**LỜI CẢM ƠN**

Trước hết, em xin chân thành cảm ơn các thầy cô, cán bộ, công nhân viên trong Khoa Viễn Thông và toàn thể cán bộ của Học Viện Công Nghệ Bưu Chính Viễn Thông đã quan tâm và tạo mọi điều kiện thuận lợi cho em trong quá trình học tập và nghiên cứu tại Học viện cũng như trong thời gian làm đồ án tốt nghiệp.

Em xin gửi lời biết ơn sâu sắc tới Thiếu tá Nguyễn Văn Quang, anh Nguyễn Hạc Tú, anh Phạm Trọng Trường cũng như toàn thể các anh chị trong Trung tâm dịch vụ Sau bán hàng – Khối 2 thuộc Tổng công ty Công nghiệp Công nghệ cao Viettel (Viettel High Tech) là những người đã tận tình chỉ bảo, định hướng cho em trong suốt quá trình nghiên cứu và thực hiện đồ án này. Với lòng nhiệt tình, sự tâm huyết đối với sinh viên cùng sự dày dặn kinh nghiệm trong nghề, các anh chị đã giúp em có thêm nhiều kiến thức bổ ích, sự quyết tâm vượt qua những khó khăn trong quá trình nghiên cứu để hoàn thành tốt đồ án tốt nghiệp này.

Em cũng xin gửi lời cảm ơn tới các thầy cô đã giảng dạy em trong suốt 4,5 năm học tập và tích lũy kiến thức tại học viện. Đó là những người đã truyền đạt cho em những kiến thức, kinh nghiệm quý báu. Những kiến thức đó đã, đang và sẽ tiếp tục là hành trang, động lực giúp em tự hoàn thiện bản thân, vượt qua những khó khăn vấp ngã trong cuộc sống và vững bước trên con đường phía trước.

Do thời gian làm đồ án có hạn, cũng như kiến thức chuyên môn vẫn còn hạn chế, chưa thể đầy đủ nên bài làm của em chắc chắn vẫn còn rất nhiều thiếu sót. Em kính mong nhận được sự nhận xét, góp ý từ phía các thầy cô và các bạn để bài đồ án của em có thể hoàn thiện hơn.

Em xin chân thành cảm ơn!

Hà Nội, tháng 12 năm 2022

Sinh viên

**Nguyễn Phương Nam**

**DANH MỤC THUẬT NGỮ VIẾT TẮT**

|  |  |  |
| --- | --- | --- |
| **Từ viết tắt** | **Giải thích** | **Ý nghĩa** |
| **BGP** | Border Gateway Protocol | Giao thức định tuyến liên miền |
| **DC** | Data Center | Trung tâm dữ liệu |
| **ECMP** | Equal Cost MultiPath Routing | Định tuyến đa đường chi phí ngang nhau |
| **EVPN** | Ethernet Virtual Private Network | Mạng riêng ảo Ethernet |
| **FW** | Firewall | Tường lửa |
| **IRB** | Integrated Route and Bridge | Định tuyến và chuyển mạch tích hợp |
| **MAC** | Media Access Control | Kiểm soát truy cập phương tiện truyền thông |
| **NAT** | Network Address Translation | Biên dịch địa chỉ mạng |
| **NLRI** | Network Layer Reachability Information | Thông tin truy cập lớp mạng |
| **UDP** | User Datagram Protocol | Giao thức dữ liệu người dùng |
| **VLAN** | Virtual Local Area Network | Mạng LAN 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 |
| **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 |
| **QoS** | Quality of Service | Chất lượng dịch vụ |
| **VACL** | VLAN Access List | Danh sách truy cập ảo |
| **ARP** | Address Resolution Protocol | Giao thức phân giải địa chỉ |
| **MPLS** | Multiprotocol Label Switching | Chuyển mạch nhãn đa giao thức |

**LỜI MỞ ĐẦU**

Ngày nay mạng máy tính ngày càng phát triển. Nhu cầu mở rộng, quản lý mạng ngày càng tăng, đòi hỏi số lượng thiết bị ngày càng lớn gây ra khó khăn cho người quản trị như khả năng mở rộng kém, lãng phí tài nguyên mạng. Mô hình truyền thống ba lớp đang được sử dụng rất nhiều trong hệ thống Data Center hiện nay là mô hình ba lớp (Core – Distribution – Access). Sự phát triển nhanh chóng về ảo hóa, cloud và BIG DATA làm cho mô hình ba lớp nở to và không còn linh hoạt nữa nên xu hướng thế giới nói chung đang chuyển sang mô hình thiết kế mới Spine – Leaf hoạt động trên giao thức VXLAN kết hợp với giao thức điều khiển BGP EVPN.

VXLAN là một kiểu đóng gói dựa trên UDP. Phần tận cùng bên ngoài của gói tin là định dạng chuẩn UDP và như vậy có thể định tuyến được bởi bất kỳ thiết bị IP nào. Phần bên trong là một phần được đóng gói bao gồm toàn bộ một Ethernet frame, địa chỉ MAC nguồn ban đầu, địa chỉ MAC đích và phần Ethernet header nguyên thủy ban đầu, thêm vào đó là gói tin IP được chứa bên trong nó. Cách đóng gói của VXLAN này là quan trọng vì nó có nghĩa là VXLAN là một công nghệ có thể hỗ trợ cho các giải pháp mạng ảo ở cả lớp 2 và lớp 3.

Do đó em quyết định chọn đề tài “Tổng quan mô hình mạng Spine-Leaf và áp dụng triển khai trong hệ thống mạng tại phòng lab của HVCNBCVT”. Mục đích chính của nghiên cứu này là đưa ra cái nhìn tổng quan về kiến trúc mạng hoàn toàn mới, tìm hiểu cách hoạt động của giao thức VXLAN với mô hình Spine-Leaf trong hệ thống mạng doanh nghiệp hiện tại.

Nội dung đồ án chia làm 4 chương:

* Chương 1: Tổng quan về mô hình mạng truyền thống và mô hình mạng Spine-Leaf.
* Chương 2: Tìm hiểu các giao thức, công nghệ sử dụng trong mô hình Spine-Leaf.
* Chương 3: Xây dựng mô hình Spine-Leaf demo trên phần mềm mô phỏng.
* Chương 4:Tìm hiểu về hạ tầng mạng đã triển khai tại phòng Lab HVCNBCVT và đề xuất chuyển đổi mô hình Spine-Leaf trong tương lai.

Em xin chân thành cảm ơn!

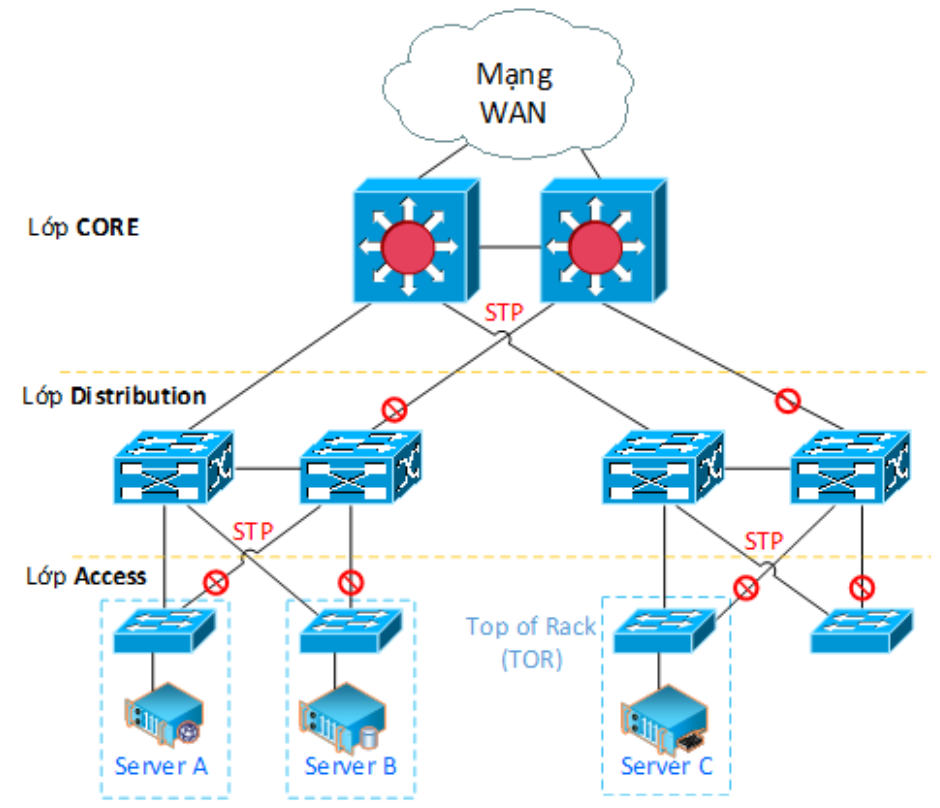
Hà Nội, tháng 12 năm 2022

Sinh viên thực hiện

**Nguyễn Phương Nam**

**CHƯƠNG 1 : TỔNG QUAN VỀ MÔ HÌNH MẠNG TRUYỀN THỐNG VÀ MÔ HÌNH SPINE – LEAF**

**1.1 Kiến trúc phân cấp 3 lớp**



**Hình 1.1: Mô hình mạng truyền thống**

Trong hơn một thập kỷ, chúng ta đã sử dụng kiến ​​trúc này với truy cập ba lớp:

* Core
* Distribution
* Access

**1.1.1 Core Layer ( Lớp mạng lõi )**

Tốc độ vận chuyển dữ liệu rất nhanh, liên kết với các lớp mạng truy cập và lớp mạng phân phối. Lớp này còn được coi là đại lộ liên kết các đường nhỏ với nhau. Nếu có một sự hư hỏng xảy ra ở lớp lõi, hầu hết các người dùng trong mạng LAN đều bị ảnh hưởng. Vì vậy, sự dự phòng là rất cần thiết tại lớp này. Do lớp lõi vận chuyển một số lượng lớn dữ liệu, nên độ trễ tại lớp lõi phải là cực nhỏ.

Tại lớp lõi, ta không nên làm bất cứ một điều gì có thể ảnh hưởng đến tốc độ chuyển mạch tại lớp lõi như là tạo các access list, routing giữa các VLAN với nhau hay packet filtering.

Các đặc điểm Lớp lõi bao gồm :

* Vận chuyển nhanh
* Độ tin cậy cao
* Có tính dự phòng
* Khả năng chịu lỗi
* Độ trễ thấp, quản lý tốt
* Chất lượng dịch vụ (QoS)

**1.1.2 Distribution layer (Lớp phân phối)**

Lớp phân phối là phần liên kết ở giữa lớp truy cập và lớp lõi, đáp ứng một số giao tiếp giúp giảm tải cho lớp lõi trong quá trình truyền thông tin trong mạng.

Chức năng chính của lớp phân phối là xử lý dữ liệu như : định tuyến (routing), lọc gói (filtering), truy cập mạng WAN, tạo access list,...

Lớp phân phối phải xác định được con đường nhanh nhất đáp ứng các yêu cầu của user. Sau khi xác định được con đường nhanh nhất, nó gửi các yêu cầu đến lớp lõi. Lớp lõi chịu trách nhiệm chuyển mạch các yêu cầu đến đúng dịch vụ cần thiết.

Lớp phân phối là nơi thực hiện các chính sách (policies) cho mạng, cung cấp tập hợp các tuyến đường đến mạng lõi. Trong phạm vi mạng LAN, lớp phân phối cung cấp định tuyến giữa các VLAN, bảo mật và QoS.

Lớp phân phối có thể có nhiều vai trò, bao gồm cả việc thực hiện các chức năng sau:

* Kết nối dựa trên chính sách (ví dụ, đảm bảo rằng lưu lượng truy cập gửi từ một mạng cụ thể được chuyển tiếp ra một giao tiếp trong khi tất cả các lưu lượng khác được chuyển tiếp ra giao tiếp khác).
* Dự phòng và cân bằng tải
* Tập hợp các kết nối LAN, WAN
* Phân địa chỉ, kết hợp các phân vùng
* Định tuyến giữa các mạng LAN ảo (VLAN)
* Truyền trung gian (ví dụ, giữa Ethernet và Token Ring)
* Có thể sử dụng một số tính năng phần mềm Cisco IOS làm phương tiện thực hiện.
* Lọc địa chỉ nguồn hoặc địa chỉ đích
* Lọc cổng đầu vào hoặc đầu ra
* Định tyến tĩnh
* Cơ chế QoS, chẳng hạn như dựa trên xếp hàng ưu tiên

**1.1.3 Access Layer (Lớp truy cập)**

Lớp truy cập được thiết kế để cung cấp các cổng kết nối đến từng máy trạm trên cùng một mạng, giúp người dùng kết nối với các tài nguyên trên mạng hoặc giao tiếp với lớp phân phối. Lớp này sử dụng các chính sách truy cập chống lại những kẻ xâm nhập bất hợp pháp, mang đến các kết nối như: WAN, Frame Relay (chuyển mạch gói), Leased Lines (đường dây thuê bao).

Lớp truy cập đặc trưng bởi các phân đoạn mạng LAN.

Đặc tính của lớp truy cập bao gồm :

* Chuyển mạch lớp 2
* Hiệu quả cao
* Bảo mật cổng
* Ngăn Broadcast
* Phân loại mức độ ưu tiên QoS
* Kiểm soát tốc độ
* Kiểm soát danh sách truy cập ảo (VACL)
* Spanning tree
* Hỗ trợ VLAN

**1.2 Kiến trúc mạng Spine-Leaf**

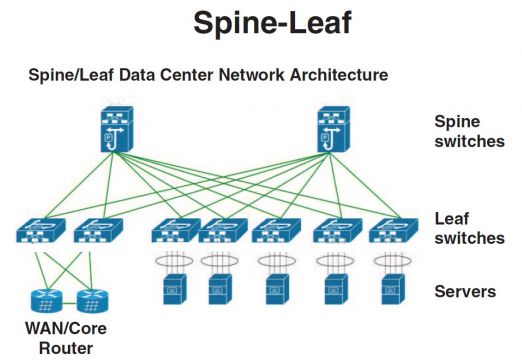
**1.2.1 Tại sao phải sử dụng mô hình mạng Spine-Leaf**

Đối với Datacenter: các công nghệ ảo hóa, điện toán đám mây, sự phát triển bùng nổ của các ứng dụng nên đòi hỏi nhu cầu về băng thông cao hơn, số lượng Server kết nối nhiều hơn, yêu cầu quản lý đơn giản hơn, chất lượng cao hơn …

Đối với doanh nghiệp: quy mô ngày càng lớn, số lượng thiết bị kết nối tăng lên, các kết nối Wifi ngày càng nhiều. Xu hướng BYOD (Bring your own device) trong các doanh nghiệp ngày càng phổ biến và đặc biệt việc triển khai áp dụng các thiết bị IoT trong doanh nghiệp sẽ tăng lên. Vì thế nhu cầu về băng thông cũng đặt ra yêu cầu thay đổi kiến trúc đối với doanh nghiệp.Phần cứng hiện tại không đủ băng thông, traffic thường phải đi qua lớp Core hoặc Distribution, Spanning-Tree block các port dự phòng và nhiều nguyên nhân khác dẫn tới nhiều hạn chế trong mô hình truyền thống.

Mô hình Spine-Leaf được phát triển để khắc phục những hạn chế của mô hình mạng cũ và đang dần trở thành xu hướng mới cho các tiêu chuẩn thiết kế Network - tạo ra một kiến ​​trúc truyền thông nhanh, có thể dự đoán, có thể mở rộng và hiệu quả trong môi trường trung tâm dữ liệu.

**1.2.2 Kiến trúc mạng Spine–Leaf**



**Hình 1.2: Mô hình kiến trúc mạng Spine-Leaf**

Kiến trúc Spine-Leaf thường được triển khai thành hai lớp: Spine (vai trò giống lớp Distribution) và Leaf (vai trò giống lớp Access). Kiến trúc Spine-Leaf cung cấp băng thông cao, độ trễ thấp và Non-Blocking giữa các Server. Đấu nối mỗi liên kết giữa Spine và Leaf có thể đạt băng thông từ 40Gb tới 100Gb tùy theo nhu cầu thiết kế.

