưu lượng từ phân hệ BCCS đến phân hệ OCS

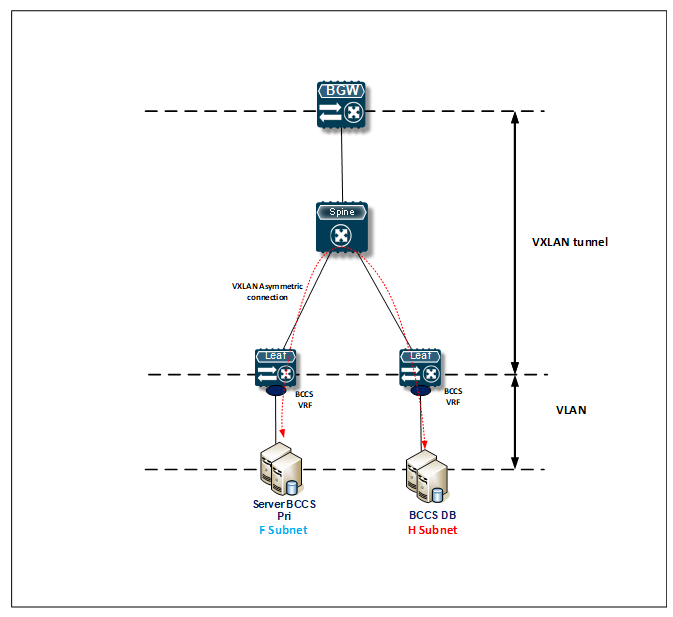
**Lưu lượng giữa BCCS và BCCS DB (hình 3.9)**

Lưu lượng yêu cầu từ máy chủ BCCS đến BCCS DB:

* Máy chủ BCCS khởi tạo kết nối đến BCCS DB (cổng Leaf BCCS).
* Do cấu hình cổng phân phối, BCCS Leaf 1 lúc này nhìn mạng con H (BCCS DB đấu nối trực tiếp) – sẽ thực hiện đóng gói VXLAN gửi tới BCCS Leaf 2 chứa BCCS DB (thông tin này học qua tuyến đường EVPN loại 2).
* BCCS Leaf 2 thực hiện tách tiêu đề VXLAN, sau đó chuyển tiếp gói tin đến BCCS DB.

Lưu lượng phản hồi từ BCCS DB về máy chủ BCCS:

* Máy chủ BCCS DB khởi tạo kết nối đến máy chủ BCCS (cổng Leaf BCCS).
* Do cấu hình cổng phân phối, BCCS Leaf 2 lúc này nhìn mạng con F (máy chủ BCCS đấu nối trực tiếp) – sẽ thực hiện đóng gói VXLAN gửi tới BCCS Leaf 1 chứa BCCS DB (thông tin này học qua tuyến đường EVPN loại 2).
* BCCS Leaf 1 thực hiện tách tiêu đề VXLAN, sau đó chuyển tiếp gói tin đến máy chủ BCCS.



##### Hình 3.9: Lưu lượng từ BCCS Server đến BCCS DB

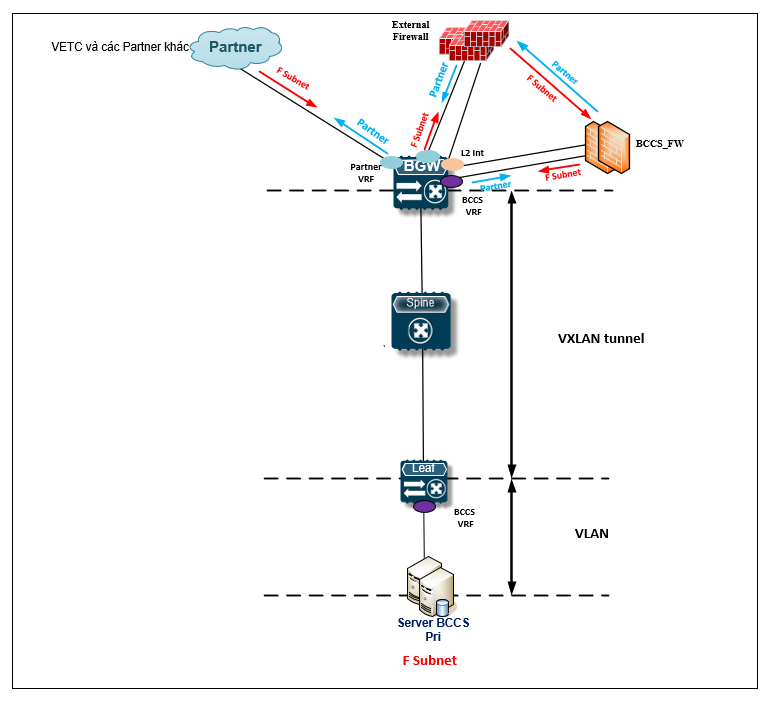
### 3.3.2. Lưu lượng truy cập giữa Backend và đối tác

Các đơn vị bên đối tác có nhu cầu truy cập vào hệ thống backend (OCS, BCCS,VMS) để phục vụ chức năng hậu kiểm. Luồng giao tiếp giữa các vùng mạng của nghiệp vụ này là như sau:

**Lưu lượng từ đối tác tới phân vùng BCCS (hình 3.10)**

Lưu lượng yêu cầu từ đối tác đến BCCS:

* Người dùng ETC truy cập vào địa chỉ IP tiêng tư tại phân vùng BCCS, trên VRF Partner thực hiện trỏ next-hop đến phân vùng BCCS qua Firewall External.
* Firewall External thực hiện lọc lưu lượng qua chính sách sau đó định tuyến lưu lượng qua Firewall Internal (BCCS).
* Firewall BCCS thực hiện chuyển tiếp lưu lượng vào LB Private (Nexthop Border Leaf – VRF BCCS).
* LB riêng tư được thiết kế theo mô hình one-arm thực hiện cân bằng tải yêu cầu vào các máy chủ vùng Backend.
* Lưu lượng đến Boder Leaf được đóng gói VXLAN gửi đến Leaf BCCS.
* Leaf BCCS tách tiêu đề VXLAN sau đó chuyển tiếp theo L2 tới máy chủ BCCS.



##### Hình 3.10: Lưu lượng từ đối tác đến phân hệ BCCS

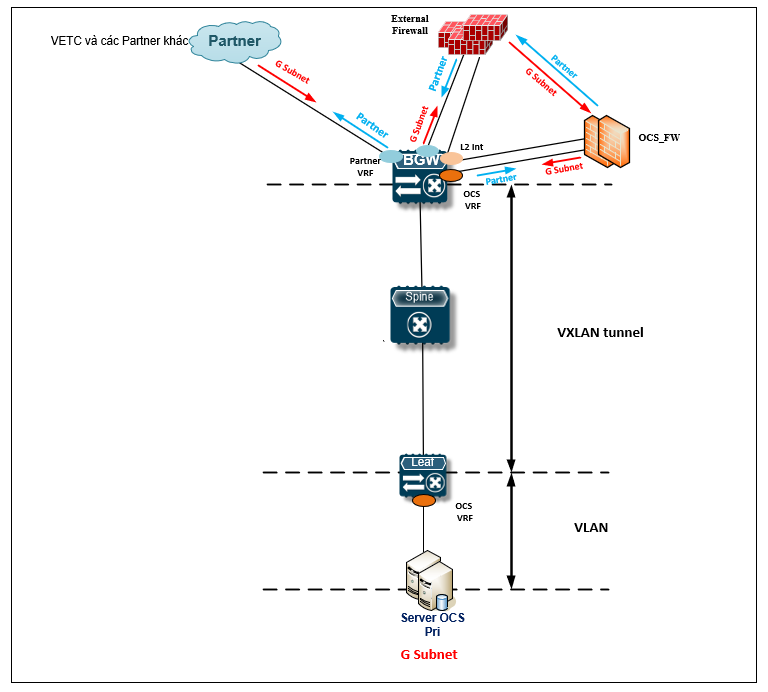
Lưu lượng phản hồi từ BCCS về đối tác:

* Máy chủ BCCS gửi bản tin phản hồi lại địa chỉ SNAT trên LB – Packet lúc này vẫn là gói tin truyền thống.
* Leaf BCCS nhận được gói tin từ máy chủ, chèn thêm tiêu đề VXLAN và gửi tới Border Leaf.
* Border Leaf tách tiêu đề VXLAN, gửi gói tin trả lại LB riêng.
* LB riêng trả gói tin ra cho người dùng ETC tới Firewall công cộng, sau đó Firewall công cộng gửi tới Border Leaf để chuyển tiếp lưu lượng tới vùng partner.

**Lưu lượng từ đối tác tới phân vùng OCS (hình 3.11)**

Lưu lượng yêu cầu từ đối tác đến OCS:

* Người dùng đối tác truy cập vào địa chỉ IP riêng thuộc mạng con OCS, luồng này sẽ đi qua Firewall công cộng kiểm duyệt sau đó gửi lưu lượng tới Firewall OCS kiểm duyệt.
* Firewall OCS thực hiện chuyển tiếp lưu lượng vào máy chủ OCS (Nexthop Border Leaf – VRF OCS).
* Lưu lượng đến Boder Leaf được đóng gói VXLAN gửi đến Leaf OCS.
* Leaf OCS tách tiêu đề VXLAN sau đó chuyển tiếp theo L2 tới Server OCS.