* Leaf Switch cung cấp các kết nối tới các Server, Firewall, Router, Load Balancing, … Các Leaf Switch được kết nối tới tất cả Spine theo mô hình Full-Mesh nhưng không kết nối ngang hàng với nhau (Leaf-Leaf) .
* Spine Switch có nhiệm vụ kết nối với tất cả Leaf Switch để điều phối và thực hiện các chức năng Routing giữa các Leaf Group, các Spine cũng không kết nối ngang hang với nhau. Liên kết giữa Leaf và Spine có thể là liên kết Layer 2 hoặc Layer 3 sử dụng các giao thức định tuyến IGP.

**1.2.3 So sánh những ưu điểm của mô hình Spine-Leaf so với mô hình mạng truyền thống**

|  |  |  |
| --- | --- | --- |
|  | **Mô hình Spine-Leaf** | **Mô hình truyền thống** |
| Băng thông | - Tất cả các liên kết đều được sử dụng để cân bằng tải nhờ sử dụng giao thức ECMP do các kết nối sử dụng các cổng có băng thông bằng nhau. Điều này giúp làm tăng băng thông của hệ thống. | - Sử dụng giao thức Spanning-Tree (STP) để phòng chống loop ở lớp 2. Giao thức này sẽ block những kết nối dự phòng nên băng thông bị giảm đi. Ngoài ra khi có sự cố, thời gian hội tụ, tính toán lại của STP rất lâu. |
| Độ trễ | - Gateway đặt ngay trên Leaf Switch cho nên lưu lượng truy cập từ server luôn phải vượt qua cùng một số thiết bị để đến máy chủ khác. Cách tiếp cận này khiến cho độ trễ thấp vì một lưu lượng chỉ phải đi tới một Leaf Switch và một Spine Switch khác để đến đích. | - Gateway đặt tại Core Switch nên lưu lượng từ server phải đi qua Access Switch, Distribution Switch rồi mới tới được Core cho nên độ trễ của nó sẽ cao hơn mô hình Spine-Leaf. |
| Tính dự phòng | - Nếu 1 thiết bị Spine bị lỗi, nó chỉ giảm 1 phần nhỏ hiệu suất của mạng mà không ảnh hưởng đến dịch vụ. Nếu 1 Leaf bị lỗi, nó chỉ ảnh hưởng đến các máy chủ đang kết nối tới Leaf Switch đó. | - Do Spanning Tree block các kết nối dự phòng nên mô hình truyền thống không thể sử dụng được các liên kết dự phòng. |
| Khả năng mở rộng | - Nếu cần thêm băng thông, chỉ cần thêm Spine Switch, nếu cần thêm nhiều máy chủ, chỉ cần thêm Leaf Switch mà không phải thiết kế lại toàn bộ hệ thống. | - Với sự phát triển nhanh chóng của các công nghệ ảo hóa, cloud, BIG DATA đòi hỏi nhu cầu về băng thông cao hơn, số lượng server kết nối nhiều hơn làm cho mô hình 3 lớp bị nở to khiến cơ sở hạ tầng khó có thể đáp ứng được. |

**1.2.4 Kết luận chương 1**

Mô hình mạng Spine-Leaf cung cấp nhiều lợi ích độc đáo hơn so với mô hình 3 lớp truyền thống. Việc sử dụng định tuyến lớp 3 với ECMP để cân bằng tải bằng cách sử dụng tất cả các liên kết có sẵn giúp làm tăng băng thông của hệ thống. Với cấu hình và thiết kế dễ điều chỉnh, Spine-Leaf đã cải thiện khả năng quản lý và khả năng mở rộng thiết kế. Việc loại bỏ Spanning Tree (STP) đã làm cho sự ổn định mạng được cải thiện một cách đáng kể.

**CHƯƠNG 2: TÌM HIỂU VỀ CÁC GIAO THỨC, CÔNG NGHỆ SỬ DỤNG TRONG MÔ HÌNH MẠNG SPINE-LEAF**

**2.Tổng quan về VLAN và VXLAN**

**2.1 VLAN truyền thống**

VLAN là một nhóm các thiết bị trên một hoặc nhiều LAN được cấu hình để giao tiếp như thể chúng được gắn vào cùng một dây, trong khi thực tế chúng được đặt trên một số phân đoạn LAN khác nhau. Bởi vì VLAN dựa trên kết nối logic thay vì kết nối vật lý nên chúng cực kỳ linh hoạt.

Hầu hết chúng ta đều đã và đang triển khai VLAN trong hệ thống mạng và chúng ta hiểu rõ về các đặc tính của VLAN. Đây là cách truyền thống để phân chia các mạng con và tăng cường yếu tố bảo mật trong việc kiểm soát được đường đi của gói tin dạng broadcast. Với tiêu chuẩn IEEE 802.1Q dùng VLAN trunking, chúng ta có tối đa 4094 Vlan. Điều này nghe như đã đủ đáp ứng nhu cầu. Tuy nhiên trong những năm gần đây khi mà người dùng chuyển qua điện toán đám mây Cloud thì con số trên không còn đáp ứng được nữa (trong một Data Center). Chúng ta cũng nên nhớ rằng VLAN chỉ hoạt động ở Layer 2 chứ không phải ở Layer 3 (theo mô hình OSI). Tuy nhiên trên thực tế có những tình huống mà chúng ta cần một giải pháp chồng lấn giữa Layer 2 và Layer 3 ví dụ như triển khai hệ thống Server Clustering để nâng cao khả năng sẵn sàng cho hệ thống mạng, hay kịch bản di chuyển các máy ảo giữa các site khác nhau theo nhu cầu hay khôi phục sau thảm họa…Lúc này thì VLAN truyền thống không thể giải quyết được vấn đề này nhưng VXLAN thì có thể.

**2.2 Khái niệm VXLAN**

VXLAN là kỹ thuật đóng gói MAC trong IP/UDP (MAC-in-UDP) với 24 bit định danh phân đoạn dưới dạng ID VXLAN cho phép mở rộng lên đến 16 triệu phân đoạn mạng trong điện toán đám mây. Ngoài ra, đóng gói IP/UDP cho phép mỗi phân đoạn mạng được mở rộng trên các mạng layer 3 hiện có sử dụng công nghệ ECMP (định tuyến đa đường chi phí ngang nhau).

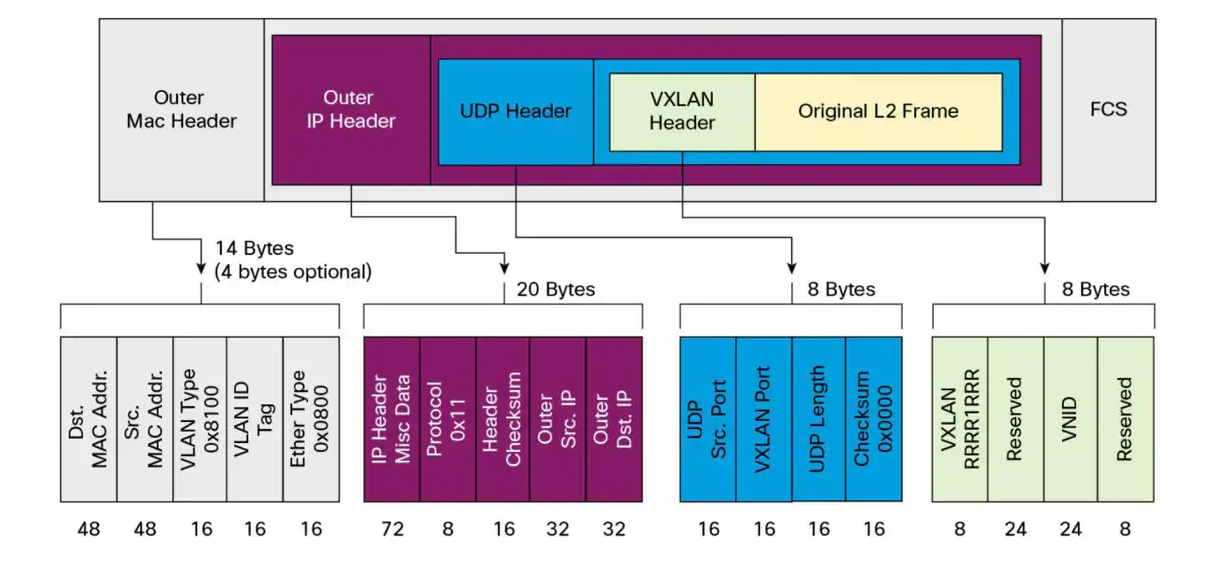
VXLAN cung cấp một cách để mở rộng các mạng lớp 2 trên cơ sở hạ tầng mạng lớp 3 bằng cách sử dụng đóng gói và đường hầm MAC-in-UDP. VXLAN cho phép các vị trí khối lượng công việc linh hoạt bằng cách sử dụng tiện ích mở rộng lớp 2. Nó cũng có thể là một cách tiếp cận để xây dựng một trung tâm dữ liệu nhiều người thuê bằng cách tách các phân đoạn lớp 2 của người thuê từ mạng lưới vận chuyển chung.

**2.3 Những ưu điểm của VXLAN so với VLAN**

|  |  |  |
| --- | --- | --- |
|  | **VXLAN** | **VLAN** |
| Về bản chất | VXLAN được coi là công nghệ mở rộng của VLAN truyền thống. Do đó, VXLAN đã và đang mang lại nhiều lợi ích hơn so với VLAN truyền thống. Nó đáp ứng được nhu cầu xử lí lưu lượng dữ liệu lớn trong môi trường Cloud hiện nay. | Với sự phát triển bùng nổ và nhanh chóng của các công nghệ ảo hóa, cloud, BIG DATA cho nên VLAN dần dần không còn đáp ứng được các nhu cầu trong các Data Center hiện nay. |
| Các giao thức | VXLAN sử dụng giao thức định tuyến lớp 3 ECMP để cân bằng tải và sử dụng tất cả các cổng có sẵn. | VLAN sử dụng giao thức Spanning-Tree (STP) chặn các đường dẫn dư thừa để tránh loop do đó chỉ cho phép sử dụng một nửa số đường dẫn có sẵn. |
| Khả năng mở rộng | VXLAN sử dụng ID 24 bit có thể mở rộng quy mô lên đến 16 triệu phân đoạn VXLAN trong một miền quản trị. | VLAN sử dụng VLAN ID 12 bit, có nghĩa là khả năng mở rộng tối đa 4094 VLAN trên một mạng. |
| Cách thức hoạt động | VXLAN sẽ sử dụng tính năng đóng gói MAC-in-UDP để mở rộng các phân đoạn lớp 2 qua các vị trí, các mạng này sẽ không bị giới hạn về mặt địa lý và mang lại khả năng mở rộng với quy mô lớn. | Một mạng LAN vật lý sẽ được chia nhỏ để giới hạn mạng trong một phạm vi địa lý nhất định. Nó sẽ sử dụng thẻ VLAN trên khung lớp 2 để đóng gói nhằm mở rộng VLAN qua các thiết bị chuyển mạch. |

**2.4 Định dạng gói tin VXLAN**

VXLAN xác định sơ đồ đóng gói MAC-in-UDP trong đó khung lớp 2 gốc có tiêu đề VXLAN được thêm vào và sau đó được đặt trong gói UDP-IP. Với đóng gói MAC-in-UDP này, VXLAN sẽ điều chỉnh mạng lớp 2 qua mạng lớp 3. Định dạng gói VXLAN được hiển thị trong hình dưới.



**Hình 2.1: Định dạng gói tin VXLAN**

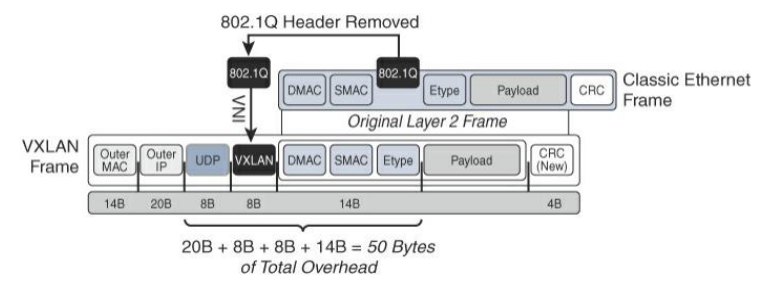
Như được hiển thị trong Hình 2.1, VXLAN Header 8 byte bao gồm 24 bit VNID và một vài bit dành riêng. VXLAN Header cùng với khung Ethernet ban đầu đi vào tải trọng UDP. 24 bit VNID được sử dụng để xác định các phân đoạn lớp 2 và để duy trì sự cô lập lớp 2 giữa các phân đoạn. Với tất cả 24 bit trong VNID, VXLAN có thể hỗ trợ 16 triệu phân đoạn LAN.

**2.5 Các thành phần và nguyên lý hoạt động của VXLAN**

Giao thức VXLAN sẽ đóng gói các khung Ethernet lớp 2 vào trong packet UDP lớp 3 để tạo các kết nối mạng lớp 2 giữa môi trường mạng lớp 3.

**2.5.1 Gói tin VXLAN**

Các giá trị bên trong của tiêu đề VXLAN đề cập đến các thành phần Lớp 2, trong khi các tiêu đề bên ngoài phản ánh các thiết bị Lớp 3. Do đó, tiêu đề VXLAN chứa gói khung Ethernet lớp 2, bao gồm địa chỉ MAC nguồn cũng như địa chỉ MAC đích, tải trọng Ethernet ban đầu và trình tự kiểm tra khung. Ngoài ra, tiêu đề 802.1Q bên trong ban đầu của khung Ethernet Lớp 2 được xóa và ánh xạ tới một VNI để hoàn thành tiêu đề VXLAN, như được minh họa trong hình sau:



**Hình 2.2 : Frame VXLAN**

**- VXLAN header**: 8 byte bao gồm các trường quan trọng sau:

* **Flags**: 8 bit, trong đó bit thứ 5 (I flag) được thiết lập là 1 để chỉ ra rằng đó là một frame có VNI có giá trị. 7 bit còn lại dùng để dự trữ và được thiết lập là 0.
* **VNI**: 24 bit cung cấp định danh duy nhất cho VXLAN segment. Các VM trong các VXLAN khác nhau không thể giao tiếp với nhau. 24 bit VNI cung cấp lên tới hơn 16 triệu VXLAN segment trong một vùng quản trị mạng.

**- Outer UDP Header**: port nguồn của Outer UDP được gán tự động và sinh ra bởi VTEP và port đích thông thường được sử dụng là port 4789 (có thể chọn port khác).

**- Outer IP Header**: Cung cấp địa chỉ IP nguồn của VTEP nguồn kết nối với VM bên trong. Địa chỉ IP outer đích là địa chỉ IP của VTEP nhận frame.

IP SA: địa chỉ IP nguồn, là địa chỉ IP của VTEP cục bộ của đường hầm VXLAN.

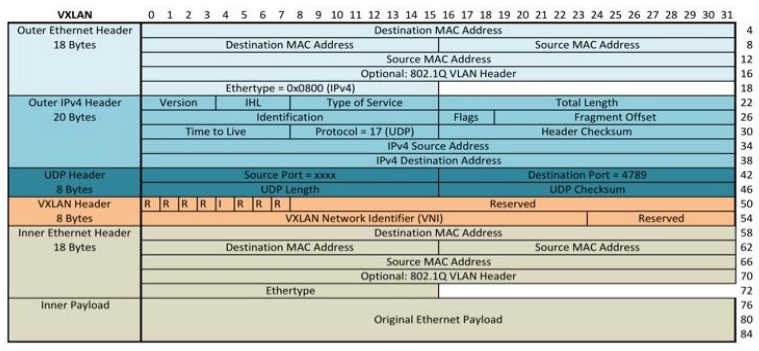
IP DA: địa chỉ IP đích, là địa chỉ IP của VTEP từ xa của đường hầm VXLAN.

Nếu Underlay Network là mạng IPv4, địa chỉ IP của VTEP là địa chỉ IPv4. Nếu Underlay Network là mạng IPv6, địa chỉ IP của VTEP là địa chỉ IPv6.

**- Outer Ethernet Header**: cung cấp địa chỉ MAC nguồn của VTEP có khung frame ban đầu. Địa chỉ MAC đích là địa chỉ của hop tiếp theo được định tuyến bởi VTEP. Outer Ethernet header có thể được gắn tag theo chuẩn 802.1Q trong quá trình vận chuyển trong mạng.