##### Hình 3.11: Lưu lượng từ đối tác đến phân hệ VMS

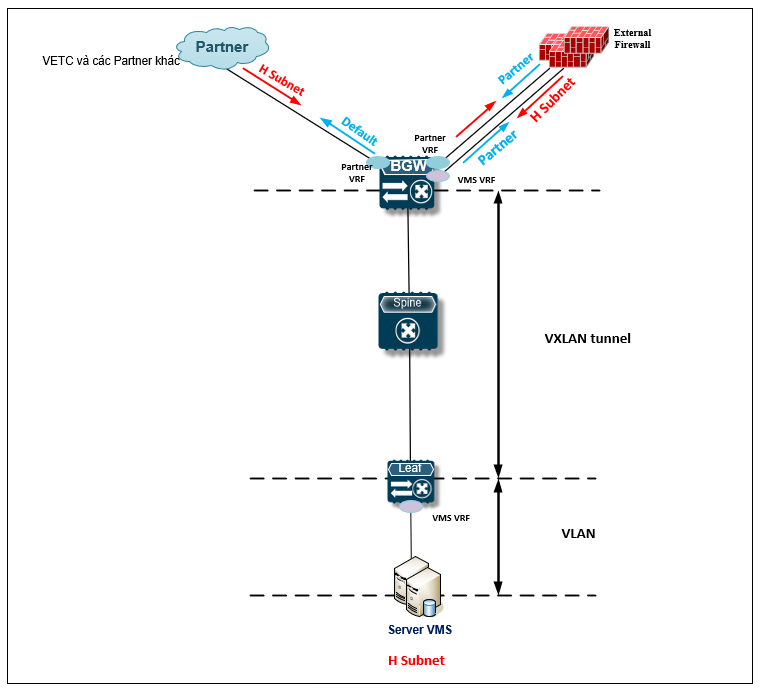
Lưu lượng phản hồi từ OCS về đối tác:

* Máy chủ OCS gửi phản hồi lại người dùng phía đối tác (cổng Leaf OCS).
* Leaf OCS nhận được gói tin từ máy chủ, chèn thêm tiêu đề VXLAN và gửi tới Border Leaf.
* Border Leaf tách tiêu đề VXLAN, gửi gói tin tới Firewall OCS.
* Firewall OCS lọc lưu lượng rồi gửi trả lại Firewall công cộng, sau đó Firewall công cộng gửi lưu lượng về đối tác qua VRF Partner.

**Lưu lượng từ đối tác tới phân vùng VMS (hình 3.11)**

Lưu lượng yêu cầu từ đối tác đến VMS:

* Người dùng đối tác truy cập vào địa chỉ IP riêng tư thuộc mạng con VMS, luồng này sẽ đi qua Firewall công cộng kiểm duyệt sau đó gửi lưu lượng tới mạng con VMS qua VRF VMS.
* Lưu lượng đến Boder Leaf được đóng gói VXLAN gửi đến Leaf VMS.
* Leaf VMS tách tiêu đề VXLAN sau đó chuyển tiếp theo L2 tới máy chủ OCS.



##### Hình 3.11: Lưu lượng từ đối tác đến phân vùng VMS

Lưu lượng phản hồi từ VMS về đối tác:

* Máy chủ VMS gửi phản hồi lại người dùng đối tác (cổng Leaf VMS)
* Leaf VMS nhận được gói tin từ máy chủ, chèn thêm tiêu đề VXLAN và gửi tới Border Leaf.
* Border Leaf tách tiêu đề VXLAN, gửi gói tin tới Firewall công cộng qua VRF VMS.
* Firewall công cộng lọc lưu lượng rồi gửi trả lại về đối tác qua VRF Partner.

### 3.3.3. Lưu lượng truy cập Backend từ các điểm bán, cửa hàng đại lý

Các điểm bán, cửa hàng đại lý có thể cần xem lịch sử của các giao dịch. Chi tiết cho dịch vụ này được tách ra làm 3 luồng riêng biệt như sau:

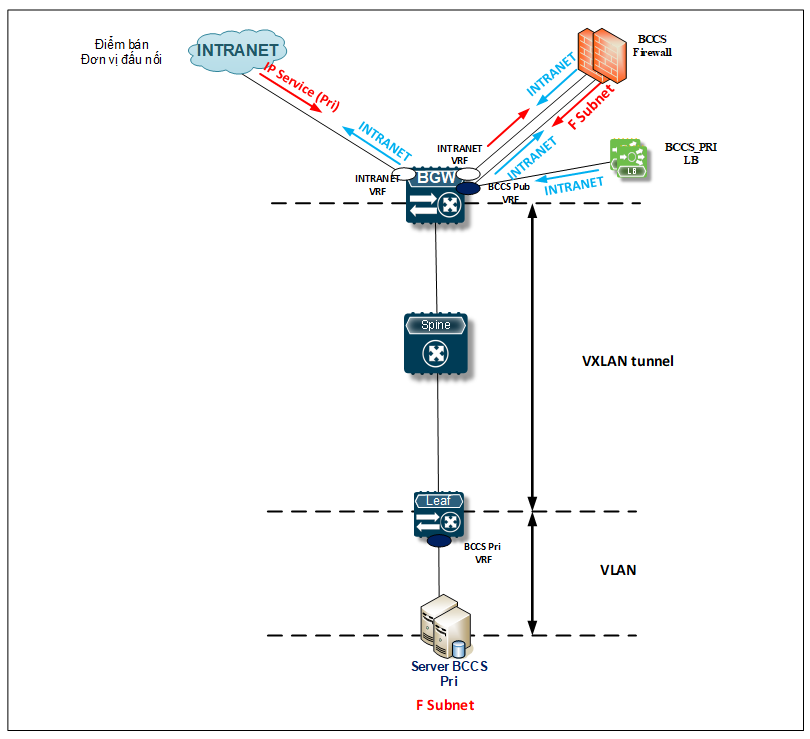
**Lưu lượng từ các điểm bán vào phân vùng BCCS (hình 3.12)**

Lưu lượng yêu cầu từ điểm bán đến BCCS:

* Người dùng điểm bán truy cập vào địa chỉ IP riêng tư (được đặt trên LB Private), luồng này sẽ đi qua Firewall BCCS kiểm duyệt.
* Firewall BCCS thực hiện chuyển tiếp lưu lượng vào LB riêng (Nexthop Border Leaf – VRF BCCS).
* LB riêng được thiết kế theo mô hình one-arm thực hiện cân bằng tải vào các máy chủ Backend.
* Lưu lượng đến Boder Leaf được đóng gói VXLAN gửi đến Leaf BCCS.
* Leaf BCCS tách tiêu đề VXLAN sau đó chuyển tiếp theo L2 tới Server BCCS.

Lưu lượng phản hồi từ BCCS về điểm bán:

* Máy chủ BCCS gửi phản hồi lại địa chỉ IP riêng trên LB – Gói tin lúc này vẫn là gói tin truyền thống.
* Leaf BCCS nhận được gói tin từ máy chủ, chèn thêm tiêu đề VXLAN và gửi tới Border Leaf.
* Border Leaf tách tiêu đề VXLAN, gửi gói tin trả lại LB riêng.
* LB riêng trả gói tin ra cho người dùng các điểm bán tới Boder Leaf, sau đó Boder Leaf chuyển tiếp gói tin tới Firewall và trả về các điểm bán.

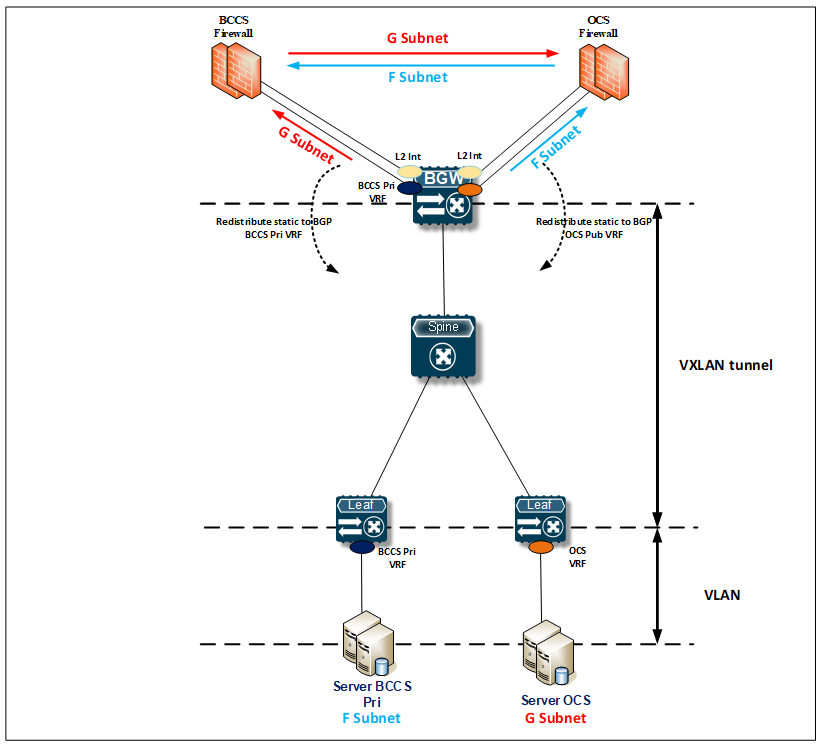


##### Hình 3.12:Lưu lượng từ Intranet vào phân hệ BCCS

**Lưu lượng giữa phân vùng BCCS và OCS (hình 3.13)**

Lưu lượng yêu cầu từ BCCS đến OCS:

* Máy chủ BCCS khởi tạo dữ liệu tới máy chủ OCS (cổng mặc định đặt trên Leaf BCCS).
* Leaf BCCS học được tuyến đường mặc định từ Boder Leaf qua tuyến đường EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF BCCS được gateway tới Firewall BCCS để ra các vùng mạng người, đóng gói truyền thống để gửi tới Firewall BCCS.
* Firewall BCCS có next-hop sang phân vùng OCS qua Firewall OCS, chuyển tiếp gói tin tới Firewall OCS qua chân đấu nối giữa 2 Firewall.
* Firewall OCS có next-hop toàn bộ mạng con phân vùng OCS là Border Leaf – VRF OCS, chuyển tiếp gói tin đến VRF OCS trên Border Leaf.
* Boder Leaf – VRF OCS có tuyến đường EVPN loại 5 tới địa chỉ server OCS, đóng gói VXLAN gửi tới Leaf OCS.
* Leaf OCS tách tiêu đề VXLAN, chuyển tiếp gói tin tới server OCS.