MAC DA: địa chỉ MAC đích, là địa chỉ MAC được ánh xạ tới địa chỉ IP của hop tiếp theo dựa trên địa chỉ VTEP đích trong bảng định tuyến của VTEP nơi VM gửi các gói nằm trong đó.

MAC SA: địa chỉ MAC nguồn, là địa chỉ MAC của VTEP mà VM gửi gói nằm trong đó.



**Hình 2.3 : Các trường trong VXLAN header**

**2.5.2 VXLAN segment**

VXLAN segment là layer 2 Overlay Network mà các VM giao tiếp, chỉ các máy ảo trong cùng phân khúc VXLAN mới có thể giao tiếp với nhau. Mỗi VXLAN segment được xác định qua VXLAN network ID (VNI) có chiều dài 24 bit.

**2.5.3 VTEP – Virtual Tunnel Endpoint**

Là các thiết bị phần cứng hoặc phần mềm, được đặt ở các vùng biên của mạng, chịu trách nhiệm khởi tạo VXLAN Tunnel và thực hiện đóng gói và giải mã VXLAN. Mỗi chức năng VTEP có hai giao diện: một là giao diện chuyển mạch trên phân đoạn LAN cục bộ để hỗ trợ điểm cuối cục bộ giao tiếp và phần còn lại là giao diện IP cho mạng IP truyền tải.

Một thiết bị VTEP có thể đóng vai trò như một VxLAN L2 gateway hoặc VxLAN L3 gateway. Các VTEP sẽ thực hiện chức năng *“Bridge”* đối với L2 VNID. Với các frame muốn đi qua các VN-ID khác (giống như chức năng routing giữa các vlan truyền thống), lúc này VTEP sẽ “route” frame đến L3 VN-ID để thực hiện chức năng định tuyến.

Cơ sở hạ tầng VLAN là 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 được đóng gói đến 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 từ xa thông qua giao diện IP của nó.

Các segment VXLAN độc lập với cấu trúc liên kết mạng bên dưới, ngược lại, mạng IP bên dưới giữa các VTEP không phụ thuộc vào lớp phủ VXLAN. Nó định tuyến các gói được đóng gói dựa trên tiêu đề địa chỉ IP bên ngoài, tiêu đề này có VTEP khởi tạo là địa chỉ IP nguồn và VTEP kết thúc là địa chỉ IP đích.

**2.5.4 VNI – Virtual Network Identifier**

Định danh mạng ảo (VNI) là một giá trị xác định một mạng ảo cụ thể trong mặt phẳng dữ liệu. Đây thường là một phần giá trị 24 bit của VXLAN Header, có thể hỗ trợ tới 16 triệu phân đoạn mạng riêng lẻ. (Giá trị VNI hợp lệ là từ 4096 đến 16.777.215.) Có hai phạm vi VNI chính:

* VNI phạm vi toàn mạng: cùng một giá trị được sử dụng để xác định mạng ảo lớp 3 cụ thể trên tất cả các thiết bị mạng biên. Phạm vi mạng này rất hữu ích trong các môi trường như trong trung tâm dữ liệu nơi các mạng có thể được tự động cung cấp bởi các hệ thống điều phối trung tâm.

Có VNI đồng nhất mỗi VPN là một cách tiếp cận đơn giản, đồng thời làm giảm bớt các hoạt động mạng (như khắc phục sự cố). Nó cũng có nghĩa là yêu cầu đơn giản hóa trên các thiết bị mạng biên, cả thiết bị vật lý và ảo. Một yêu cầu quan trọng đối với loại phương pháp này là có một số lượng rất lớn các giá trị định danh mạng được đưa ra theo phạm vi toàn mạng.

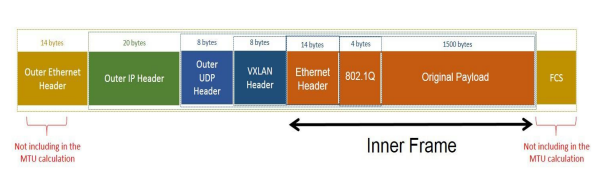
* VNI được chỉ định cục bộ: Theo cách tiếp cận thay thế được hỗ trợ theo RFC 4364, mã định danh có ý nghĩa cục bộ đối với thiết bị mạng biên quảng bá tuyến đường. Trong trường hợp này, tác động của quy mô mạng ảo được xác định trên cơ sở mỗi nút so với cơ sở mạng.

Khi nó được sử dụng cục bộ và sử dụng cùng một ngữ nghĩa hiện có như nhãn MPLS VPN, các hành vi chuyển tiếp tương tự như được chỉ định trong RFC 4364 có thể được sử dụng. Do đó, phạm vi này cho phép các khâu của VPN liền mạch với nhau.

**2.5.5 VSI (Virtual Switch Instance)**

VSI - Virtual Switch Instance là khái niệm một vùng Layer 2 (Ethernet) Broadcast. Khái niệm này tương đương với khái niệm VLAN truyền thống.

Đặc trưng của mỗi VSI sẽ là VXLAN ID có chiều dài 24 bit so với 12 bit của VLAN ID. Chính vì vậy, trong 1 DC thế hệ mới, số lượng vùng Broadcast Layer 2 có thể phân chia nhiều hơn rất nhiều (16 triệu).



**Hình 2.4: Chiều dài gói tin VXLAN**

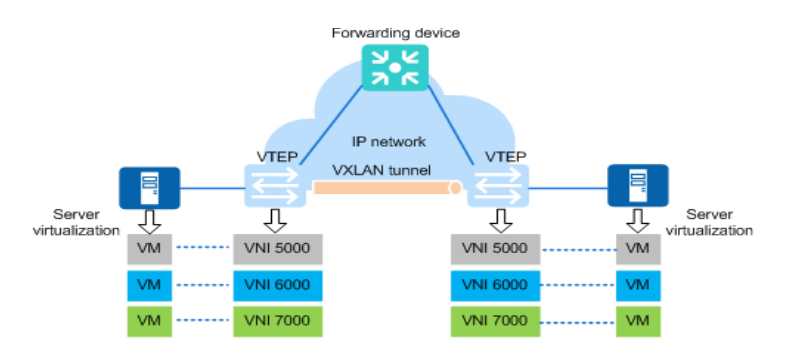
**2.5.6 VSI Interface**

VSI Interface là một interface logic (ảo) được tạo ra nhằm mục đích định tuyến cho các host thuộc một VSI. Khái niệm này tương đương với khái niệm VLAN interface truyền thống. Thông thường, VSI Interface sẽ là default gateway cho các host thuộc VSI tương ứng.

**2.5.7 Tenant/VRF**

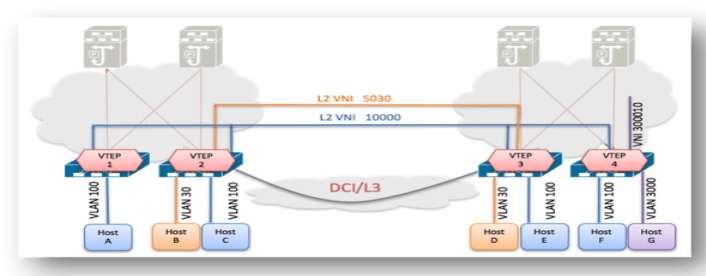
Tenant/VRF là khái niệm một đối tượng khách hàng/mạng phân tách riêng rẽ nhau về mặt định tuyến (layer 3) và chuyển mạch (layer 2). Mỗi một Tenant/VRF sẽ được đặc trưng bởi RD (Route-Distinguisher), việc học các thông tin IP/MAC được kiểm soát bởi giao thức MP-BGP, thường dựa vào các RT (Route-Target) để thực hiện. Các Tenant có thể trùng lặp về IP/VLAN/MAC.

**2.5.8 L2-VNI, L3-VNI**Layer 2 VXLAN ID (L2 VNI) là một thông số giúp các VTEP phân biệt traffic giữa các vùng layer 2 Broadcast với nhau. L2-VNI có quan hệ mật thiết và mapping 1-1 với VLAN. Khi VTEP nhận được traffic Ethernet từ thiết bị đầu cuối (Servers, Network devices,…) gửi tới một thiết bị thuộc cùng vùng Layer 2 Broadcast ở một VTEP khác,VTEP sẽ dựa vào VLAN-ID để xử lý đóng gói thêm VXLAN Header tương ứng.



**Hình 2.5: Layer 2 VXLAN ID (L2 VNI)**

Layer 3 VXLAN ID (L3 VNI) là một thông số giúp các VTEP hiểu và phân biệt các traffic cần xử lý định tuyến giữa các vùng layer 2 Broadcast thuộc các VTEP khác nhau hoặc giữa vùng layer 2 Broadcast với mạng IP truyền thống.



**Hình 2.6 : Layer 3 VXLAN ID (L3 VNI)**

**2.5.9 Distributed Gateway**

Distributed Layer 3 Gateway có nhiều ưu điểm trong mạng VXLAN/EVPN như sau:

- Phân tán các gateway tới tất cả các switch. Mỗi thiết bị đầu cuối có thể sử dụng thiết bị VTEP kết nối trực tiếp để định tuyến các traffic layer 3. Trong trường hợp đích đến của traffic thuộc cùng một thiết bị VTEP, flow dữ liệu sẽ đi trực tiếp trên cùng thiết bị VTEP đó thay vì phải đi vòng ra thiết bị chứa Gateway riêng biệt.

- Khi kết hợp với ARP suppression, nó giảm thiểu tối đa các traffic/ các bản tin flooding trong toàn mạng.

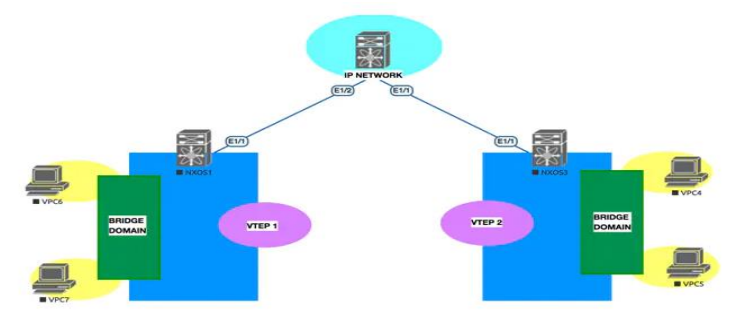
- Trong trường hợp một host cần chuyển dịch giữa các Leaf hoặc DC, host đó không cần thay đổi IP hoặc gửi ARP để tìm kiếm MAC của default gateway.

**2.5.10 Nguyên lý hoạt động của VXLAN**

VXLAN xác định Điểm cuối đường hầm VXLAN (VTEP) chứa tất cả các chức năng cần thiết để cung cấp các dịch vụ Ethernet lớp 2 cho các hệ thống đầu cuối được kết nối. VTEP được thiết kế để ở vùng biên của mạng, thường kết nối một access switch (ảo hoặc vật lý) với mạng truyền tải IP.Người ta mong đợi rằng chức năng VTEP sẽ được tích hợp vào access switch, nhưng nó tách biệt về mặt logic với access switch.

Mỗi hệ thống đầu cuối kết nối với cùng một thiết bị chuyển mạch truy cập giao tiếp thông qua access switch. Access switch hoạt động bằng cách làm “tràn” ra các cổng của nó khi nó không biết MAC đích hoặc gửi ra một cổng duy nhất khi nó đã biết hướng nào dẫn đến trạm cuối như được xác định bởi học MAC nguồn.

Lưu lượng phát sóng được gửi ra tất cả các cổng. Hơn nữa, access switch có thể hỗ trợ nhiều miền cầu nối thường được xác định là VXLAN với ID VLAN liên quan được mang trong tiêu đề 802.1Q trên cổng trung kế. Trong trường hợp chuyển đổi kích hoạt VXLAN, thay vào đó, miền cầu nối sẽ được liên kết với một ID VXLAN.



**Hình 2.7: Mô tả hoạt động VXLAN**

Mỗi VXLAN có hai giao diện: Một là cổng trung kế Bridge Domain kết nối tới switch truy cập và cổng còn lại là giao diện IP với mạng IP.

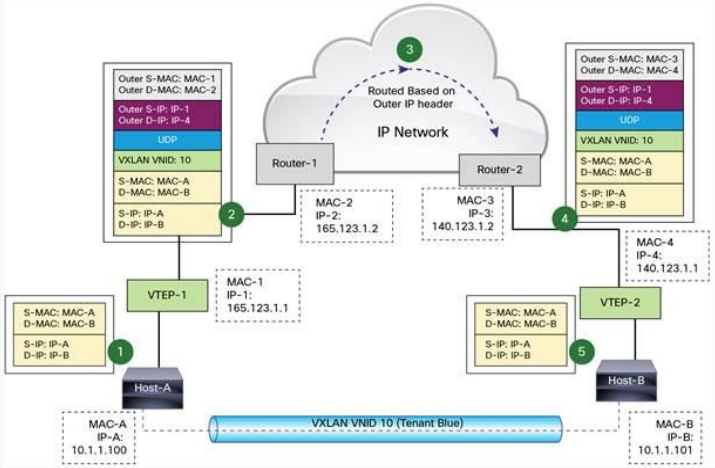
VTEP hoạt động như trong máy chủ IP của mạng IP. Nó được cấu hình với một địa chỉ IP dựa trên mạng con mà giao diện OP của nó được kết nối. VTEP sử dụng giao diện IP này để trao đổi các gói IP mang khung Ethernet được đóng gói với các VTEP khác.

VTEP cũng hoạt động như một máy chủ IP bằng cách sử dụng IGMP để tham gia nhóm phát đa hướng IP.

Ngoài một ID VXLAN được chuyển qua giao diện IP giữa VTEP, mỗi VXLAN được liên kết với một nhóm phát đa hướng IP. Nhóm phát đa hướng IP được sử dụng làm bus giao tiếp giữa mỗi VTEP để truyền các khung phát sóng, đa hướng và các khung unicast không xác định tới VTEP tham gia vào VXLAN.

**Lưu lượng gói tin VXLAN Unicast**

Giả định rằng việc học địa chỉ đã được thực hiện ở cả hai phía và ánh xạ MAC-to-VTEP tương ứng tồn tại trên cả hai VTEP. Gói tin Unicast truyền giữa hai VTEP được mô tả như trong Hình 2.7:



**Hình 2.8 :Lưu lượng gói tin VXLAN Unicast**

Bước 1: Host-A và Host-B trong VXLAN segment 10 giao tiếp với nhau thông qua VXLAN tunnel giữa VTEP-1 và VTEP-2. Ví dụ giả định rằng việc học địa chỉ đã được thực hiện ở cả hai phía và ánh xạ MAC-to-VTEP tương ứng tồn tại trên cả hai VTEP.

Bước 2: Khi Host-A gửi lưu lượng đến Host-B, nó tạo thành các khung Ethernet có địa chỉ MAC-B của Host-B làm địa chỉ MAC đích và gửi ra VTEP-1.

Bước 3: VTEP-1 với ánh xạ MAC-B to VTEP-2 trong bảng ánh xạ của nó, thực hiện đóng gói VXLAN bằng cách thêm VXLAN header, UDP header và Outer IP address header vào nó. Trong Outer IP address header, địa chỉ IP nguồn là địa chỉ IP của VTEP-1 và địa chỉ IP đích là địa chỉ IP của VTEP-2.

Bước 4: VTEP-1 sau đó thực hiện tra cứu địa chỉ IP cho địa chỉ IP của VTEP-2 để giải quyết bước nhảy tiếp theo trong mạng chuyển tiếp và sau đó sử dụng địa chỉ MAC của thiết bị hop kế tiếp để đóng gói các gói trong khung Ethernet để gửi đến thiết bị hop tiếp theo.

Bước 5: Các gói được chuyển hướng đến VTEP-2 thông qua mạng truyền tải IP dựa trên Outer IP address-header có địa chỉ IP của VTEP-2 làm địa chỉ đích. Sau khi VTEP-2 nhận được các gói, nó loại bỏ các tiêu đề Ethernet, IP, UDP và VXLAN headers và chuyển tiếp các gói tới Host-B, dựa trên địa chỉ MAC đích ban đầu trong khung Ethernet.