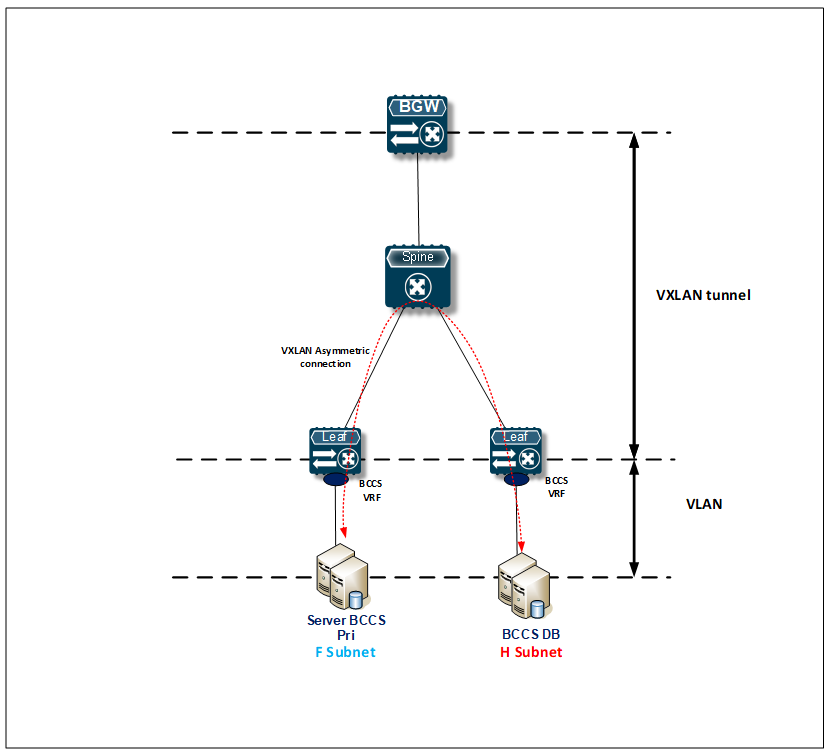


##### Hình 3.13: Lưu lượng giữa phân hệ BCCS và OCS

Lưu lượng phản hồi từ OCS về BCCS:

* Máy chủ OCS khởi tạo lưu lượng tới máy chủ BCCS (cổng mặc định đặt trên Leaf OCS).
* Leaf OCS học được tuyến đường mặc định từ Boder Leaf qua tuyến đường EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF OCS có cổng tới Firewall OCS để ra các vùng mạng người, đóng gói truyền thống để gửi tới Firewall OCS.
* Firewall OCS có next-hop sang phân vùng BCCS qua Firewall BCCS, chuyển tiếp gói tin tới Firewall BCCS qua chân đấu nối giữa 2 Firewall.
* Firewall BCCS có next-hop toàn bộ mạng con phân vùng BCCS là Border Leaf – VRF BCCS, chuyển tiếp gói tin đến VRF BCCS trên Border Leaf.
* Boder Leaf – VRF BCCS có tuyến đường EVPN loại 5 tới địa chỉ máy chủ BCCS, đóng gói VXLAN gửi tới Leaf BCCS.
* Leaf BCCS tách tiêu đề VXLAN, chuyển tiếp gói tin tới máy chủ BCCS.

**Lưu lượng giữa BCCS và BCCS DB (hình 3.14)**



##### Hình 3.14: Lưu lượng giữa BCCS Server và BCCS DB

Lưu lượng yêu cầu từ máy chủ BCCS đến BCCS DB:

* Máy chủ BCCS khởi tạo kết nối đến BCCS DB (cổng Leaf BCCS).
* Do cấu hình cổng phân phối, BCCS Leaf 1 lúc này nhìn mạng con H (BCCS DB đấu nối trực tiếp) – sẽ thực hiện đóng gói VXLAN gửi tới BCCS Leaf 2 chứa BCCS DB (thông tin này học qua tuyến đường EVPN loại 2).
* BCCS Leaf 2 thực hiện tách tiêu đề VXLAN, sau đó chuyển tiếp gói tin đến BCCS DB.

Lưu lượng phản hồi từ BCCS DB về máy chủ BCCS:

* Máy chủ BCCS DB khởi tạo kết nối đến máy chủ BCCS (cổng BCCS Leaf).
* Do cấu hình cổng phân phối, BCCS Leaf 2 lúc này nhìn mạng con F (BCCS Server đấu nối trực tiếp) – sẽ thực hiện đóng gói VXLAN gửi tới BCCS Leaf 1 chứa BCCS DB (thông tin này học qua tuyến đường EVPN loại 2).
* BCCS Leaf 1 thực hiện tách tiêu đề VXLAN, sau đó chuyển tiếp gói tin đến máy chủ BCCS.

### 3.3.4. Lưu lượng truy cập xem video từ trung tâm giám sát

Các đơn vị trung tâm giám sát, đối soát có thể cần theo dõi camera của các trạm hoặc xem lại video, lịch sử của các giao dịch. Thông qua phân hệ VMS, chi tiết dịch vụ này được tách ra làm 3 luồng riêng biệt như sau:

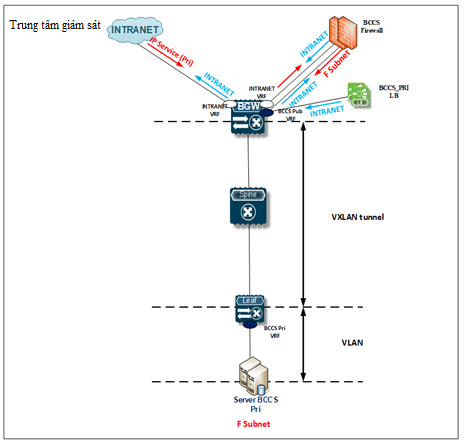
**Lưu lượng từ trung tâm giám sát vào phân vùng BCCS (hình 3.15)**

Lưu lượng yêu cầu từ trung tâm đến BCCS:

* Người dùng trung tâm truy cập vào địa chỉ IP riêng tư (được đặt trên LB Private), luồng này sẽ đi qua Firewall BCCS kiểm duyệt.
* Firewall BCCS thực hiện chuyển tiếp lưu lượng vào LB riêng (Nexthop Border Leaf – VRF BCCS).
* LB riêng được thiết kế theo mô hình one-arm thực hiện cân bằng tải vào các máy chủ Backend.
* Traffic đến Boder Leaf được đóng gói VXLAN gửi đến Leaf BCCS.
* Leaf BCCS tách tiêu đề VXLAN sau đó chuyển tiếp theo L2 tới Server BCCS.

Lưu lượng phản hồi từ BCCS về trung tâm:

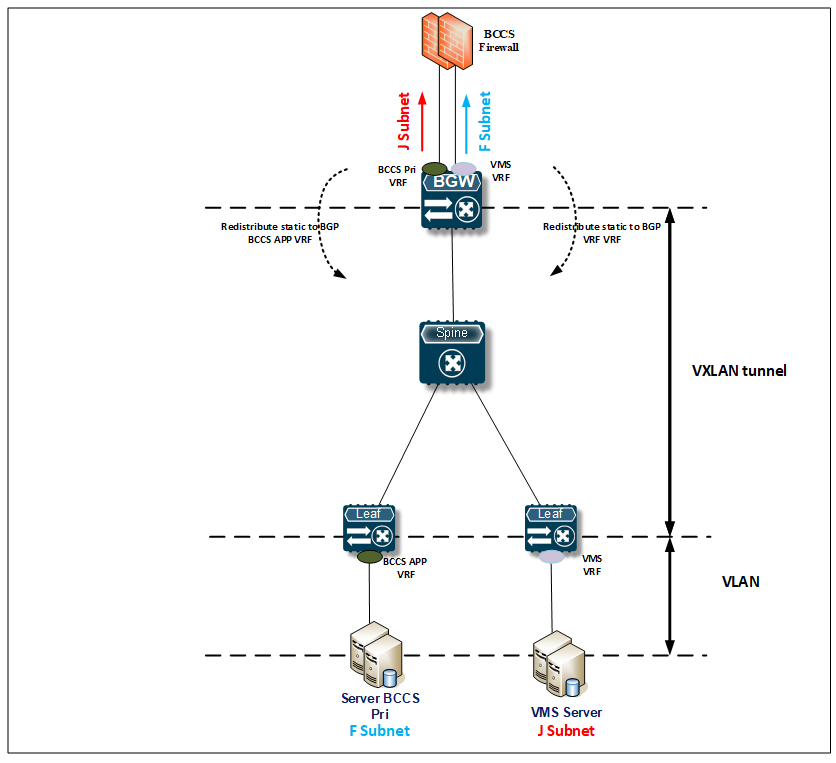
* Máy chủ BCCS gửi phản hồi lại địa chỉ IP riêng trên LB – Gói tin lúc này vẫn là gói tin truyền thống.
* Leaf BCCS nhận được gói tin từ máy chủ, chèn thêm tiêu đề VXLAN và gửi tới Border Leaf.
* Border Leaf tách tiêu đề VXLAN, gửi gói tin trả lại LB riêng.
* LB riêng trả gói tin ra cho người dùng các điểm bán tới Boder Leaf, sau đó Boder Leaf chuyển tiếp gói tin tới Firewall và trả về trung tâm.



##### Hình 3.15: Lưu lượng từ Intranet vào phân hệ BCCS

**Lưu lượng giữa phân vùng BCCS và VMS (hình 3.16)**

Lưu lượng yêu cầu từ BCCS đến VMS:

* Máy chủ BCCS khởi tạo lưu lượng máy chủ VMS (cổng mặc định đặt trên Leaf BCCS).
* Leaf BCCS học được tuyến mặc định từ Boder Leaf qua tuyến đường EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF BCCS có cổng mặc định tới Firewall BCCS để ra các vùng mạng người, đóng gói truyền thống để gửi tới Firewall BCCS.
* Firewall BCCS có next-hop sang phân vùng VMS qua Border Leaf – VRF VMS , chuyển tiếp gói tin tới Border Leaf – VMS qua chân đấu nối giữa Firewall và Border Leaf.
* Boder Leaf – VRF OCS có tuyến đường EVPN loại 5 tới địa chỉ máy chủ VMS, đóng gói VXLAN gửi tới Leaf VMS.
* Leaf VMS tách tiêu đề VXLAN, chuyển tiếp gói tin tới máy chủ VMS.

##### Hình 3.16: Flow traffic giữa phân hệ BCCS và VMS

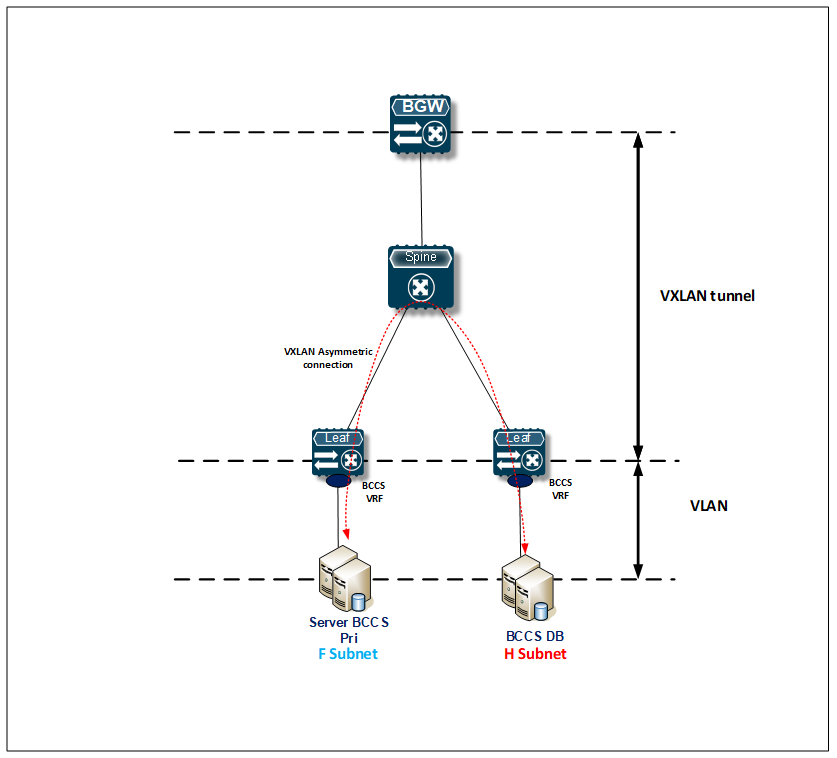
Lưu lượng phản hồi từ VMS về BCCS:

* Máy chủ VMS khởi tạo lưu lượng tới máy chủ BCCS (cổng mặc định đặt trên Leaf VMS).
* Leaf VMS học được tuyến đường mặc định từ Boder Leaf qua EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF VMS có cổng mặc định tới Firewall BCCS để ra các vùng mạng BCCS, đóng gói truyền thống để gửi tới Firewall BCCS.
* Firewall BCCS có next-hop sang phân vùng BCCS qua Border Leaf - VRF BCCS, chuyển tiếp gói tin tới Border Leaf- VRF BCCS qua chân đấu nối.
* Boder Leaf – VRF BCCS có tuyến đường EVPN loại 5 tới địa chỉ server BCCS, đóng gói VXLAN gửi tới Leaf BCCS.
* Leaf BCCS tách tiêu đề VXLAN, chuyển tiếp tới máy chủ BCCS.

**Lưu lượng giữa BCCS và BCCS DB (hình 3.17)**

Lưu lượng yêu cầu từ máy chủ BCCS đến BCCS DB

* Máy chủ BCCS khởi tạo kết nối đến BCCS DB (cổng Leaf BCCS).
* Do cấu hình cổng phân phối, BCCS Leaf 1 lúc này nhìn mạng con H (BCCS DB đấu nối trực tiếp) – sẽ thực hiện đóng gói VXLAN gửi tới BCCS Leaf 2 chứa BCCS DB (thông tin này học qua tuyến đường EVPN loại 2).
* BCCS Leaf 2 thực hiện tách tiêu đề VXLAN, sau đó chuyển tiếp gói tin đến BCCS DB.



##### Hình 3.17: Flow traffic giữa BCCS Server và BCCS DB

Lưu lượng phản hồi từ BCCS DB về máy chủ BCCS:

* Máy chủ BCCS DB khởi tạo kết nối đến máy chủ BCCS (cổng leaf BCCS).
* Do cấu hình cổng phân phối, BCCS Leaf 2 lúc này nhìn mạng con F (BCCS Server đấu nối trực tiếp) – sẽ thực hiện đóng gói VXLAN gửi tới BCCS Leaf 1 chứa BCCS DB (thông tin này học qua tuyến đường EVPN loại 2).
* BCCS Leaf 1 thực hiện tách tiêu đề VXLAN, sau đó chuyển tiếp gói tin đến máy chủ BCCS.

### 3.3.5. Lưu lượng đồng bộ video từ Frontend về Backend

Luồng dưới đây được thiết kế phục vụ đồng bộ Video từ Frontend về Backend gồm 2 phần chính (mô tả trong hình 3.18):

- Kết nối giữa Frontend và Backend thông qua IPsec VPN.

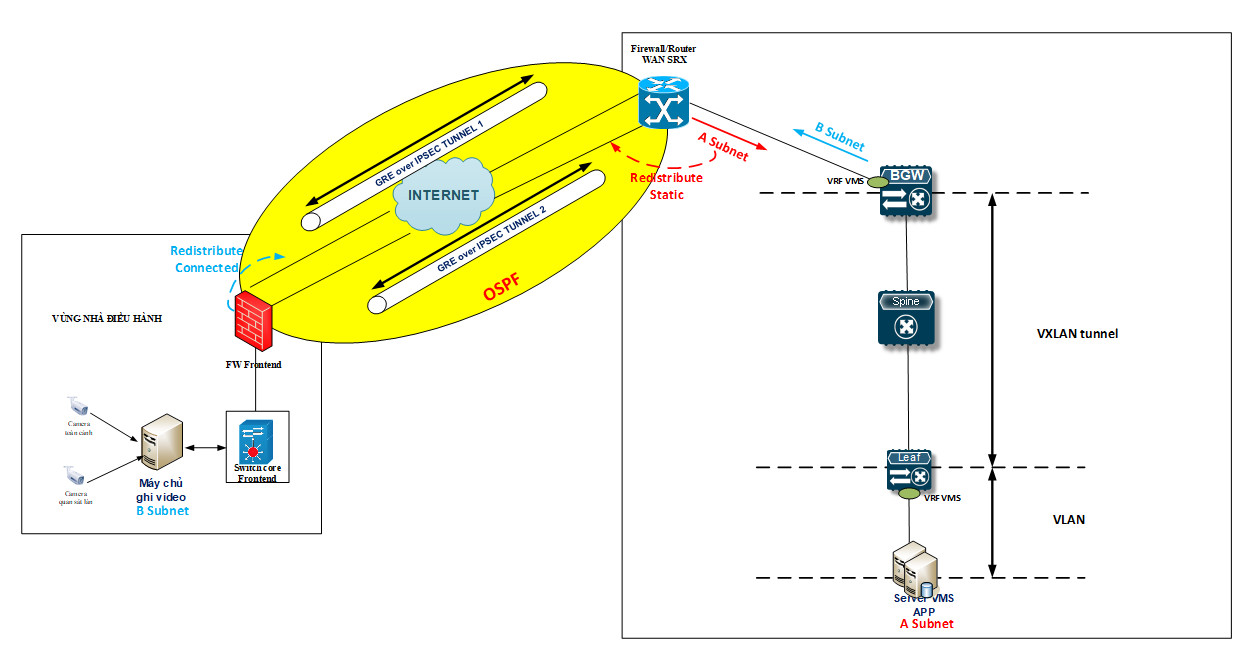
- Kết nối nội bộ trong Backend (từ Router backend tới máy chủ VMS).

Kết nối yêu cầu từ Frontend tới VMS:

* Lưu lượng Video từ Frontend được gửi tới Router BE, traffic này được gửi tới next-hop và Border Leaf – VRF VMS.
* Border Leaf học được tuyến đường loại 5 từ Leaf VMS quảng bá, Boder Leaf đóng gói VXLAN để gửi tới Leaf VMS.
* Leaf VMS tách VXLAN và gửi traffic tới VMS Server.

Kết nối phản hồi từ VMS tới Frontend:

* Máy chủ VMS khởi tạo kết nối đến Frontend (cổng Leaf VMS).
* Leaf VMS học được tuyến đường loại 5 từ Border Leaf quảng bá, Leaf VMS đóng gói VXLAN để gửi tới Border Leaf.
* Border Leaf gửi lưu lượng đến next-hop là router backend.