**ECMP và cân bằng tải LACP với VXLAN**

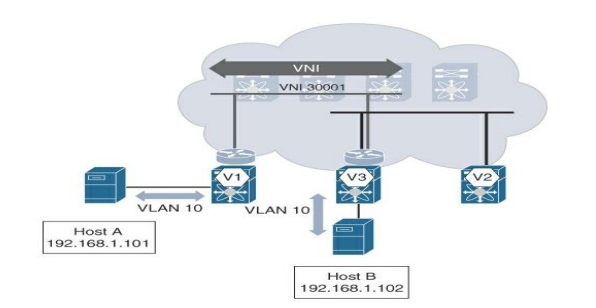
Các gói VXLAN được đóng gói và được chuyển tiếp giữa các VTEP dựa trên các chuyển tiếp gốc của mạng truyền tải. Hầu hết các mạng truyền tải trung tâm dữ liệu được thiết kế và triển khai với nhiều đường dẫn dự phòng và tận dụng các công nghệ chia sẻ tải đa luồng khác nhau để phân phối tải lưu lượng trên tất cả các đường dẫn có sẵn.

Mạng truyền tải VXLAN điển hình là mạng định tuyến IP sử dụng IP ECMP tiêu chuẩn để cân bằng tải lưu lượng giữa nhiều đường dẫn tốt nhất. Để tránh chuyển tiếp gói ngoài chuỗi, ECMP dựa trên luồng thường được triển khai. Luồng ECMP được xác định bởi các địa chỉ IP nguồn và đích và tùy chọn các cổng TCP hoặc UDP nguồn và đích trong tiêu đề gói IP.

Bởi vì tất cả các gói VXLAN chuyển tiếp giữa một cặp VTEP có cùng địa chỉ IP nguồn và đích bên ngoài và tất cả các thiết bị VTEP phải sử dụng một cổng UDP đích giống hệt nhau có thể là cổng UDP được phân bổ theo số (IANA) 4789 hoặc một cổng được cấu hình bởi khách hàng, thành phần biến duy nhất trong định nghĩa luồng ECMP có thể phân biệt các luồng VXLAN với quan điểm của mạng truyền tải là cổng UDP nguồn. Tình huống tương tự đối với giao thức điều khiển liên kết (LACP) xảy ra nếu giao diện đầu ra được giải quyết dựa trên quyết định định tuyến và ECMP là kênh cổng LACP. LACP sử dụng tiêu đề gói ngoài VXLAN để băm chia sẻ tải liên kết, dẫn đến cổng UDP nguồn là thành phần duy nhất có thể xác định duy nhất một luồng VXLAN.

**2.6 Flood and Learn VXLAN**

Thông qua cơ chế này, lưu lượng truy cập đa điểm bị “tràn” trên VXLAN giữa các VTEP để tìm hiểu về các máy chủ MAC nằm phía sau các VTEP để lưu lượng truy cập tiếp theo có thể được unicast. Đây được gọi là cơ chế Flood & Learn. Với điều này, giao tiếp trong VXLAN VNI tuân theo ngữ nghĩa chuyển tiếp Lớp 2 thông thường, giống như trường hợp trong một VLAN (xem Hình 2.8). Sự khác biệt chính là Miền VXLAN VNI trải dài các mạng Lớp 3 trên lớp phủ IP.



**Hình 2.9: Food and Learn VXLAN**

Nhiều VNI có thể chia sẻ cùng một nhóm phát đa hướng. Điều này là do có 16 triệu các tùy chọn VNI tiềm năng và có ánh xạ 1: 1 duy nhất giữa VNI và đa hướng nhóm yêu cầu một số lượng tài nguyên phần mềm và phần cứng quá lớn, khiến nó không thực tế để triển khai. Trên thực tế, trong hầu hết các triển khai thực tế, tối đa là 512 hoặc 1024 nhóm đa hướng được hỗ trợ. Điều này có nghĩa là VTEP có thể nhận được lưu lượng truy cập đa điểm đến được đóng gói VXLAN cho một VNI không được cấu hình cục bộ. Cô lập tại cấp độ mạng vẫn được duy trì, bởi vì mọi phân đoạn mạng VXLAN đều có VNI duy nhất và khóa tra cứu luôn là [VNI, DMAC]. Một giải pháp thay thế cho IP multicast để xử lý lưu lượng truy cập đa đích trong môi trường VXLAN là sử dụng tính năng nhập sao chép (IR), hoặc sao chép đầu-cuối. Với IR, mọi VTEP phải biết về các VTEP có tư cách thành viên trong một VNI nhất định.

VTEP nguồn tạo ra n bản sao cho mọi khung nhiều người tham gia, với từng mục tiêu dành cho các VTEP khác có tư cách thành viên trong VNI tương ứng. Điều này tạo thêm gánh nặng cho VTEP, nhưng nó có lợi ích của việc đơn giản hóa vì không cần phải chạy đa hướng trong lớp dưới IP.

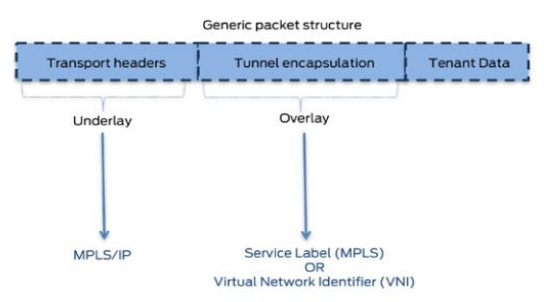
VXLAN là một trong số ít các lớp phủ được sử dụng làm lớp phủ máy chủ và lớp phủ mạng. Với sự phát triển của SDN, VXLAN đã được sử dụng nhiều như một máy chủ lưu trữ lớp phủ, với máy chủ vật lý lưu trữ phần mềm VTEP thực hiện VXLAN đóng gói và giải mã cho lưu lượng giữa các khối lượng công việc được ảo hóa. Thành lập giao tiếp giữa các khối lượng công việc cũ không hỗ trợ VXLAN, Lớp 2 Cổng VXLAN đi đầu, cung cấp khả năng dịch giữa các miền dựa trên VXLAN cũ và dựa trên 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 như một mạng lớp phủ. Điều này đã mở đường cho một môi trường kết hợp, nơi cả phần mềm và VTEP phần cứng có thể được hỗ trợ đồng thời trong cùng một môi trường trung tâm dữ liệu.

**2.7 Underlay và Overlay Network**

Underlay network là thuật ngữ chỉ hạ tầng cung cấp kết nối nền tảng của mạng. Trong mô hình Spine-Leaf, thuật ngữ Underlay thường được sử dụng để chỉ việc định tuyến Layer 3 (IP) giữa các thiết bị với nhau. Mục tiêu của Underlay là cung cấp kết nối any-to-any giữa tất cả các thiết bị. Nhờ đó, các thiết bị có thể truyền các giao thức báo hiệu sử dụng cho mạng Overlay.

Các giao thức Underlay phổ biến trong mạng DC EVPN/VXLAN thường thấy là eBGP hoặc OSPF.

Overlay network là thuật ngữ chỉ cách thức đóng gói traffic và truyền nó qua môi trường mạng nền tảng (Underlay). Giao thức phổ biến trong môi trường mạng DC là VXLAN. Giao thức này sẽ đóng gói các frame Ethernet Layer 2 vào trong packet UDP Layer 3 để tạo các kết nối mạng Layer 2 giữa môi trường mạng Layer 3.



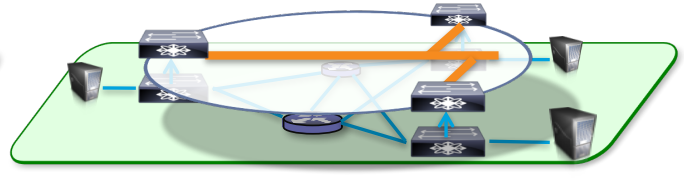
**Hình 2.10: Giao thức triển khai trên mạng Underlay và Overlay**

**2.7.1 Overlay Network**

**2.7.1.1 Giới thiệu về Overlay Network**

Các kết cấu trung tâm dữ liệu ảo hóa hiện đại phải đáp ứng các yêu cầu nhất định để tăng tốc triển khai ứng dụng và hỗ trợ các nhu cầu của DevOps. Ví dụ: các phân vùng mạng cần hỗ trợ mở rộng bảng chuyển tiếp, mở rộng phân đoạn mạng, mở rộng phân đoạn lớp 2, tính di động của thiết bị ảo, tối ưu hóa đường dẫn chuyển tiếp và mạng ảo hóa để hỗ trợ nhiều đối tượng trên cơ sở hạ tầng vật lý được chia sẻ.

Mặc dù khái niệm về lớp phủ mạng không phải là mới, nhưng sự quan tâm đến lớp phủ mạng đã tăng lên trong vài năm qua vì tiềm năng của chúng để giải quyết một số yêu cầu này. Mối quan tâm đến mạng lớp phủ cũng tăng lên với sự ra đời của các định dạng khung đóng gói mới được xây dựng đặc biệt cho trung tâm dữ liệu. Các định dạng này bao gồm mạng LAN ảo có thể mở rộng (VXLAN), Ảo hóa mạng sử dụng đóng gói định tuyến chung (NVGRE), Kết nối liên kết trong suốt của rất nhiều liên kết (TRILL) và Giao thức phân tách định vị / vị trí (LISP). Lớp phủ mạng là mạng ảo gồm các nút được kết nối với nhau chia sẻ mạng vật lý bên dưới, cho phép triển khai các ứng dụng yêu cầu cấu trúc liên kết mạng cụ thể mà không cần sửa đổi mạng bên dưới .



**Hình 2.11: Overlay Network**

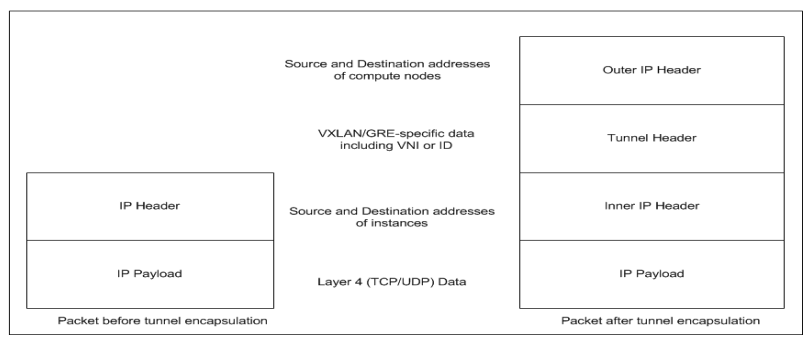
**Lợi ích của lớp phủ ảo hóa mạng bao gồm:**

* Các chức năng của thiết bị được tối ưu hóa: Mạng lớp phủ cho phép phân tách (và chuyên môn hóa) các chức năng của thiết bị dựa trên vị trí thiết bị đang được sử dụng trong mạng. Thiết bị biên hoặc thiết bị dạng Leaf có thể tối ưu hóa các chức năng và tất cả các giao thức liên quan của nó dựa trên thông tin và quy mô trạng thái cuối và thiết bị lõi hoặc thiết bị lõi có thể tối ưu hóa các chức năng và giao thức của nó dựa trên cập nhật trạng thái liên kết, tối ưu hóa với sự hội tụ nhanh chóng.
* Khả năng mở rộng và tính linh hoạt của kết cấu: Công nghệ lớp phủ cho phép mạng mở rộng quy mô bằng cách tập trung mở rộng quy mô vào các thiết bị biên của lớp phủ mạng. Với các lớp phủ được sử dụng ở biên của phân vùng mạng, các thiết bị lõi và thiết bị Spine được giải phóng khỏi nhu cầu thêm thông tin máy chủ lưu trữ cuối vào bảng chuyển tiếp của chúng.
* Định địa chỉ lớp phủ: Hầu hết các công nghệ lớp phủ được sử dụng trong trung tâm dữ liệu cho phép ID mạng ảo có phạm vi duy nhất và xác định các mạng riêng lẻ. Phạm vi này cho phép chồng chéo tiềm năng trong địa chỉ MAC và IP giữa các bên thuê. Việc đóng gói lớp phủ cũng cho phép không gian địa chỉ cơ sở hạ tầng bên dưới được quản lý tách biệt với không gian địa chỉ đối tượng thuê.

**2.7.1.2 Đặc điểm Overlay**

Các công nghệ như GRE hay VXLAN là khi một mạng overlay tạo nên thì sẽ có một ID duy nhất để định danh cho mạng đó đồng thời được sử dụng để đóng gói lưu lượng (traffic encapsulation). Mỗi gói tin giữa các instances trên các host vật lý khác nhau được đóng gói trên một host và gửi tới các host khác thông qua point-to-point GRE hoặc VXLAN tunnel. Khi gói tin tới host đích, các tunnel header sẽ bị loại bỏ (tại tunnel endpoint) và gói tin được chuyển tiếp tới bridge kết nối với instances.

Sơ đồ sau đây thể hiện gói tin trước và sau khi được đóng gói bởi host.



**Hình 2.12: Gói tin trước và sau khi được đóng gói bởi host**

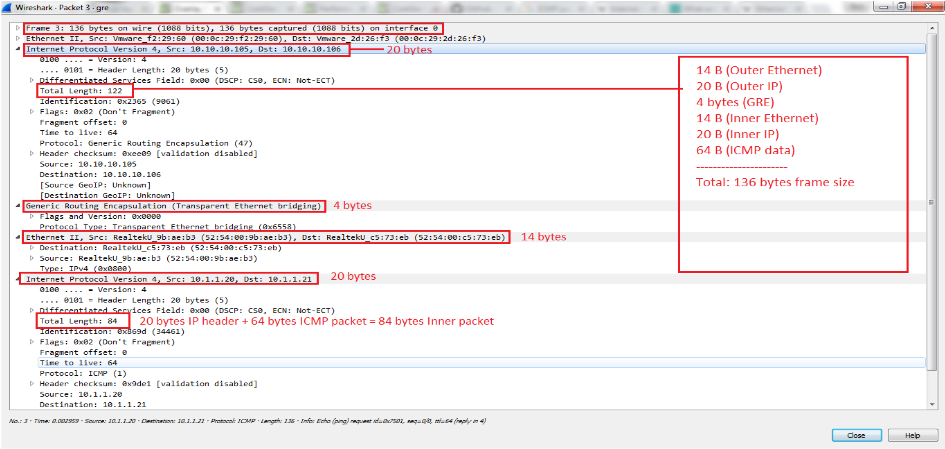
Theo hình mô tả ở trên, địa chỉ nguồn và đích trong outer IP header sẽ định danh cho endpoint của tunnel. Tunnel endpoints bao gồm compute nodes và bất kì host nào chạy các dịch vụ L3 và DHCP, có thể là controller node hoặc network node. Các địa chỉ nguồn và đích trong Inner IP header định danh cho các instances gửi và nhận payload.

**2.7.2 Một số công nghệ Overlay Network**

**2.7.2.1 GRE (Generic Routing Encapsulation)**

GRE-Generic Routing Encapsulation là giao thức được phát triển đầu tiên bởi Cisco. Giao thức này sẽ đóng gói một số kiểu gói tin vào bên trong các IP tunnels để tạo thành các kết nối điểm-điểm (point-to-point) ảo. Các IP tunnel chạy trên hạ tầng mạng công cộng. Trong đó trường Key (chiếm 4 octects tương đương 32 bits) trong GRE header sử dụng để mang Tenant Network Identifier (TNI - định danh mạng khách hàng) và được sử dụng để cô lập các logical segment khác nhau.

GRE đóng gói inner frame sử dụng IP protocol số 47 để truyền thông chứ không sử dụng TCP hay UDP.