##### Hình 3.18: Lưu lượng đồng bộ video từ frontend về backend

### 3.3.6. Lưu lượng tính cước từ Frontend về Backend

Lưu lượng dưới đây được thiết kế phục vụ tính cước từ Frontend về Backendgồm 2 phần chính (mô tả trong hình 3.19):

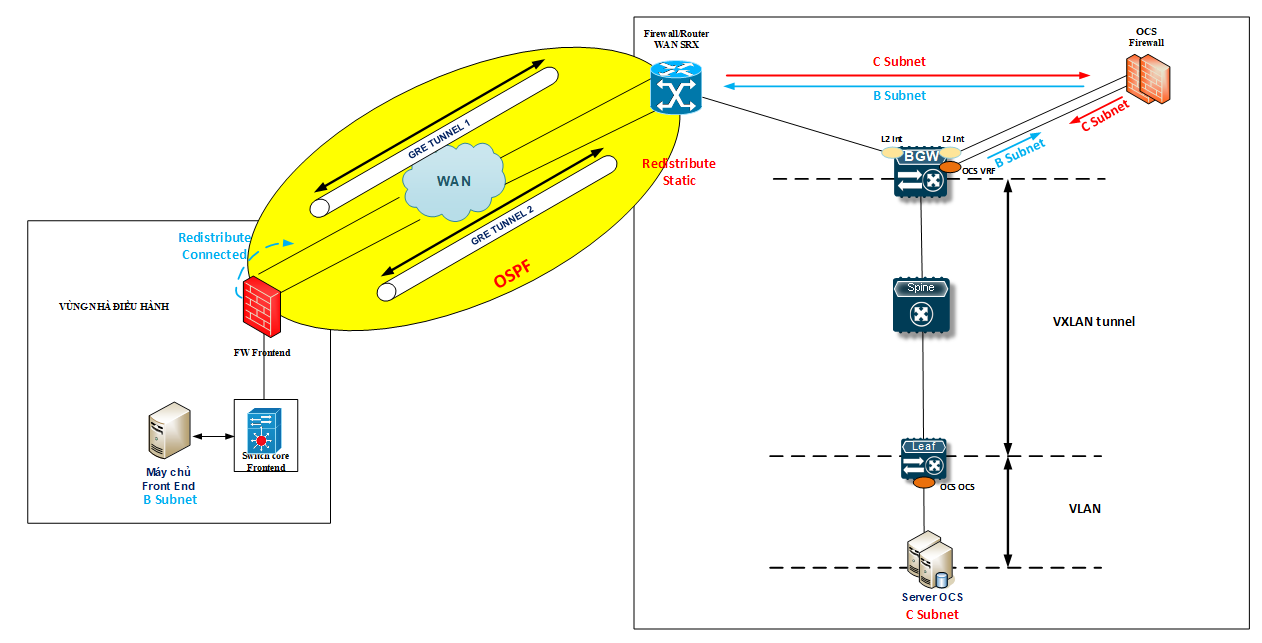
* Kết nối giữa Frontend và Backend thông qua IPsec VPN (Frontend: Fortigate Firewall và Backend: SRX).
* Kết nối nội bộ trong Backend (từ Router backend tới VMS Server).

Lưu lượng chiều yêu cầu:

* Frontend truy cập vào địa chỉ IP riêng thuộc subnet OCS, lưu lượng này sẽ đi qua Firewall OCS kiểm duyệt.
* Firewall OCS thực hiện chuyển tiếp lưu lượng vào máy chủ OCS (Nexthop Border Leaf – VRF OCS).
* Lưu lượng đến Boder Leaf được đóng gói VXLAN đến Leaf OCS.
* Leaf OCS tách header VXLAN sau đó chuyển tiếp theo L2 tới máy chủ OCS.

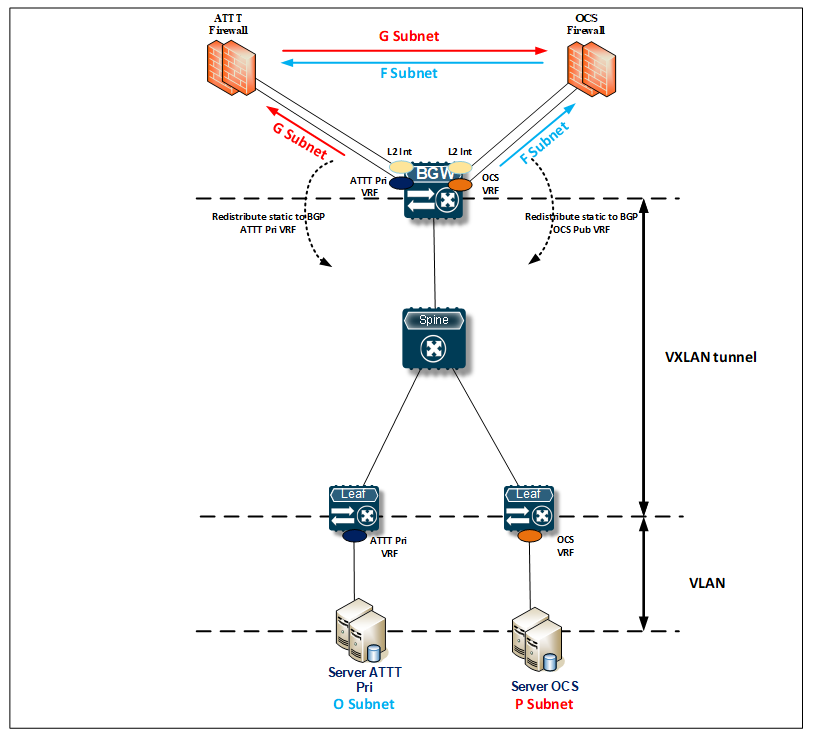
Lưu lượng chiều phản hồi:

* Máy chủ OCS gửi phản hồi lại Frontend (cổng Leaf OCS).
* Leaf OCS nhận được gói tin từ máy chủ, chèn thêm tiêu đề VXLAN và gửi tới Border Leaf.
* Border Leaf tách tiêu đề VXLAN, gửi gói tin tới Firewall OCS
* Firewall OCS lọc lưu lượng rồi gửi trả lại Frontend qua router backend.



##### Hình 3.19: Lưu lượng tính cước từ frontend về backend

### 3.3.7. Lưu lượng truy cập từ phân vùng ATTT



##### Hình 3.20: Kết nối giữa ATTT và OCS

**Lưu lượng giữa phân vùng ATTT và BCCS (hình 3.20)**

Lưu lượng chiều từ ATTT đến OCS:

* Máy chủ ATTT khởi tạo lưu lượng tới máy chủ OCS (cổng mặc định đặt trên Leaf ATTT).
* Leaf ATTT học được tuyến mặc định từ Boder Leaf qua tuyến EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF ATTT có cổng mặc định tới Firewall ATTT để ra các vùng mạng người, đóng gói truyền thống để gửi tới Firewall ATTT.
* Firewall ATTT có next-hop sang phân vùng OCS qua Firewall OCS, chuyển tiếp gói tin tới Firewall OCS qua chân đấu nối giữa 2 Firewall.
* Firewall OCS có next-hop toàn bộ mạng con phân vùng OCS là Border Leaf – VRF OCS, chuyển tiếp gói tin đến VRF OCS trên Border Leaf
* Boder Leaf – VRF OCS có tuyến EVPN loại 5 tới địa chỉ máy chủ OCS, đóng gói VXLAN gửi tới Leaf OCS.
* Leaf OCS tách tiêu đề VXLAN, chuyển tiếp gói tin tới máy chủ OCS.

Lưu lượng từ OCS về ATTT:

* Máy chủ OCS khởi tạo dữ liệu tới máy chủ ATTT (cổng mặc định đặt trên Leaf OCS).
* Leaf OCS học được tuyến mặc định từ từ Boder Leaf qua tuyến EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF OCS có cổng mặc định tới Firewall OCS để ra các vùng mạng khác, đóng gói truyền thống để gửi tới Firewall OCS.
* Firewall OCS có next-hop sang phân vùng ATTT qua Firewall ATTT, chuyển tiếp gói tin tới Firewall ATTT qua chân đấu nối giữa 2 Firewall.
* Firewall ATTT có next-hop toàn bộ mạng con phân vùng ATTT là Border Leaf – VRF ATTT, chuyển tiếp gói tin đến VRF ATTT trên Border Leaf.
* Boder Leaf – VRF ATTT có tuyến đường EVPN loại 5 tới địa chỉ máy chủ ATTT, đóng gói VXLAN gửi tới Leaf ATTT.
* Leaf ATTT tách tiêu đề VXLAN, chuyển tiếp gói tin tới máy chủ ATTT.

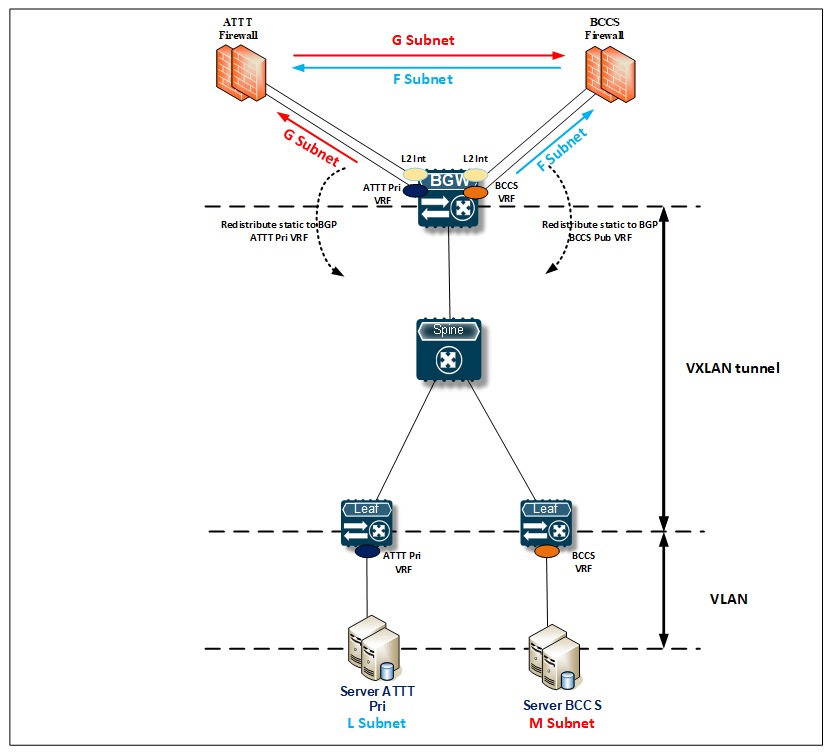
**Lưu lượng giữa phân vùng ATTT và BCCS (hình 3.21)**

Lưu lượng chiều từ ATTT đến BCCS:

* Máy chủ ATTT khởi tạo lưu lượng tới máy chủ BCCS (cổng mặc định đặt trên Leaf ATTT).
* Leaf ATTT học được tuyến mặc định từ Boder Leaf qua tuyến EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF ATTT có cổng tới Firewall ATTT để ra các vùng mạng người, đóng gói truyền thống để gửi tới Firewall ATTT.
* Firewall ATTT có next-hop sang phân vùng BCCS qua Firewall BCCS, chuyển tiếp gói tin tới Firewall BCCS qua chân đấu nối giữa 2 Firewall.
* Firewall BCCS có next-hop toàn bộ mạng con phân vùng BCCS là Border Leaf – VRF BCCS, chuyển tiếp gói tin đến VRF BCCS trên Border Leaf.
* Boder Leaf – VRF BCCS có tuyến EVPN loại 5 tới địa chỉ máy chủ BCCS, đóng gói VXLAN gửi tới Leaf BCCS.
* Leaf BCCS tách tiêu đề VXLAN, chuyển tiếp gói tin tới máy chủ BCCS

Lưu lượng từ BCCS về ATTT:

* Máy chủ BCCS khởi tạo dữ liệu tới máy chủ ATTT (cổng mặc định đặt trên Leaf BCCS).
* Leaf BCCS học được tuyến mặc định từ từ Boder Leaf qua tuyến EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF BCCS có cổng tới Firewall BCCS để ra các vùng mạng khác, đóng gói truyền thống để gửi tới Firewall BCCS.
* Firewall BCCS có next-hop sang phân vùng ATTT qua Firewall ATTT, chuyển tiếp gói tin tới Firewall ATTT qua chân đấu nối giữa 2 Firewall.
* Firewall ATTT có next-hop toàn bộ mạng con phân vùng ATTT là Border Leaf – VRF ATTT, chuyển tiếp gói tin đến VRF ATTT trên Border Leaf.
* Boder Leaf – VRF ATTT có tuyến đường EVPN loại 5 tới địa chỉ máy chủ ATTT, đóng gói VXLAN gửi tới Leaf ATTT.
* Leaf ATTT tách tiêu đề VXLAN, chuyển tiếp gói tin tới máy chủ ATTT.



##### Hình 3.21: Kết nối giữa ATTT và BCCS

**Lưu lượng giữa phân vùng ATTT và VMS**

Lưu lượng từ ATTT đến VMS:

* Máy chủ ATTT khởi tạo dữ liệu tới máy chủ VMS (cổng mặc định đặt trên Leaf ATTT).
* Leaf ATTT học được tuyến đường mặc định từ Boder Leaf qua EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF ATTT có cổng tới Firewall ATTT để ra các vùng mạng người, đóng gói truyền thống để gửi tới Firewall ATTT.
* Firewall ATTT có next-hop sang phân vùng VMS qua Border Leaf – VRF VMS , chuyển tiếp gói tin tới Border Leaf – VMS qua chân đấu nối giữa Firewall và Border Leaf.
* Boder Leaf – VRF OCS có tuyến đường EVPN loại 5 tới địa chỉ máy chủ VMS, đóng gói VXLAN gửi tới Leaf VMS.
* Leaf VMS tách tiêu đề VXLAN, chuyển tiếp gói tin tới máy chủ VMS.



##### Hình 3.22: Kết nối giữa ATTT và VMS

Lưu lượng từ VMS về ATTT:

* Máy chủ VMS khởi tạo lưu lượng tới máy chủ ATTT (cổng mặc định đặt trên Leaf VMS).
* Leaf VMS học được tuyến mặc định từ Boder Leaf qua tuyến đường EVPN loại 5, đóng gói VXLAN chuyển tới Boder Leaf.
* Boder Leaf – VRF VMS có cổng tới Firewall ATTT để ra các vùng mạng ATTT, đóng gói truyền thống để gửi tới Firewall ATTT.
* Firewall ATTT có next-hop sang phân vùng ATTT qua Border Leaf - VRF ATTT, chuyển tiếp gói tin tới Border Leaf- VRF ATTT qua chân đấu nối.
* Boder Leaf – VRF ATTT có tuyến đường EVPN loại 5 tới địa chỉ máy chủ ATTT, đóng gói VXLAN gửi tới Leaf ATTT.
* Leaf ATTT tách tiêu đề VXLAN, chuyển tiếp gói tin tới máy chủ ATTT.

## 3.4. Kết luận chương

Trạm thu phí không dừng là vấn đề rất cần thiết cho nhu cầu, tình hình hiện nay tăng tính minh bạch trong việc thu phí, góp phần giảm thiểu ùn tắc giao thông, bảo vệ môi trường. Dựa trên sở lý thuyết về nguyên tắc hoạt động của VXLAN cùng với sự kết hợp giao thức điều khiển BGP EVPN, mạng trung tâm dữ liệu thế hệ mới được xây dựng thành công nhằm triển khai dịch vụ thu phí tự động. Bằng những mô hình đang áp dụng thực tế trong mạng đã cho thấy tầm quan trọng của mạng trung tâm dữ liệu Spine-Leaf chạy trên giao thức VXLAN BGP EVPN đang dần thay thế mạng 3 lớp truyền thống.

# KẾT LUẬN

. Thông qua nội dung của ba chương đồ án trên đã trình bày về các khía cạnh cơ bản của công nghệ VXLAN, mạng trung tâm dữ liệu thế hệ mới N-DCN và ứng dụng VXLAN trong mạng N-DCN triển khai trạm thu phí không dừng. Qua đó có thể nắm rõ những khái niệm, phương thức hoạt động của VXLAN cũng như kiến trúc của mạng N-DCN. Ngoài ra xây dựng được mô hình triển khai trạm thu phí không dừng đã góp phần tiết kiệm thời gian, chi phí, giảm sự tắc nghẽn giao thông và đặc biệt có ảnh hưởng tốt đến việc giảm thiểu ô nhiễm môi trường. VXLAN cũng như mạng trung tâm dữ liệu thế hệ mới N-DCN không còn là những khái niệm xa lạ, tuy nhiên nó chỉ mới được đưa vào khai thác sử dụng trong một vài năm gần đây. Với sự bùng nổ mạnh mẽ của máy chủ ảo hóa đã đặt ra nhu cầu ngày càng cao đối với cơ sở hạ tầng mạng vật lý hiện nay, các công nghệ cũ dường như bộc lộ rất nhiều hạn chế và yếu điểm. Do đó VXLAN và mạng N-DCN như một bước đi đúng đắn và thích hợp của các nhà cung cấp dịch vụ và triển khai mạng lưới. Với những ưu điểm vượt trội của mình, trong tương lai không xa VXLAN và mạng N-DCN sẽ có thể thay thế hoàn toàn những công nghệ đã lạc hậu, góp phần thúc đẩy cuộc cách mạng 4.0, là yếu tố quan trọng báo hiệu cho một sự chuyển dịch mạnh mẽ của mạng Internet trong tương lai. Kết quả của đồ án là một bài báo đã được chấp nhận đăng trên tạp chí Bưu chính Viễn thông với tên là “Công nghệ VXLAN và ứng dụng trong mạng DCN”.

Để tiếp tục cho sự chuyển dịch về mạng sau này, việc nghiên cứu chuyên sâu hơn về ứng dụng và tăng tính bảo mật của mạng trung tâm dữ liệu thế hệ mới N-DCN dựa trên giao thức VXLAN kết hợp BGP EVPN. Tiếp tục tối ưu hóa giải pháp xây dựng trung tâm dữ liệu để đạt được quy mô lớn hơn, hiệu năng tốt hơn, tăng cường tính bảo mật, đáp ứng được lượng thông tin ngày càng nhiều mà không làm ảnh hưởng đến trải nghiệm người dùng.

# DANH MỤC TÀI LIỆU THAM KHẢO

1. *Lukas Krattiger, Shyam Kapadia & David Jansen “Building Data Centers with -VXLAN BGP EVPN”, Cisco Systems, 2017*
2. *Cisco “Cisco Data Center Spine-and-Leaf Architecture: Design Overview”, Cisco Systems, 2017*
3. *Santa Clara, “Data Center Interconnection with VXLAN”, 2017*
4. *Yves “VxLAN Evolution in the Context of DCI Requirements”, 2015*
5. *Cisco*  “*Cisco Programmable Fabric with VXLAN BGP EVPN Configuration Guide”, 2019*
6. *RFC4270, “Multiprotocol Extensions for BGP”, 2007*
7. *Dr. Jim Metzzler, Ashton Metzler & Casado, “VXLAN Configuration Guide”, Huawei System, 2017*
8. *Cisco “Ethernet VPN (EVPN) and Provider Backbone Bridging-EVPN: Next Generation Solutions for MPLS-based Ethernet Services”, 2014*
9. *RFC4271, “A Border gateway protocol 4 (BGP 4)”, 2006*
10. *Maksim Sisov, “Cisco Data Center Spine-and-Leaf Architecture: Design Overview”, Cisco System, 2016*

# PHỤ LỤC: bài báo: “Công nghệ VXLAN và ứng dụng trong mạng DCN”

**CÔNG NGHỆ VXLAN VÀ ỨNG DỤNG TRONG MẠNG DCN**

Đặng Nguyễn Minh Hưng, Phạm Anh Thư

***Tóm tắt:***

Bài báo này trình bày về sự cải tiến và lợi ích của mạng LAN ảo mở rộng (VXLAN) so với công nghệ VLAN truyền thống. Ngoài ra, phương thức hoạt động của công nghệ VXLAN và ứng dụng triển khai trong mạng trung tâm dữ liệu (DCN) cũng được phân tích.