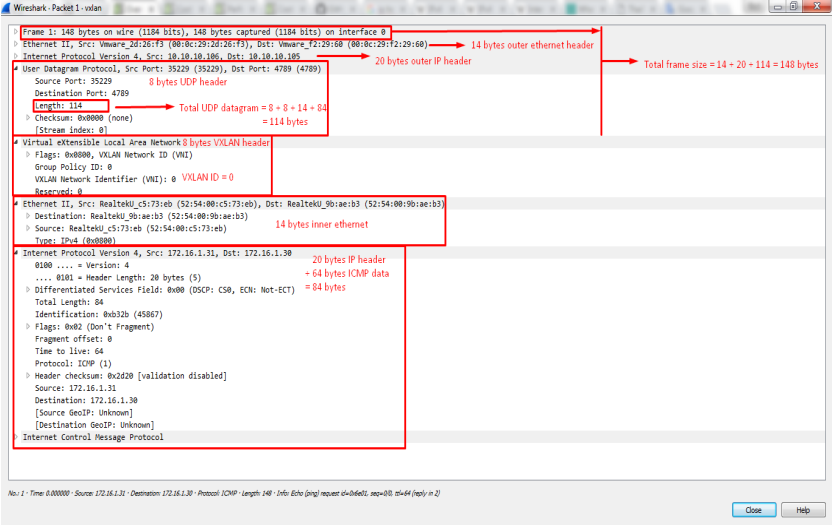


**Hình 2.13: Gói tin được đóng gói bởi GRE**

**2.7.2.2 VXLAN**

VXLAN hay Virtual Extensible LAN (VXLAN) là giao thức sử dụng UDP (cổng 4789) để truyền thông và một segment ID độ dài 24 bit còn gọi là VXLAN Network Identifier (VNID).

Với VXLAN, hash của header trong inner frame được sử dụng để làm cổng nguồn của UDP. Kết quả là, một VXLAN flow sẽ là duy nhất với các địa chỉ IP và UDP ports kết nối với nhau trong IP header khi duyệt qua mạng vật lý bên dưới.



**Hình 2.14: Gói tin được đóng gói bởi VXLAN**

**2.7.3 So sánh Underlay Network và Overlay Network**

Underlay Network là thuật ngữ chỉ hạ tầng cung cấp kết nối nền tảng của mạng. Trong mô hình Clos, thuật ngữ Underlay thường được sử dụng để chỉ việc định tuyến Layer 3 (IP) giữa các thiết bị với nhau. Mục tiêu của Underlay là cung cấp kết nối any-to-any giữa tất cả các thiết bị. Nhờ đó, các thiết bị có thể truyền các giao thức báo hiệu sử dụng cho mạng Overlay.

Overlay Network là thuật ngữ chỉ cách thức đóng gói traffic và truyền nó qua môi trường mạng hạ tầng (Underlay).

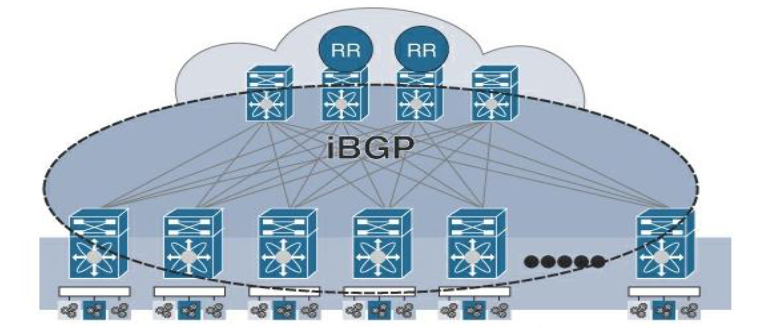
* **Bảng So sánh Underlay Network và Overlay Network**

|  |  |  |
| --- | --- | --- |
| **Tham số** | **Underlay Network** | **Overlay Network** |
| Định nghĩa | Mạng Underlay là cơ sở hạ tầng vật lý trên mạng Overlay được xây dựng sẵn. | Một mạng Overlay là một mạng ảo được xây dựng nằm phía trên của một cơ sở hạ tầng mạng (các Underlay). |
| Các giao thức liên quan | Chuyển mạch Ethernet, VLAN định tuyến. | VXLAN, OTV, VPLS… |
| Khả năng mở rộng | Ít khả năng mở rộng do giới hạn công nghệ. | Được thiết kế để cung cấp khả năng mở rộng hơn Underlay Network.Ví dụ: VLAN (Underlay Network) cung cấp 4094 VLAN trong khi VXLAN (Overlay Network) cung cấp số nhận dạng tối đa 16.000.000. |
| Điểu khiển gói tin | Do phần cứng. | Do phần mềm. |
| Chuyển gói tin | Chịu trách nhiệm phân phối các gói. | Giảm tải từ phân phối các gói tin. |
| Đóng gói tin | Gửi gói xảy ra ở lớp 3 và lớp 4. | Đóng gói bản tin suốt từ nguồn và đích đến, do đó có thêm tiêu đề. |
| Lưu lượng chuyển tiếp đa đường | Ít khả năng mở rộng các tùy chọn của đa trợ đa đường dẫn chuyển tiếp. | Hỗ trợ đa đường dẫn chuyển tiếp trong ảo hóa. |
| Thời gian triển khai | Ít khả năng mở rộng và tốn thời gian hoạt động để thiết lập các dịch vụ và chức năng mới. | Khả năng triển khai nhanh chóng và dễ mở rộng. |
| Lưu lượng truyền | Truyền các gói đi qua mạng thiết bị như bộ chuyển mạch và bộ định tuyến. | Truyền gói dọc theo các liên kết ảo giữa các nút lớp phủ. |

**2.8 Giao thức VXLAN BGP EVPN**

**2.8.1 Giới thiệu về BGP.**

Trong BGP, khi sử dụng một hệ thống tự trị (AS), sự ngang hàng giữa các speak BGP được gọi là BGP nội bộ (iBGP). iBGP chủ yếu được sử dụng để trao đổi thông tin giữa tất cả các speak theo kiểu đồng bộ trên toàn bộ AS. iBGP yêu cầu mạng ngang hàng đầy đủ vì thông tin khả năng tiếp cận phải được học từ một speak iBGP và không được chuyển tiếp đến bất kỳ speak iBGP nào khác. Để thay đổi hành vi này, iBGP có thể sử dụng chức năng của một bộ phản xạ tuyến đường (RR). RR được phép nhận iBGP thông tin và sau đó "phản ánh" thông tin cho bất kỳ khách hàng RR nào của nó - nghĩa là tất cả hàng xóm iBGP được biết đến. Với chức năng iBGP và RR, BGP peerings có thể được tối ưu hóa để yêu cầu sử dụng đồng đẳng BGP toàn lưới có thể bị bỏ. Tương tự, điều này cũng cho phép mặt phẳng điều khiển được mở rộng quy mô cao hơn.



**Hình 2.15: VXLAN BGP EVPN**

Với BGP bên ngoài (eBGP), việc ngang hàng giữa các BGP speak đạt được giữa các hệ thống tự trị khác nhau. Ví dụ, điều này có nghĩa là một BGP speak trong AS 65500 có thể được xếp ngang hàng với BGP speak hàng xóm của nó, nằm trong AS 65501.Trong eBGP, các quy tắc trao đổi tuyến đường hơi khác một chút.Không giống như với iBGP, yêu cầu mạng ngang hàng đầy đủ mà không sử dụng RR, khả năng truy cập BGP thông tin nhận được từ một eBGP speak luôn được gửi đến tất cả các eBGP láng giềng của nó. Ngoài ra, so với iBGP, eBGP sẽ tự cập nhật next-hop-self khi nó đi qua một AS còn iBGP thì phải cấu hình thủ công.

BGP đóng một vài vai trò quan trọng trong việc thực hiện các yêu cầu mở rộng quy mô EVPN. Trong số đó có:

* Loại bỏ việc học địa chỉ MAC khỏi mặt phẳng dữ liệu. Trong mạng Ethernet tiêu chuẩn, địa chỉ MAC được học bằng cách chuyển mạch quan sát các khung (gói) khi chúng được truyền đi. Bất cứ khi nào một bộ chuyển mạch gặp một địa chỉ MAC đích không xác định, nó sẽ làm “tràn” khung, tức là sao chép nó đến tất cả các điểm đến có thể. Loại flood khung hình unicast không xác định này có thể dẫn đến tắc nghẽn giao thông, có khả năng bao gồm các cơn bão phát sóng và quy mô rất kém.
* Giảm tải lưu lượng truyền phát và phát đa hướng. Tương tự như flood unicast không xác định, lưu lượng phát sóng và đa hướng có thể tạo ra tải trọng lớn trên mạng - và khả năng mất mạng do bão phát sóng. BGP EVPN có thể giảm bớt những vấn đề này bằng cách chuyển tiếp lưu lượng truy cập đó một cách chọn lọc.
* Cho phép chuyển tiếp, cân bằng tải và hội tụ tối ưu trên mạng IP bên dưới.

**2.8.2 MP-BGP EVPN**

Multiprotocol Border Gateway Protocol (MP-BGP) Ethernet Virtual Private Network (EVPN) là giao thức điều khiển phù hợp và được khuyên dùng cho VXLAN. MP-BGP EVPN được định nghĩa bởi IETF như một chuẩn chung cung cấp cơ chế cho phép thiết bị VTEP tìm kiếm và học các thông tin từ các host, server, network devices của các VTEP khác.

Sử dụng băng thông hiệu quả và khả năng phục hồi với tính năng đa hệ số Active-Active. VXLAN được hỗ trợ với PortChannel ảo (vPC). Điều này cho phép khả năng phục hồi trong kết nối cho các máy chủ được gắn với các thiết bị chuyển mạch truy cập với việc sử dụng hiệu quả băng thông khả dụng.

Hoạt động của BGP EVPN Vxlan: EVPN sử dụng BGP để trao đổi thông tin địa chỉ IP và MAC của điểm cuối giữa các VTEP. Khi một máy chủ gửi một gói đến một điểm cuối, bộ chuyển mạch sẽ tìm kiếm bảng định tuyến để tìm một kết quả phù hợp. Nếu nó tìm thấy một kết quả phù hợp tồn tại đằng sau một VTEP khác, thì gói tin sẽ được đóng gói bằng các tiêu đề VXLAN, UDP và được đóng gói lại với các tiêu đề IP và Ethernet bên ngoài để vận chuyển qua hệ thống mạng Spine-Leaf. Khi gói tin đến VTEP đích, các tiêu đề Ethernet, IP, UDP và VXLAN bên ngoài được loại bỏ và bộ chuyển mạch sẽ gửi gói ban đầu đến điểm cuối.

Đối với VXLAN, các gói tin BUM (Broadcast, Unknow unicast, Multicast) vẫn quảng bá tự do như trong các Broadcast domain và có nghĩa việc học địa chỉ MAC đối với VXLAN vẫn như đối với VLAN. Một hạn chế khác là việc cấu hình VXLAN tunnel (VTEP-VXLAN Tunnel End Point) được thực hiện bằng tay nên rất khó quản lý nếu số lượng Leaf lớn.

Nói một cách khác VXLAN chỉ được xem như là một Data-plane dùng để vận chuyển dữ liệu thuần túy, điều còn thiếu cho giải pháp này là một giao thức đóng vai trò Control-plane nhằm tự động thiết lập các Tunnel và tự trao đổi MAC học được cho nhau mà không cần chờ gói tin ARP broadcast.

Lúc này giao thức MP-BGP (Multi protocol BGP) được áp dụng như cách thức nó đã được áp dụng thành công trong các hệ thống mạng MPLS. Vì thế giải pháp này còn được gọi là VXLAN BGP EVPN và có thể nói đây là giải pháp triệt để nhất hiện nay áp dụng cho những hệ thống mạng hoặc Data Center lớn hoặc nằm phân tán.

Control Plane MP-BGP EVPN cung cấp những lợi ích chính sau:

* Giao thức MP-BGP EVPN dựa trên các tiêu chuẩn công nghiệp, cho phép khả năng tương tác đa động cơ.
* Nó cho phép học trên bình diện điều khiển thông tin khả năng tiếp cận Lớp 2 và Lớp 3 của máy chủ lưu trữ cuối, cho phép các tổ chức xây dựng mạng lớp phủ VXLAN có khả năng mở rộng và mạnh mẽ hơn.
* Nó sử dụng công nghệ MP-BGP EVPN hàng thập kỷ để hỗ trợ các mạng lớp phủ VXLAN có thể mở rộng.
* Họ địa chỉ EVPN mang cả thông tin khả năng truy cập Lớp-2 và Lớp-3, do đó cung cấp kết nối và định tuyến tích hợp trong mạng lớp phủ VXLAN.
* Nó giảm thiểu tình trạng “flood” mạng thông qua phân phối tuyến MAC / IP máy chủ và ngăn chặn Giao thức phân giải địa chỉ (ARP) trên các VTEP cục bộ.
* Nó cung cấp chuyển tiếp tối ưu cho lưu lượng East-West (dữ liệu trao đổi giữa các máy chủ) và North-South (dữ liệu từ client đến server) và hỗ trợ tính di động của khối lượng công việc với chức năng anycast phân tán.
* Nó cung cấp khả năng phát hiện và xác thực ngang hàng VTEP, giảm thiểu rủi ro về các VTEP giả mạo trong mạng lớp phủ VXLAN.
* Nó cung cấp các cơ chế để xây dựng đa hiệu năng tích cực ở Lớp 2.

**VTEP chạy MP-BGP EVPN** :VTEP chạy MP-BGP EVPN cần hỗ trợ cả chức năng mặt phẳng điều khiển và mặt phẳng dữ liệu. Trong mặt phẳng điều khiển, họ bắt đầu các tuyến MP-BGP EVPN để quảng bá máy chủ cục bộ của họ. Họ nhận các bản cập nhật MP-BGP EVPN từ các đồng nghiệp của họ và cài đặt các tuyến EVPN trong bảng chuyển tiếp của họ. Đối với chuyển tiếp dữ liệu, chúng đóng gói lưu lượng người dùng trong VXLAN và gửi nó qua mạng Underlay IP. Theo hướng ngược lại, họ nhận được lưu lượng được đóng gói VXLAN từ các VTEP khác, giải mã nó và chuyển tiếp lưu lượng với tính năng đóng gói Ethernet gốc về phía máy chủ.

Nền tảng chuyển mạch chính xác cần được chọn cho các vai trò mạng khác nhau. Đối với các thiết bị truyền tải IP, phần mềm cần hỗ trợ mặt phẳng điều khiển MP-EVPN, nhưng phần cứng không cần hỗ trợ các chức năng mặt phẳng dữ liệu VXLAN. Đối với VTEP, cần hỗ trợ cả chức năng mặt phẳng điều khiển và mặt phẳng dữ liệu.

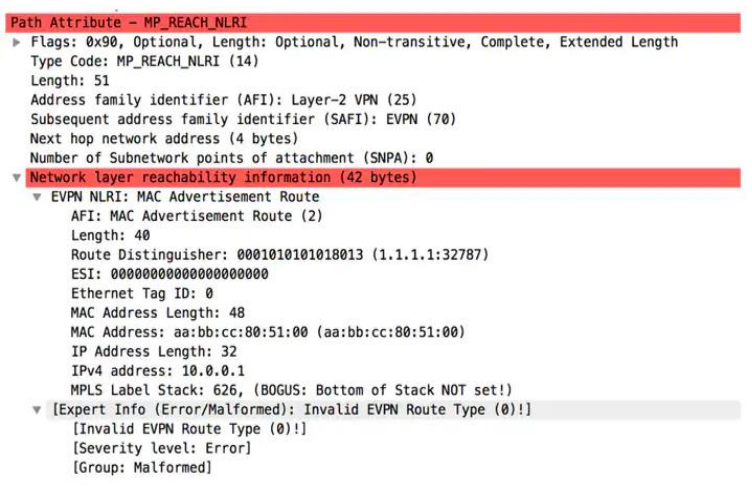
**MP-BGP EVPN NLRI và họ địa chỉ L2VPN EVPN**

Giống như các giao thức điều khiển định tuyến mạng khác, MP-BGP EVPN được thiết kế để phân phối thông tin khả năng tiếp cận lớp mạng (NLRI) cho mạng. Một tính năng độc đáo của EVPN NLRI là nó bao gồm cả thông tin khả năng truy cập Lớp-2 và Lớp-3 cho các máy chủ cuối nằm trong mạng lớp phủ EVPN VXLAN. Nói cách khác, nó quảng cáo cả địa chỉ MAC và IP của máy chủ cuối EVPN VXLAN. Khả năng này tạo cơ sở cho hỗ trợ định tuyến và bắc cầu tích hợp VXLAN.