1. **Giới thiệu chung**

Việc phân đoạn mạng truyền thống cung cấp bởi các mạng LAN ảo (VLAN) được tiêu chuẩn hóa theo IEEE 802.1Q. VLAN đưa ra sự tách biệt hợp lý giữa ranh giới lớp 2 hoặc miền quảng bá. Tuy nhiên, do việc sử dụng không hiệu quả các liên kết mạng có sẵn, các yêu cầu khắt khe về vị trí thiết bị trong mạng trung tâm dữ liệu và chỉ có khả năng mở rộng tối đa 4094 VLAN, việc sử dụng VLAN đã trở thành một yếu tố hạn chế đối với các bộ phận công nghệ thông tin và nhà cung cấp đám mây vì họ cần tạo ra những trung tâm dữ liệu quy mô lớn cho nhiều đối tượng. Vì vậy, công nghệ mạng LAN ảo mở rộng (VXLAN) đã được đề xuất như một giải pháp để đáp ứng những hạn chế, thách thức mà công nghệ VLAN truyền thống gặp phải.

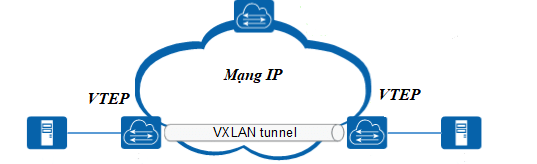
Như tên gọi của nó, VXLAN được thiết kế để cung cấp các dịch vụ mạng Ethernet lớp 2 giống như VLAN hiện nay, nhưng với khả năng mở rộng và tính linh hoạt cao hơn. So với VLAN, VXLAN mang lại những lợi ích sau:

* Sự sắp xếp linh hoạt các phân đoạn của nhiều đối tượng mạng trong toàn bộ trung tâm dữ liệu: Nó cung cấp giải pháp để mở rộng các phân đoạn lớp 2 để dữ liệu có thể được vận chuyển trên hạ tầng vật lý trong trung tâm dữ liệu.
* Khả năng mở rộng cao hơn để giải quyết nhiều phân đoạn lớp 2: Các VLAN sử dụng ID VLAN 12 bit để định danh cho các phân đoạn lớp 2, điều này dẫn đến việc hạn chế khả năng mở rộng của chỉ 4094 VLAN như đã nói ở trên. VXLAN sử dụng 24 bit để định danh mạng VXLAN (VNID), cho phép tối đa 16 triệu phân đoạn VXLAN cùng tồn tại trong cùng một miền quản trị.
* Tận dụng tốt hơn các tuyến đường có sẵn trong cơ sở hạ tầng mạng vật lý. VLAN sử dụng giao thức STP (Spanning-tree) để ngăn chặn vòng lặp, dẫn đến việc không sử dụng một nửa số liên kết mạng trong mạng bằng cách chặn các tuyến đường còn lại. Ngược lại, các gói VXLAN được chuyển qua mạng vật lý dựa trên tiêu đề lớp 3 của nó và có thể tận dụng hoàn toàn lợi ích của định tuyến Lớp 3, định tuyến đa đường chi phí ngang bằng (ECMP) và các giao thức liên kết tổng hợp để sử dụng tất cả các đường dẫn có sẵn.

Với các ưu điểm ở trên, công nghệ VXLAN được xem là giải pháp tiềm năng cho các mạng trung tâm dữ liệu DCN. Vì vậy, trong bài báo này, chúng tôi sẽ đưa ra cái nhìn tổng quan về công nghệ VXLAN và khả năng ứng dụng của nó trong mạng DCN. Bài báo có bố cục như sau. Sau phần giới thiệu, công nghệ VXLAN được trình bày ở phần II. Phần III đưa ra khả năng ứng dụng của VXLAN trong mạng DCN. Cuối bài báo là phần kết luận.

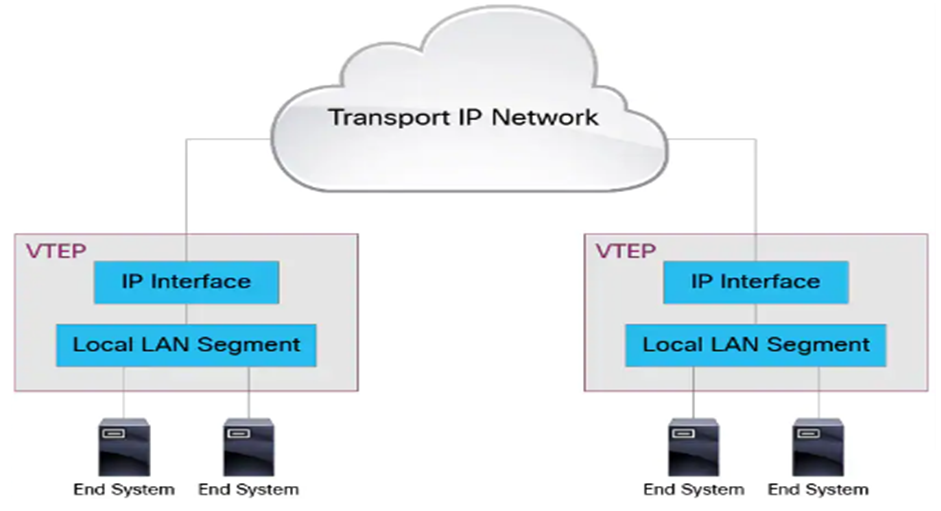
**II. Công nghệ VXLAN**

Về cốt lõi, VXLAN chỉ đơn giản là một sơ đồ đóng gói MAC trong UDP (đóng gói một khung Ethernet lớp 2 trong IP) cho phép tạo ra các mạng con ảo lớp 2 hoạt động trên mạng IP vật lý lớp 3. VXLAN cho phép kết nối giữa hai hoặc nhiều mạng lớp 3, điều này cho phép các thiết bị đầu cuối hoạt động trong các mạng riêng biệt trong khi hoạt động như thể chúng được gắn vào cùng một mạng con lớp 2. VXLAN là lớp mạng ảo Overlay. Hình 1 mô tả họat động của công nghệ VXLAN.



**Hình 1: Công nghệ VXLAN**

VXLAN sử dụng thiết bị điểm cuối đường hầm VXLAN (VTEP) để ánh xạ thiết bị đầu cuối của thuê bao tới các phân đoạn VXLAN và thực hiện đóng gói và hủy đóng gói VXLAN. Mỗi chức năng VTEP có hai giao diện: một là giao diện chuyển mạch trên phân đoạn LAN cục bộ để hỗ trợ giao tiếp điểm cuối cục bộ thông qua cầu nối, và giao diện còn lại là giao diện IP với mạng truyền tải IP. Giao diện IP có một địa chỉ IP duy nhất xác định thiết bị VTEP trên mạng truyền tải IP. Thiết bị VTEP sử dụng địa chỉ IP này để đóng gói các khung Ethernet và truyền các gói được đóng tới mạng truyền tải thông qua giao diện IP. Một thiết bị VTEP cũng phát hiện ra các VTEP từ xa cho các phân đoạn VXLAN của nó và học các ánh xạ địa chỉ MAC đến VTEP từ xa thông qua giao diện IP của nó. Các thành phần chức năng của VTEP và cấu trúc liên kết logic được tạo ra cho kết nối lớp 2 trên mạng truyền tải IP được thể hiện trong hình 2.

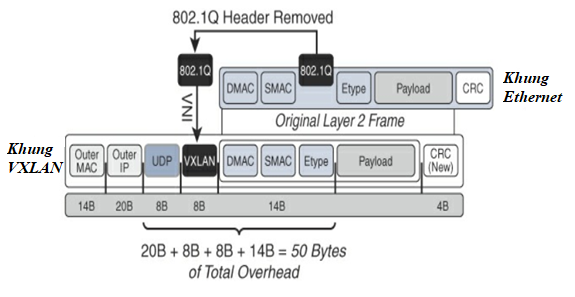


**Hình 2: Chức năng và giao diện VTEP**

Các phân đoạn VXLAN độc lập với cấu trúc liên kết mạng vật lý. Ngược lại, mạng IP vật lý giữa các VTEP độc lập với lớp mạng ảo VXLAN. Nó định tuyến các gói được đóng dựa trên tiêu đề địa chỉ IP bên ngoài, có VTEP khởi tạo là địa chỉ IP nguồn và VTEP kết thúc là địa chỉ IP đích.

VXLAN định nghĩa một lược đồ đóng gói MAC trong UDP, ở đó khung lớp 2 ban đầu có thêm tiêu đề VXLAN và sau đó được đặt trong một gói UDP-IP. Với cách đóng gói này, VXLAN là đường hầm mạng lớp 2 qua mạng lớp 3. Định dạng gói VXLAN được hiển thị trong hình 3.

Bên trong tiêu đề VXLAN đề cập đến các thành phần lớp 2, trong khi các tiêu đề bên ngoài là thông tin các thiết bị biên lớp 3. Do đó, tiêu đề VXLAN chứa khung Ethernet lớp 2 ban đầu, bao gồm địa chỉ MAC nguồn cũng như địa chỉ MAC đích bên trong, thông tin dữ liệu Ethernet và chuỗi kiểm tra lỗi (FCS). Ngoài ra, nhãn 802.1Q bên trong khung Ethernet lớp 2 ban đầu được loại bỏ và ánh xạ tới một VNI để hoàn thành tiêu đề VXLAN. Tiêu đề VXLAN đứng trước những tiêu đề như UDP, Outer IP và Outer MAC. Cổng đích của tiêu đề UDP bên ngoài được đặt với giá trị 4789. Số cổng nguồn UDP được tạo ra dựa trên các trường của tiêu đề ban đầu. Trong tiêu đề Outer IP thiết lập địa chỉ của VTEP nguồn là địa chỉ IP nguồn, địa chỉ VTEP đích là địa chỉ IP đích. Nguồn gốc của tiêu đề Outer MAC dựa trên tra cứu định tuyến trên tiêu đề Outer IP. Do đó, VXLAN cần thêm 50 byte vào khung Ethernet hiện tại hoặc 54 byte, nếu nhãn IEEE 802.1Q được thêm vào.



**Hình 3: Định dạng gói tin VXLAN**

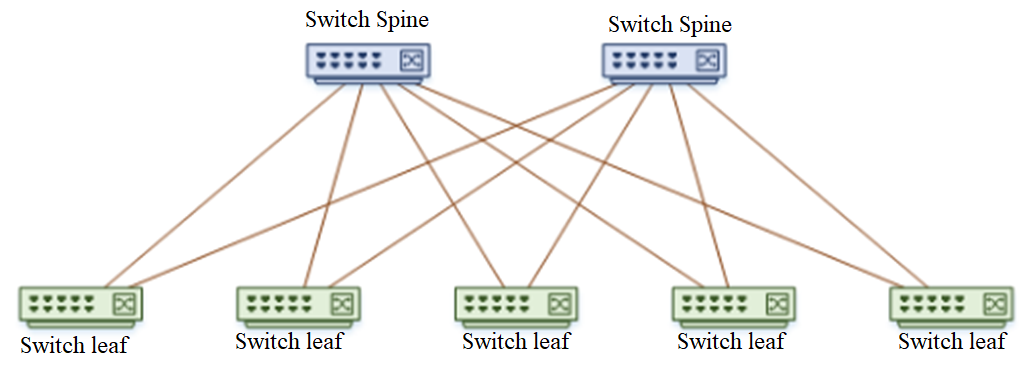
Thông thường, trong mạng lớp 3 với ECMP, cân bằng tải đạt được để chuyển tiếp lưu lượng bằng cách chọn một trong những tuyến đường có giá trị Cost như nhau dựa trên 5 thuộc tính bao gồm địa chỉ IP nguồn, địa chỉ IP đích, giao thức lớp 4 (L4), cổng nguồn L4 và cổng đích L4. Với VXLAN, đối với lưu lượng giữa các thiết bị đầu cuối phía sau cặp VTEP nguồn và đích, tất cả các trường trong tiêu đề bên ngoài, ngoại trừ UDP đều giống hệt nhau. Vì cổng nguồn UDP thay đổi dựa trên nội dung của gói/khung bên trong, nó cung cấp thông tin cần thiết để chọn các tuyến đường khác nhau cho các thiết bị đầu cuối giữa cùng một cặp VTEP nguồn và đích. Do đó, việc tạo giá trị cổng nguồn UDP một cách thích hợp là một tính năng quan trọng mà VTEP cần hỗ trợ, bất kể chúng được triển khai ở phần mềm hay phần cứng. Đây là điểm khác biệt quan trọng của VXLAN so với các lớp mạng overlay khác.

1. **Ứng dụng công nghệ VXLAN trong mạng trung tâm dữ liệu**

Trong phần này, chúng tôi đưa ra một ứng dụng của công nghệ VXLAN, đó là triển khai VXLAN trong mạng trung tâm dữ liệu spine-leaf. Cấu trúc liên kết spine-leaf thường được sử dụng trong lớp mạng IP vật lý. Nó hoàn toàn khác với hệ thống phân cấp truyền thống. Đây là một kiến trúc không dành riêng cho VXLAN. Các công nghệ khác như FabricPath cũng có thể sử dụng mô hình spine-leaf. Kiến trúc này sử dụng hai lớp:

* Lớp dưới (leaf) là nơi máy chủ và các thiết bị khác kết nối. Lớp Leaf xử lý tất cả các chức năng VXLAN, như tạo mạng ảo và ánh xạ VLAN với VNI.
* Lớp spine được sử dụng để vận chuyển. Thứ duy nhất kết nối với spine chính là các switch leaf. Tất cả các liên kết đến spine là các liên kết được định tuyến. Điều này tạo ra nhiều tuyến đường ECMP, có thể được quản lý bởi một giao thức định tuyến. Các switch spine chuyển tiếp lưu lượng đi qua chúng và không nhận biết được sự xuất hiện của VXLAN.

Nếu nhìn vào sơ đồ cấu trúc liên kết hình 4 dưới đây, ta sẽ thấy rằng các switch spine không kết nối với nhau. Các switch leaf cũng không kết nối với nhau. Điều này là để làm cho mạng trở nên nhất quán. Bằng cách này, mọi dữ liệu được truyền chỉ cần đi qua hai bước, vì tất cả các liên kết đều có cùng tốc độ và tất cả các điểm đến đều được xác định. Loại mạng này thường được gọi là mạng fabric. Một fabric như thế này rất tốt. Nếu bạn cần thêm nhiều máy chủ hơn, chỉ cần bổ sung các switch leaf. Nếu bạn cần thêm băng thông, chỉ cần bổ sung switch spine vào fabric. Khi VXLAN được sử dụng trong kiến trúc này, lợi ích là việc mở rộng lớp mạng vật lý không ảnh hưởng đến mạng Overlay (không cần thiết kế lại).



**Hình 4: Mô hình spine-leaf**

Như chúng ta đã biết, lưu lượng VXLAN được đóng gói trước khi nó được gửi qua mạng. Điều này tạo ra các đường hầm trên toàn mạng, từ switch nguồn đến switch đích. Trong kiến trúc spine-leaf này, các switch leaf đóng vai trò là các VTEP có nhiệm vụ đóng gói và mở đóng gói VXLAN. Giờ ta hay cùng xem lưu lượng truy cập qua một mạng VXLAN trên mô hình spine-leaf:

* Khung đến trên cổng của switch leaf từ một máy chủ. Cổng này là một cổng (truy cập) không được gắn thẻ thông thường, sẽ chỉ định một VLAN cho lưu lượng.
* Switch leaf xác định rằng khung cần được chuyển tiếp đến một vị trí khác.Leaf đầu xa được kết nối bằng mạng IP với số bước đến đích là 2.
* VLAN được liên kết với VNI, vì vậy tiêu đề VxLAN được áp dụng. Leaf (VTEP) đóng gói lưu lượng trong các tiêu đề UDP và IP. Cổng UDP 4789 được sử dụng làm cổng đích. Lưu lượng được gửi qua mạng IP.
* Leaf đầu xa nhận gói tin và giải mã nó. Còn lại một khung lớp 2 thông thường với VLAN ID
* Switch leaf chọn một cổng ra để gửi khung ra ngoài, điều này dựa trên tra cứu MAC thông thường. Phần còn lại của quá trình diễn ra bình thường.

Ngoài ra, trong kiến trúc spine-leaf, VXLAN với cổng anycast IP phân tán áp dụng khái niệm mạng anycast “một cho liên kết gần nhất”. Anycast là một phương pháp định tuyến và địa chỉ mạng trong đó lưu lượng dữ liệu từ một điểm cuối được định tuyến theo cấu trúc liên kết đến nút gần nhất trong một nhóm các cổng đều được xác định bởi cùng một địa chỉ IP đích. Với cổng anycast IP phân tán, cổng mặc định được di chuyển đến gần điểm cuối - cụ thể là đến switch leaf nơi mỗi điểm cuối được gắn trực tiếp vào. Cổng anycast đang hoạt động trên mỗi thiết bị biên/VTEP trên toàn bộ kết cấu mạng, loại bỏ yêu cầu phải có các gói/giao thức hello truyền thống trên kết cấu mạng. Do đó, cùng một cổng cho một mạng con có thể tồn tại đồng thời ở nhiều switch leaf, đồng thời tăng tính di động các máy chủ đứng sau leaf.

1. **Kết luận**

Bài báo chỉ ra VXLAN là một công cụ mạnh mẽ để mở rộng mạng con lớp 2 qua ranh giới mạng lớp 3. Nó giải quyết các hạn chế về tính di động của các máy chủ bằng cách đóng gói lưu lượng và mở rộng nó qua các cổng lớp 3. VXLAN cũng có thể chứa nhiều mạng con trên cơ sở hạ tầng trung tâm dữ liệu. Với phần cứng phù hợp, VXLAN có thể vượt qua giới hạn 4094 VLAN của 802.1Q mà không ảnh hưởng đến sự ổn định của mạng. VXLAN được chuyển qua mạng vật lý dựa trên tiêu đề lớp 3 của nó và có thể tận dụng hoàn toàn lợi ích của định tuyến lớp 3 ECMP và các giao thức liên kết tổng hợp để sử dụng tất cả các tuyến đường có sẵn.

**TÀI LIỆU THAM KHẢO**

*[1] Cisco “Cisco Data Center Spine-and-Leaf Architecture: Design Overview”, Cisco Systems, 2017.*

*[2] Lukas Krattiger, Shyam Kapadia & David Jansen “Building Data Centers with -VXLAN BGP EVPN”, Cisco Systems, 2017.*

*[3] Yves “VxLAN Evolution in the Context of DCI Requirements”, 2015*

*[4] Cisco “Cisco Programmable Fabric with VXLAN BGP EVPN Configuration Guide”, 2019.*