Địa chỉ MAC Lớp-2 cần được phân phối vì VXLAN là công nghệ mở rộng Lớp 2. Không giống như một VLAN truyền thống, được giới hạn ở một vị trí cụ thể trong mạng và vẫn nằm trong ranh giới Lớp 2 và Lớp 3, VNI là một phân đoạn Lớp 2 ảo trong mạng Overlay. Tuy nhiên, từ quan điểm của mạng Underlay, nó có thể trải dài nhiều trang web không liền nhau, vượt ra ngoài ranh giới Lớp-2 và Lớp-3 của cơ sở hạ tầng mạng Underlay. Lưu lượng giữa các máy chủ cuối trong cùng một VNI cần được bắc cầu trong mạng Overlay, có nghĩa là các thiết bị VTEP trong một VNI nhất định cần biết về các địa chỉ MAC khác của máy chủ cuối trong VNI này. Việc phân phối địa chỉ MAC thông qua BGP EVPN cho phép giảm hoặc loại bỏ tình trạng “flood” unicast không xác định trong VXLAN.

Địa chỉ IP máy chủ lưu trữ lớp-3 được quảng bá thông qua MP-BGP EVPN để lưu lượng truy cập giữa các VXLAN có thể được định tuyến đến máy chủ lưu trữ cuối đích thông qua một đường dẫn tối ưu. Đối với lưu lượng giữa các VXLAN cần được định tuyến đến máy chủ cuối đích, định tuyến IP dựa trên máy chủ có thể cung cấp đường dẫn chuyển tiếp tối ưu đến vị trí chính xác của máy chủ đích.

MP-BGP EVPN cũng có thể quảng bá các tuyến tiền tố mạng con IP của các VNI. Các tuyến tiền tố có thể được sử dụng để định tuyến lưu lượng truy cập đến máy chủ đích khi các tuyến IP máy chủ bị thiếu: ví dụ: khi các tuyến IP máy chủ chưa được VTEP học thông qua MP-BGP. VTEP cũng có thể quảng bá các tuyến tiền tố ra bên ngoài mạng VXLAN nếu các mạng con cần được định tuyến và được biết đến bên ngoài mạng VXLAN.



**Hình 2.16: EVPN NLRI**

EVPN NLRI được mang trong BGP bằng cách sử dụng phần mở rộng đa giao thức BGP với họ địa chỉ mới được gọi là Layer-2 VPN (L2VPN) EVPN. Tương tự như họ địa chỉ VPNv4 trong IP VPN dựa trên BGP MPLS (RFC 4364), họ địa chỉ L2VPN EVPN cho EVPN sử dụng bộ phân biệt tuyến (RD) để duy trì tính duy nhất giữa các tuyến giống nhau trong các trường hợp VRF khác nhau và sử dụng mục tiêu tuyến ( RTs) để xác định các chính sách xác định cách các tuyến đường được quảng bá và chia sẻ bởi các phiên bản VRF khác nhau.

**Cổng phân phối IP Anycast**

Trong MP-BGP EVPN, mọi VTEP trong VNI có thể là cổng anycast phân tán cho các máy chủ cuối trong mạng con IP của nó bằng cách hỗ trợ cùng một địa chỉ IP cổng ảo và địa chỉ MAC của cổng ảo. Với chức năng cổng anycast trong EVPN, các máy chủ cuối trong VNI có thể sử dụng VTEP cục bộ của chúng cho VNI này làm cổng mặc định để gửi lưu lượng ra khỏi mạng con IP của chúng. Khả năng này cho phép tối ưu lưu lượng truy cập từ các máy chủ cuối trong mạng Overlay VXLAN.

Một cổng anycast phân tán cũng cung cấp tính di động của máy chủ trong suốt mạng Overlay VXLAN. Do địa chỉ IP cổng ảo và địa chỉ MAC ảo được cung cấp giống hệt nhau trên tất cả các VTEP trong cùng một VNI nên khi máy chủ cuối chuyển lưu lượng từ VTEP này sang VTEP khác, nó không cần phải gửi một yêu cầu ARP để tìm lại địa chỉ MAC của cổng.

Nói chung Distributed Anycast Layer 3 Gateway có nhiều ưu điểm trong mạng VXLAN/EVPN như sau:

- Phân tán các gateway tới tất cả các switch. Mỗi thiết bị đầu cuối có thể sử dụng thiết bị VTEP kết nối trực tiếp để định tuyến các traffic layer 3. Trong trường hợp đích đến của traffic thuộc cùng một thiết bị VTEP, flow dữ liệu sẽ đi trực tiếp trên cùng thiết bị VTEP đó thay vì phải đi vòng ra thiết bị chứa Gateway riêng biệt.

- Khi kết hợp với ARP suppression, nó giảm thiểu tối đa các traffic/ các bản tin flooding trong toàn mạng.

- Trong trường hợp một host cần chuyển dịch giữa các Leaf hoặc các DC, host đó không cần thay đổi IP hoặc gửi ARP để tìm kiếm MAC của Default gateway.

**2.8.3 Sự phát triển của BGP EVPN VXLAN**

Theo truyền thống, VLAN là phương pháp tiêu chuẩn để cung cấp phân đoạn mạng trong các mạng cơ sở. Các VLAN sử dụng các kỹ thuật ngăn chặn vòng lặp như giao thức Spanning-Tree (STP), áp đặt các hạn chế về thiết kế mạng và khả năng phục hồi. Hơn nữa, do giới hạn về số lượng VLAN có thể được sử dụng để giải quyết các phân đoạn lớp 2 (4094 VLAN), VLAN là một yếu tố hạn chế đối với các bộ phận CNTT và các nhà cung cấp đám mây-những người xây dựng mạng khuôn viên lớn và phức tạp.

VXLAN được thiết kế để khắc phục những hạn chế cố hữu của VLAN và STP. Đây là tiêu chuẩn IETF được đề xuất [RFC 7348] để cung cấp các dịch vụ mạng Ethernet Lớp 2 tương tự như các VLAN, nhưng có tính linh hoạt cao hơn. Về mặt chức năng, nó là một giao thức đóng gói MAC-in-UDP chạy như một lớp phủ ảo trên mạng Lớp 3 hiện có.

Tuy nhiên, bản thân VXLAN không cung cấp khả năng chuyển mạch và định tuyến tối ưu trong mạng, bởi vì cơ chế “Flood and Learn” mà nó sử dụng làm hạn chế khả năng mở rộng của nó (đối với một máy chủ có thể truy cập được, thông tin của máy chủ bị “tràn ngập” trên mạng). Lớp phủ VXLAN, yêu cầu:

* Một mạng truyền tải cơ bản thực hiện chuyển tiếp mặt phẳng dữ liệu để liên lạc unicast giữa các điểm cuối được kết nối với kết cấu.
* Một mặt phẳng điều khiển có khả năng phân phối thông tin về khả năng truy cập máy chủ Lớp 2 và Lớp 3 trên toàn mạng.

Để đáp ứng các yêu cầu bổ sung này, đề xuất MP-BGP, có tính năng Thông tin khả năng tiếp cận lớp mạng (NLRI), mang cả MAC và thông tin IP cùng lúc. Với thông tin MAC và IP có sẵn cùng nhau cho các quyết định chuyển tiếp làm cho việc định tuyến và chuyển mạch trong mạng được tối ưu hóa. Điều này cũng giảm thiểu việc sử dụng cơ chế flood thông thường, hạn chế khả năng đóng cặn của fabric VXLAN. Phần mở rộng cho phép BGP vận chuyển thông tin IP Lớp 2 và Lớp 3 là EVPN.

**2.9 Kết luận chương 2**

Overlay networking là công nghệ cho phép tạo ra các mạng ảo trên hệ thống mạng vật lý bên dưới (Underlay network) mà không ảnh hưởng hoặc ảnh hưởng không đáng kể tới hạ tầng mạng bên dưới. Chương 2 đã đưa ra chi tiết các thành phần trong mạng VXLAN cũng như các giao thức và nguyên lý hoạt động của nó. Việc VXLAN kết hợp với giao thức điều khiển BGP-EVPN là điều khuyên dùng trong mạng trung tâm dữ liệu hiện nay, đem lại rất nhiều lợi ích cho các nhà cung cấp dịch vụ trong việc triển khai và mở rộng trung tâm dữ liệu.

**CHƯƠNG 3:XÂY DỰNG MÔ HÌNH SPINE-LEAF DEMO TRÊN PHẦN MỀM MÔ PHỎNG**

**3.1 Topology**



|  |  |  |  |
| --- | --- | --- | --- |
| **Thiết bị** | **Interface** | **IP address** |  |
| Spine-1 | Ethernet 1 | 10.10.1.1/24 | AS 100 |
| Ethernet 2 | 10.10.2.1/24 |
| Ethernet 3 | 10.10.3.1/24 |
| Spine-2 | Ethernet 1 | 20.20.1.1/24 | AS 100 |
| Ethernet 2 | 20.20.2.1/24 |
| Ethernet 3 | 20.20.3.1/24 |
| Leaf-1 | Ethernet 1 | 10.10.1.2/24 | AS 201 |
| Ethernet 2 | 20.20.1.2/24 |
| Loopback 1 | 101.1.1.1/24 |
| Leaf-2 | Ethernet 1 | 10.10.2.2/24 | AS 202 |
| Ethernet 2 | 20.20.2.2/24 |
| Loopback 1 | 102.1.1.1/24 |
| Leaf-3 | Ethernet 1 | 10.10.3.2/24 | AS 203 |
| Ethernet 2 | 20.20.3.2/24 |
| Loopback 1 | 103.1.1.1/24 |
| VPC 6 | Gateway 31.1.1.254 | 31.1.1.1/24 | VLAN 31  VNI 3311 |
| VPC 8 | 31.1.1.2/24 | VLAN 31  VNI 3311 |
| VPC 7 | Gateway 41.1.1.254 | 41.1.1.1/24 | VLAN 41  VNI 4411 |
| VPC 9 | 41.1.1.2/24 | VLAN 41  VNI 4411 |
| VPC 10 | Gateway 51.1.1.254 | 51.1.1.1/24 | VLAN 51 |
| VPC 11 | 51.1.1.2/24 | VLAN 51 |

**3.2 Lời nói đầu**

Bài lab bao gồm 5 switch: 2 Spine và 3 Leaf. Leaf-1, 2 ta tạo Vlan 31, 41. Còn Leaf-3 tạo Vlan 51. Công nghệ Vxlan sẽ mở rộng mạng lớp 2 qua mạng lớp 3. Bài lab trình bày mạng lớp dưới BGP và thử nghiệm khả năng dự phòng cao trong trường hợp có một Spine bị lỗi hoặc lỗi bất kỳ liên kết nào.

Tiếp theo trình bày cách định cấu hình VXLAN trong nhiều thiết bị chuyển mạch Leaf, cấu hình VNI, gán cổng cho VLAN, định cấu hình peer VTEPS cho host và mở rộng Lớp 2 qua lớp 3 trong các thiết bị Leaf.

Cuối cùng trình bày cách định cấu hình cổng VXLAN và định tuyến giữa lớp Underlay và lớp Overlay. Hiển thị cấu hình định tuyến VXLAN trực tiếp bằng câu lệnh "ip add virtual". Tạo một VLAN underlay 51 và xác thực giao tiếp.

**3.3 Mục đích bài lab**

Mục đích của bài lab này là cấu hình Vxlan trên các switch Leaf-1, 2, 3 với giao thức BGP và ECMP. Hiểu rõ hơn về một số ưu điểm của mô hình Spine-Leaf như: Tính dự phòng cao, không sử dụng STP, mở rộng mạng lớp 2 qua mạng lớp 3.

Trong bài lab này em sẽ cấu hình trên con Switch Arista và được thiết kế trên phần mềm mô phỏng EVE-NG.

**3.4 Các bước cấu hình và show kết quả**

**Step 1: Cấu hình IP cho các Switch như trên hình**

**Spine-1:**

interface ethernet 1

no switchport

ip address 10.10.1.1 255.255.255.0

no shutdown

interface ethernet 2

no switchport

ip address 10.10.2.1 255.255.255.0

no shutdown

interface ethernet 3

no switchport

ip address 10.10.3.1 255.255.255.0

no shutdown

**Spine-2:**

interface ethernet 1

no switchport

ip address 20.20.1.1 255.255.255.0

no shutdown

interface ethernet 2

no switchport

ip address 20.20.2.1 255.255.255.0

no shutdown

interface ethernet 3

no switchport

ip address 20.20.3.1 255.255.255.0

no shutdown

**Leaf-1:**

interface ethernet 1

no switchport

ip address 10.10.1.2 255.255.255.0

no shutdown

interface ethernet 2

no switchport

ip address 20.20.1.2 255.255.255.0

no shutdown

interface loopback 1

ip address 101.1.1.1 255.255.255.0

no shutdown

vlan 31

interface vlan 31

ip address 31.1.1.254 255.255.255.0

no shutdown

vlan 41

interface vlan 41

ip address 41.1.1.254 255.255.255.0

no shutdown

interface ethernet 3

switchport mode access

switchport access vlan 31

interface ethernet 4

switchport mode access

switchport access vlan 41

**Leaf-2:**

interface ethernet 1

no switchport

ip address 10.10.2.2 255.255.255.0

no shutdown

interface ethernet 2

no switchport

ip address 20.20.2.2 255.255.255.0

no shutdown

interface loopback 1

ip address 102.1.1.1 255.255.255.0

no shutdown

vlan 31

interface vlan 31

ip address 31.1.1.254 255.255.255.0

no shutdown

vlan 41

interface vlan 41

ip address 41.1.1.254 255.255.255.0

no shutdown

interface ethernet 3

switchport mode access

switchport access vlan 31

interface ethernet 4

switchport mode access

switchport access vlan 41

**Leaf-3:**

interface ethernet 1

no switchport

ip address 10.10.3.2 255.255.255.0

no shutdown

interface ethernet 2

no switchport

ip address 20.20.3.2 255.255.255.0

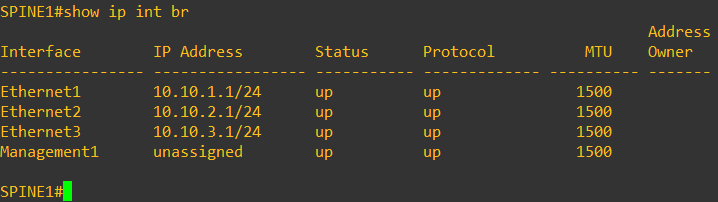
no shutdown

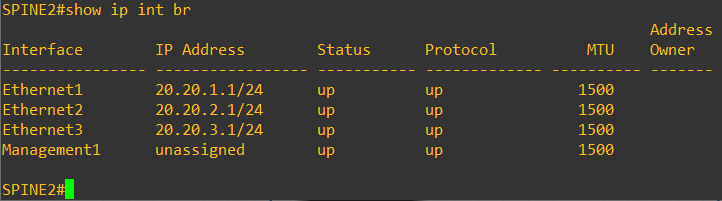
interface loopback 1

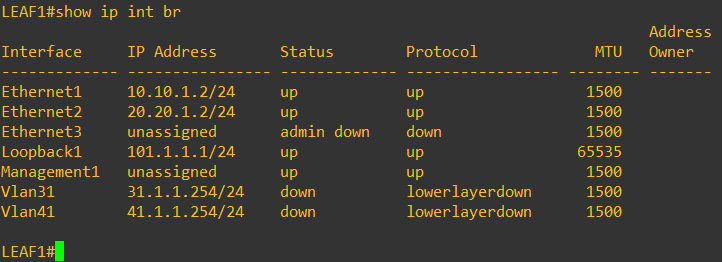
ip address 103.1.1.1 255.255.255.0

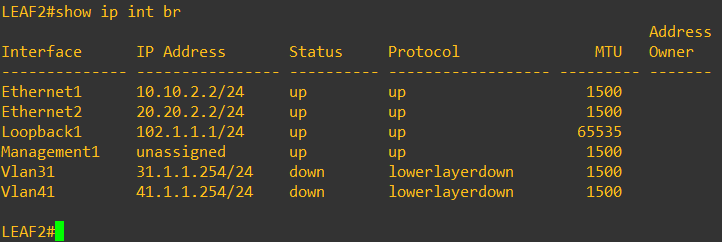
no shutdown

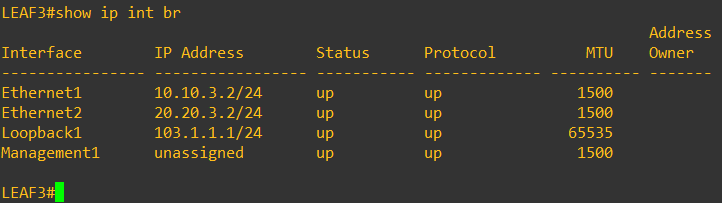
- Sau khi cấu hình xong:











**Step 2: Cấu hình giao thức BGP trên các switch**

**Spine-1:**

ip routing

router bgp 100

maximum-paths 2 ecmp 2

neighbor 10.10.1.2 remote-as 201

neighbor 10.10.1.2 maximum-routes 12000

neighbor 10.10.2.2 remote-as 202

neighbor 10.10.2.2 maximum-routes 12000

neighbor 10.10.3.2 remote-as 203

neighbor 10.10.3.2 maximum-routes 12000

redistribute connected

**Spine-2:**

router bgp 100

maximum-paths 2 ecmp 2

neighbor 20.20.1.2 remote-as 201

neighbor 20.20.1.2 maximum-routes 12000

neighbor 20.20.2.2 remote-as 202

neighbor 20.20.2.2 maximum-routes 12000

neighbor 20.20.3.2 remote-as 203

neighbor 20.20.3.2 maximum-routes 12000

redistribute connected

**Leaf-1:**

ip routing

router bgp 201

maximum-paths 2 ecmp 2

neighbor 10.10.1.1 remote-as 100

neighbor 10.10.1.1 maximum-routes 12000

neighbor 20.20.1.1 remote-as 100

neighbor 20.20.1.1 maximum-routes 12000

redistribute connected

**Leaf-2:**

ip routing

router bgp 202

maximum-paths 2 ecmp 2

neighbor 10.10.2.1 remote-as 100

neighbor 10.10.2.1 maximum-routes 12000

neighbor 20.20.2.1 remote-as 100

neighbor 20.20.2.1 maximum-routes 12000

redistribute connected

**Leaf-3:**

ip routing

router bgp 203

maximum-paths 2 ecmp 2

neighbor 10.10.3.1 remote-as 100

neighbor 10.10.3.1 maximum-routes 12000

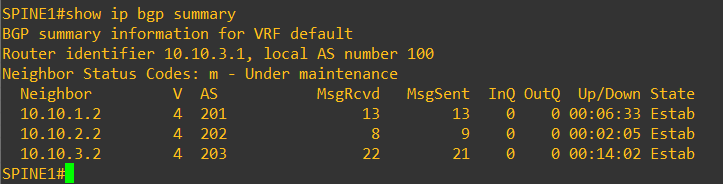
neighbor 20.20.3.1 remote-as 100

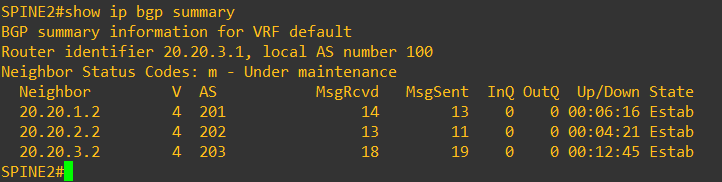
neighbor 20.20.3.1 maximum-routes 12000

redistribute connected

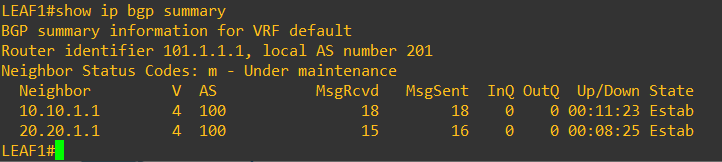
**Show kết quả khi thiết lập thành công phiên BGP:**

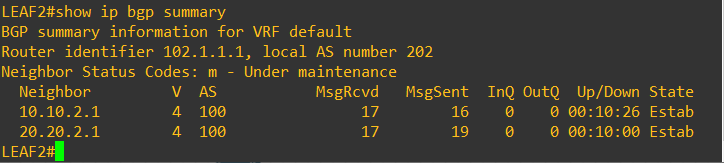
- Spine 1,2 thiết lập thành công với 3 Leaf bên dưới:

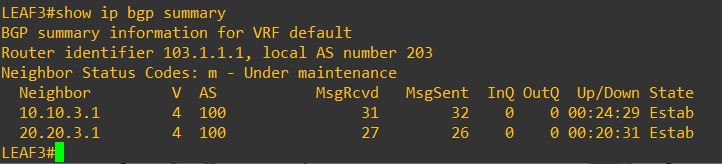




- Tương tự như Spine thì Leaf cũng vậy:

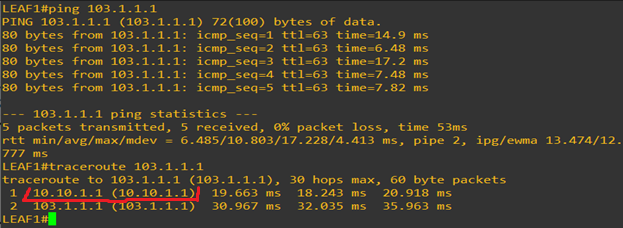




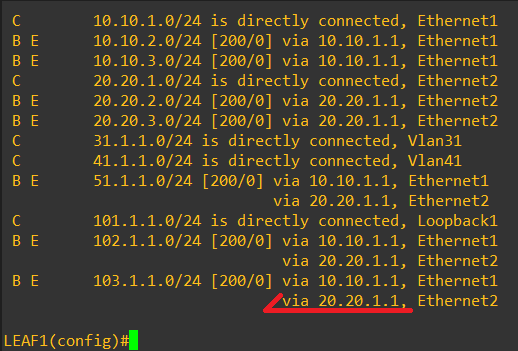


**Kiểm tra tính dự phòng của mô hình:**

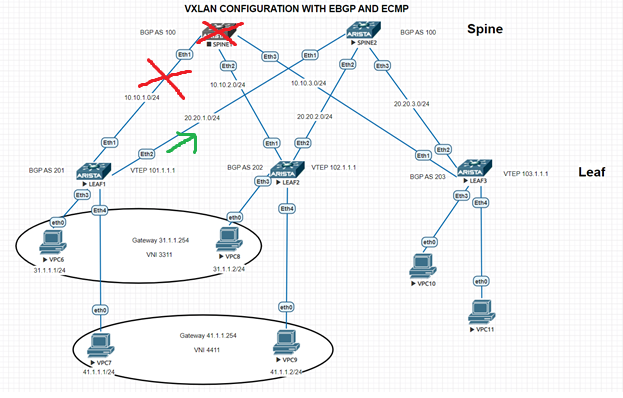
- Như hình bên dưới, ta thấy khi hoạt động bình thường thì khi ping từ Leaf-1 đến địa chỉ 103.1.1.1 (Leaf-3) thì nó sẽ đi qua Spine-1 rồi mới tới Leaf-3:

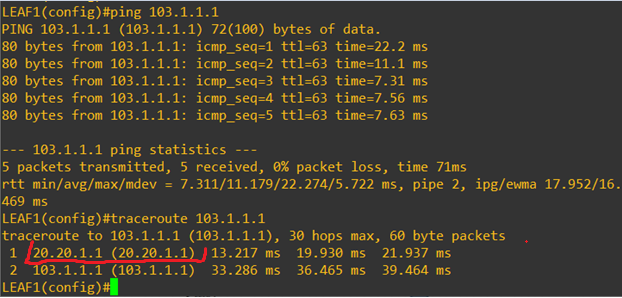


- Lệnh “show ip route” càng cho ta thấy rõ tính dự phòng của mô hình hơn, có thể thấy khi ping tới địa chỉ 103.1.1.1 ta có thêm cả 1 đường đi dự phòng nữa đó là đi qua cổng 20.20.1.1 nằm trên con Spine-2:



- Sau khi Shut down cổng Ethernet 1 hay khi Spine-1 xảy ra sự cố nào đó thì trên Leaf-1 khi ta ping tới địa chỉ 103.1.1.1 (Leaf-3) thì tuyến đường của nó sẽ chuyển sang Spine-2 rồi mới tới Leaf-3:





- Khi ta “show ip route” lại thì nó đã mất cổng 10.10.1.1 trên Spine-1, chỉ còn mỗi cổng 20.20.1.1 của con Spine-2:



**Step 3: Cấu hình Vxlan cho các Leaf để có thể kết nối các PC ở layer 2 thông qua layer 3**

**Leaf-1:**

interface Vxlan 1

vxlan source-interface Loopback 1

vxlan udp-port 4789

vxlan vlan 31 vni 3311

vxlan vlan 41 vni 4411

vxlan flood vtep 102.1.1.1 103.1.1.1

**Leaf-2:**

interface Vxlan 1

vxlan source-interface Loopback 1

vxlan udp-port 4789

vxlan vlan 31 vni 3311

vxlan vlan 41 vni 4411

vxlan flood vtep 101.1.1.1 103.1.1.1

**Leaf-3:**

interface Vxlan 1

vxlan source-interface Loopback 1

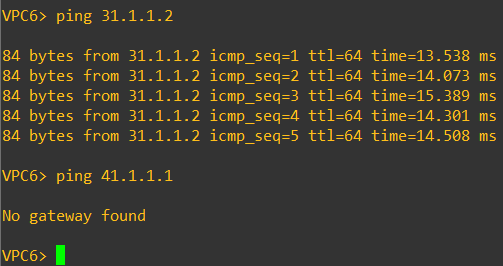
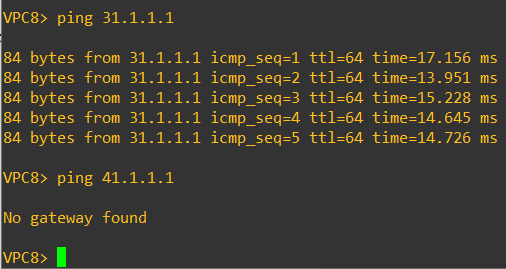
vxlan udp-port 4789

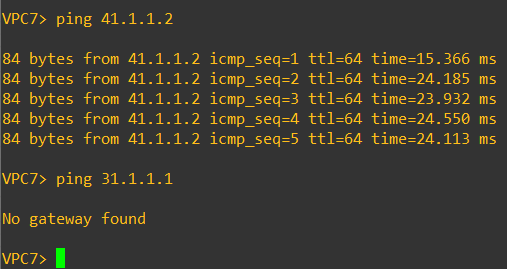
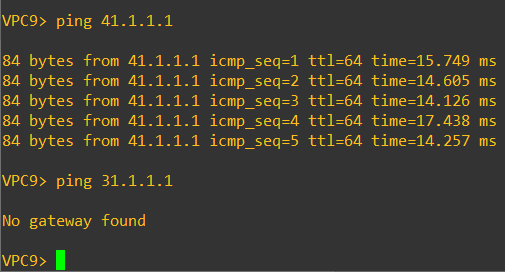
vxlan vlan 31 vni 3311

vxlan vlan 41 vni 4411

vxlan flood vtep 101.1.1.1 102.1.1.1

- Trước khi cấu hình gateway thì các PC thuộc vlan 31 hay vlan 41 chỉ có thể ping được đến các PC trong cùng một vlan:

**Cấu hình ip add virtual :**

**Leaf-1, Leaf-2 :**

interface Vlan 31

ip address virtual 31.1.1.254/24

interface Vlan 41

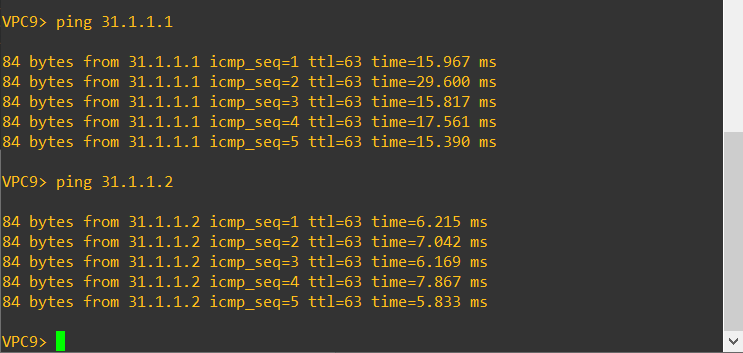
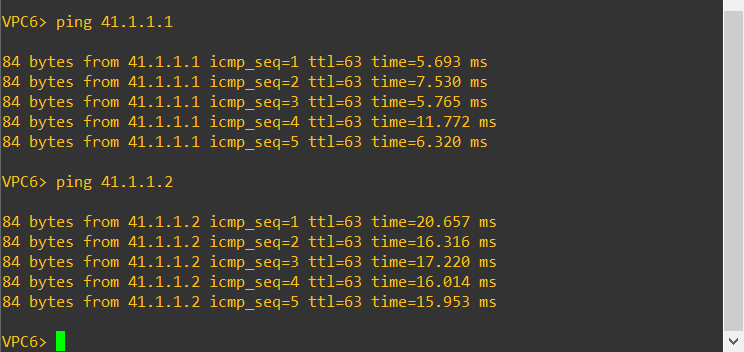
ip address virtual 41.1.1.254/24

**Leaf-3 :**

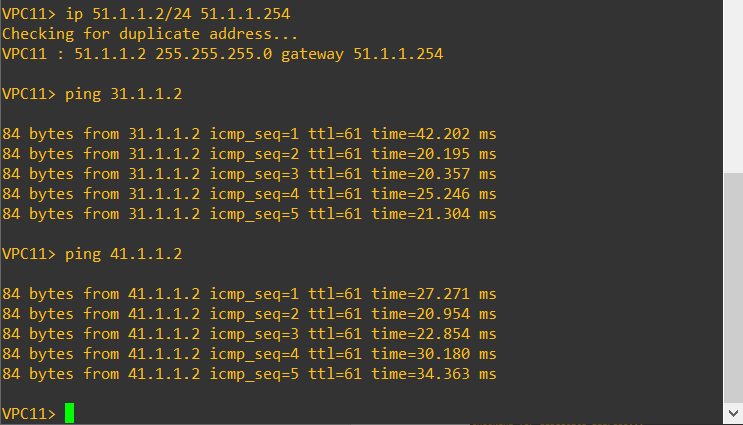
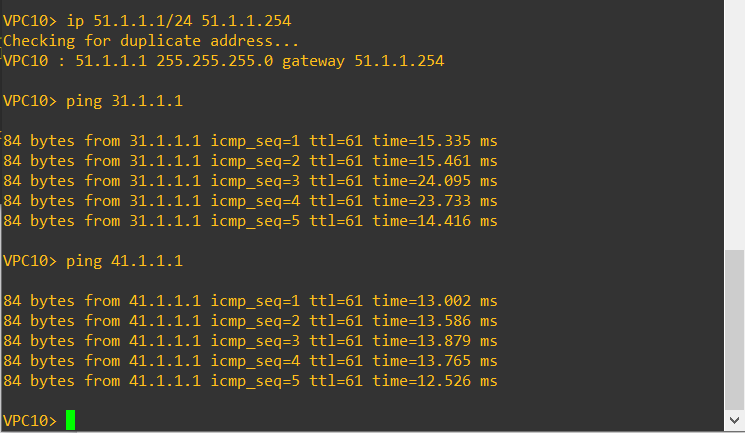
interface Vlan 51

ip address virtual 51.1.1.254/24

- Sau khi cấu hình gateway xong ta thấy từ vlan 31 ping được tới vlan 41 và ngược lại:



- Trên Leaf-3 tạo Vlan 51 và gán nó cho 2 PC 10 và 11. Mặc dù Leaf-3 không chạy Vxlan nhưng các PC thuộc Vlan 51 vẫn có thể kết nối được các PC thuộc Vlan 31 và 41:

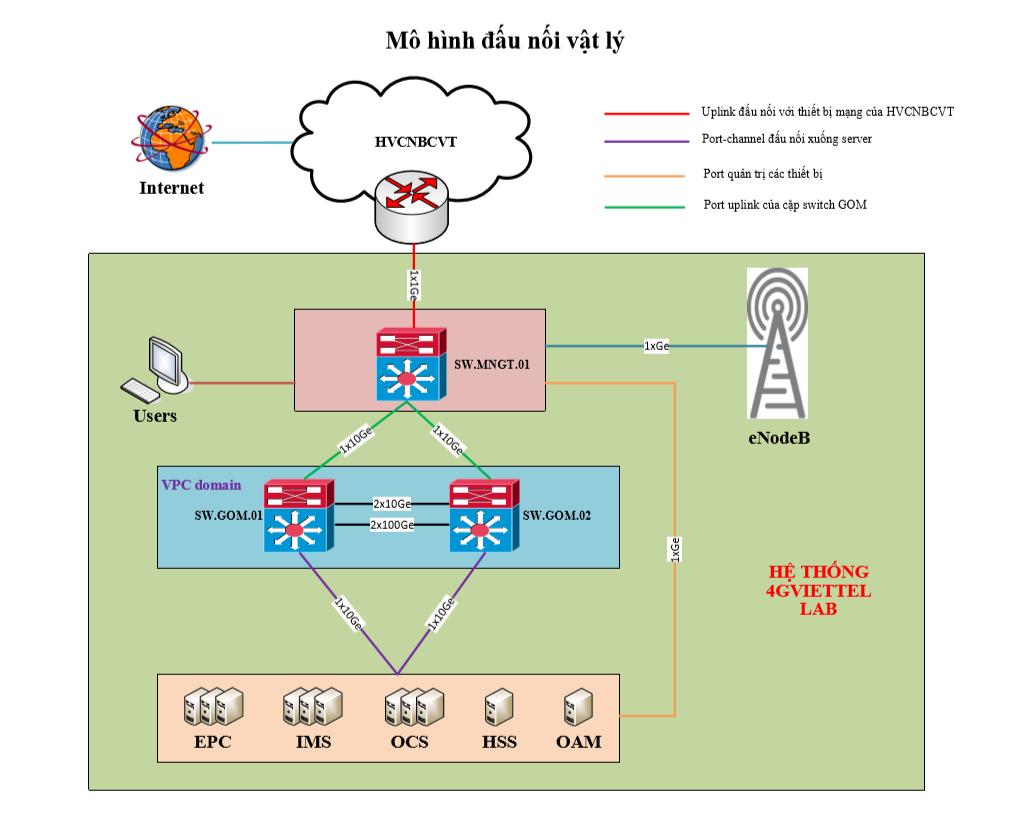


**3.5 Kết luận chương 3**

Qua cơ sở lý thuyết về nguyên lý hoạt động của VXLAN kết hợp với mô hình Spine-Leaf. Chương 3 đã thiết kế thành công một bài lab trên phần mềm mô phỏng EVE-NG để cho chúng ta thấy được tầm quan trọng trong việc chuyển dịch từ mô hình mạng trung tâm dữ liệu 3 lớp truyền thống sang mô hình mạng Spine-Leaf chạy VXLAN.

**CHƯƠNG 4: TÌM HIỂU VỀ HẠ TẦNG MẠNG ĐÃ TRIỂN KHAI TẠI PHÒNG LAB HVCNBCVT VÀ ĐỀ XUẤT CHUYỂN ĐỔI MÔ HÌNH SPINE-LEAF TRONG TƯƠNG LAI**

**4.1 Mô hình hạ tầng mạng đang triển khai tại phòng Lab HVCNBCVT**



Mặc dù mô hình này đã được triển khai và sử dụng tại phòng lab của HVCNBCVT trong nhiều năm, các thầy cô, các bạn sinh viên, các kỹ sư mạng đã quá quen thuộc với thiết kế này cũng như các giao thức mà nó sử dụng. Tuy nhiên, các trung tâm dữ liệu hiện đại rất khác so với chỉ một thời gian ngắn trước đây. Cơ sở hạ tầng mạng đã chuyển từ các máy chủ vật lý tại chỗ truyền thống sang cơ sở hạ tầng ảo hóa hỗ trợ các ứng dụng và khối lượng công việc trên các nhóm cơ sở hạ tầng mạng vật lý. Sự phát triển này đã diễn ra nhanh chóng và trong một khoảng thời gian ngắn mang đến các công nghệ phổ biến như ảo hóa, đám mây (private, public, hybrid), mạng định nghĩa bằng phần mềm (SDN).

Trong bối cảnh mới, nơi tất cả các khối lượng công việc và ứng dụng đang được chuyển đến các trung tâm dữ liệu, dần dần mô hình 3 lớp truyền thống không còn đủ để giải quyết tất cả các yêu cầu và các thách thức mà các Data Center hiện nay đặt ra:

* Khả năng mở rộng: Khả năng mở rộng là rất quan trọng, đặc biệt là trong dữ liệu dựa trên đám mây trung tâm. Một trung tâm dữ liệu có thể chứa hàng ngàn người thuê. Giới hạn mạng 4094 của 12-trường Vlan bit không đủ để hỗ trợ các trung tâm dữ liệu đa tầng lớn.
* Tính khả dụng: Một trung tâm dữ liệu phải có khả năng hoạt động liên tục (24/7). Ngoài ra, quyền truy cập vào các ứng dụng cần có sẵn từ tất cả các loại thiết bị như máy tính bảng, điện thoại thông minh và đồng hồ thông minh.
* Bảo mật: Đặc biệt là trong việc triển khai trung tâm dữ liệu nhiều tầng, yêu cầu chính là áp dụng các chính sách bảo mật hiệu quả để đảm bảo lưu lượng truy cập từ người thuê này được cách ly hoàn toàn với người thuê khác. Một số yêu cầu khác liên quan đến bảo mật bao gồm thực thi chính sách ứng dụng, ngăn chặn truy cập trái phép, phát hiện các mối đe dọa, cách ly thiết bị bị nhiễm, phân phối các bản vá bảo mật cho các thiết bị bị ảnh hưởng và các ứng dụng chính sách nhất quán giữa private cloud và public cloud.
* Hỗ trợ cho việc triển khai kết hợp: Cả doanh nghiệp và nhà cung cấp dịch vụ đã áp dụng mô hình đám mây ở một mức độ nào đó. Kết quả là, một trong những yêu cầu chính của trung tâm dữ liệu là hỗ trợ cho việc triển khai hybrid cloud, nơi các tài nguyên từ public cloud có thể được mở rộng đến trung tâm dữ liệu doanh nghiệp tư nhân một cách linh hoạt.
* Chi phí thấp.
* Dễ sử dụng, dễ quản lý.

Mô hình Spine-Leaf với nhiều ưu điểm hơn mô hình cũ và được phát triển để khắc phục các nhược điểm còn tồn tại của mô hình mạng cũ đang dần trở thành xu hướng được sử dụng trong các Data Center hiện nay.

**4.2 Những ưu điểm của mô hình Spine-Leaf so với mô hình mạng cũ**

- Loại bỏ giao thức Spanning-Tree (STP) để sử dụng giao thức định tuyến lớp 3 ECMP do đó tất cả các liên kết đều được sử dụng để cân bằng tải do các kết nối sử dụng các cổng có băng thông bằng nhau và có chính xác 2 bước nhảy giữa các Leaf Switch. Với kiến trúc Spine-Leaf, bất kể máy chủ nào được kết nối với máy chủ nào, lưu lượng truy cập của nó luôn phải vượt qua cùng một số thiết bị để đến máy chủ khác (trừ khi máy chủ khác nằm trên cùng một Leaf). Cách tiếp cận này giữ độ trễ ở mức có thể dự đoán được vì một lưu lượng chỉ phải đi tới một Spine Switch và một Leaf Switch khác để đến đích.

- Tính dự phòng cao: Nếu 1 Spine Switch bị lỗi, nó chỉ giảm 1 phần nhỏ hiệu suất của mạng mà không ảnh hưởng đến dịch vụ. Nếu 1 Leaf Switch bị lỗi, nó chỉ ảnh hưởng đến các máy chủ đang kết nối tới Leaf switch đó.

- Tính mở rộng cao: Nếu cần thêm băng thông, chỉ cần thêm Spine Switch, nếu cần thêm nhiều máy chủ, chỉ cần thêm Leaf Switch mà không phải thiết kế lại toàn bộ hệ thống.

- Sử dụng VXLAN: VXLAN sử dụng ID 24 bit có thể mở rộng quy mô lên đến 16 triệu phân đoạn VXLAN trong một miền quản trị, trái ngược với con số 4094 của VLAN. Bởi vậy, công nghệ VXLAN cho phép cô lập mạng một cách hiệu quả những người thuê hàng loạt trong điện toán đám mây.

- Hỗ trợ SDN: Sử dụng công nghệ SDN ta có thể đặt máy ảo tại bất kỳ vị trí nào trong hệ thống mạng nhưng vẫn giữ được kết nối Layer 2 với các máy ảo khác. Có nghĩa là tách kiến trúc ảo (Virtual Topology) ra khỏi kiến trúc vật lý (Physical Topology). Kiến trúc mạng 3 lớp không cho phép tạo ra hệ thống hạ tầng mạng vật lý (Underlay Network) để có thể xây dựng hệ thống mạng logic thượng tầng (Overlay Network) đáp ứng yêu cầu của kiến trúc mạng SDN.

**4.3 Đề xuất thiết kế mô hình Spine-Leaf kết nối 2 site Hà Nội và Hà Chí Minh của HVCNBCVT**

Kiến trúc multisite có tính năng chia sẻ các loại Fabric hoàn toàn độc lập tài nguyên cụ thể và giao tiếp thông qua nhà cung cấp dịch vụ trung gian mạng. Giả sử bạn có DC1 và DC2 trong cấu hình nhiều site. DC1 thực hiện mối quan hệ láng giềng eBGP với nhà cung cấp dịch vụ của bạn, trong đó nhà cung cấp dịch vụ mang lưu lượng EVPN đến DC2 đích của bạn.

Khía cạnh ngang hàng xảy ra ở Leaf Boder Gateway (BGW). Ở đây Leaf là nơi bạn giao tiếp với BGP nội bộ (iBGP) rồi tới eBGP AS bên ngoài, sau đó chuyển tiếp đến nhà cung cấp dịch vụ. Trong DC2 thì kiến trúc cũng tương tự. Leaf BGW ngang hàng với iBGP AS của bạn, tương đương với số eBGP AS của bạn, minh họa trong Hình dưới:

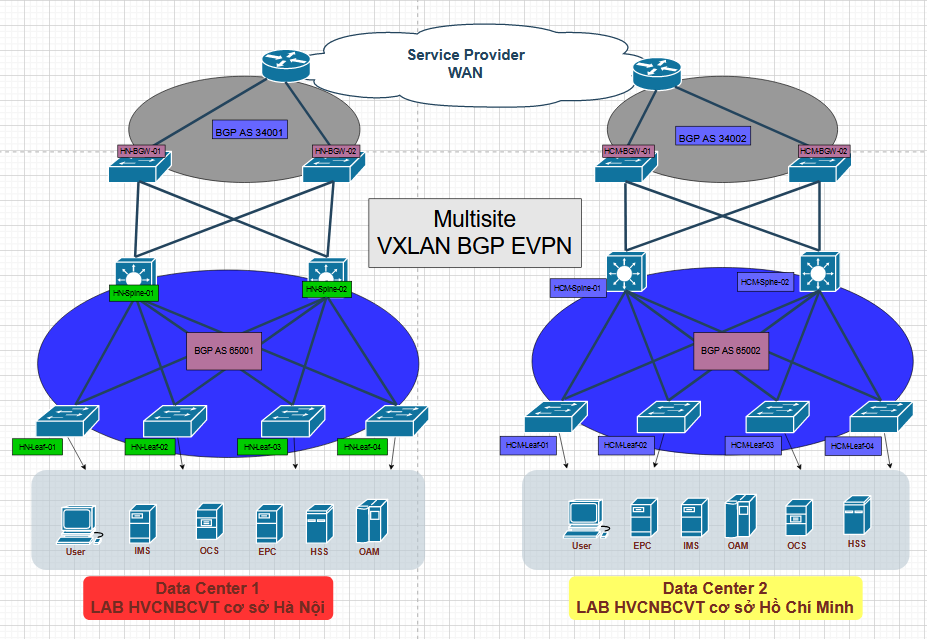


**Hình 4.1: Đường dẫn BGP giữa hai trung tâm dữ liệu trong một cấu hình**

**Multisite**

Vai trò của Leaf BGW là thực hiện sự ngang hàng giữa iBGP và eBGP và thực hiện AS NAT hoặc là BGP AS number từ internal sang external cho lưu lượng truy cập cụ thể EVPN.

Giả sử thiết kế mô hình 2 site Hà Nội và Hồ Chí Minh của HVCNBCVT với các thông số như sau: DC1 Fabric AS là 65001, DC1 eBGP AS là 34001, DC2 Fabric  
AS là 65002 và DC2 eBGP AS là 34002. Thời điểm ta gửi một yêu cầu liên lạc của mình tới DC2, nguồn của giao tiếp thay đổi từ DC1 65001 sang DC1 34001  
eBGP AS. BGW chuyển 65001 thành 34001 để chuyển sang đám mây của nhà  
cung cấp dịch vụ. Khi gói đến DC2, nó được ghi lại thành 65002, là Fabric của DC2 AS. Lưu lượng truy cập đến sẽ được giả mạo để trông giống như trên cùng một loại Fabric ( Xem hình bên dưới ).



**Hình 4.2 : Multisite Vxlan BGP EVPN**

Mô hình Spine-Leaf sử dụng công nghệ VXLAN để mở rộng các phân đoạn lớp 2 qua các vị trí cho nên các mạng này sẽ không bị giới hạn về mặt địa lý nên mặc dù các server ở 2 site Hà Nội và Hồ Chí Minh nhưng vẫn có thể dùng chung một hệ thống để quản lý, điều khiển.

**4.4 Kết luận chương 4**

Ngày nay mạng máy tính đang ngày càng phát triển dẫn đến các nhu cầu mở rộng, nhu cầu quản lý mạng ngày càng tăng, đòi hỏi số lượng thiết bị ngày càng lớn gây ra khó khăn cho người quản trị như khả năng mở rộng kém, lãng phí tài nguyên mạng. Mô hình truyền thống ba lớp đang được sử dụng rất nhiều trong hệ thống Data Center hiện nay là mô hình ba lớp Core - Distribution - Access kết hợp với giao thức STP, VPC. Sự phát triển nhanh chóng về ảo hóa, cloud và BIG DATA làm cho mô hình 3 lớp nở to và không còn linh hoạt nữa nên xu hướng thế giới nói chung đang chuyển sang mô hình thiết kế mới là mô hình Spine – Leaf hoạt động trên giao thức VXLAN để tạo ra một kiến trúc với các ưu điểm truyền thông nhanh, có thể dự đoán, có thể mở rộng và hiệu quả trong môi trường trung tâm dữ liệu.

**KẾT LUẬN**

Đồ án đã cho chúng ta thấy rõ hơn về những ưu điểm của Công nghệ Vxlan và Mô hình Spine-Leaf trong mạng Data Center.

**Sau quá trình tìm hiểu và nghiên cứu, đồ án đã thực hiện được những nội dung sau:**

* Giới thiệu tổng quan về mạng trung tâm dữ liệu ba lớp truyền thống cũng như nhược điểm của mạng này. Qua đó, đồ án đã làm nổi bật lên ưu điểm của mô hình mạng Spine – Leaf dựa trên giao thức VXLAN.
* Giới thiệu tổng quan về công nghệ VXLAN và những lợi ích của VXLAN trong mạng doanh nghiệp đang dần thay thế cho VLAN.
* Xây dựng một bài lab demo cơ bản về kiến trúc mạng Spine-Leaf sử dụng VXLAN để tối ưu hơn so với mô hình truyền thống 3 lớp.
* Tìm hiểu về mô hình hạ tầng mạng đang triển khai tại phòng Lab của HVCNBCVT và đề xuất chuyển đổi sang mô hình Spine-Leaf trong tương lai thông qua những ưu điểm vượt trội mà nó đem lại.

**TÀI LIỆU THAM KHẢO**

1.<https://networklessons.com/cisco/ccna-200-301/spine-and-leaf-architecture>

2.<https://www.cisco.com/c/en/us/td/docs/switches/datacenter/sw/nx-os/vxlan/configuration/guide/b_NX-OS_VXLAN_Configuration_Guide/overview.pdf>

3.<https://www.cisco.com/c/dam/m/sl_si/events/2016/cisco_dan_inovativnih_resitev/pdf/cisco_day_slovenia_2016_vxlan_marian_klas_final.pdf>

4.<https://viettelidc.com.vn/tin-tuc/vxlan-va-vlan-dau-moi-la-su-lua-chon-phu-hop-nhat-cho-cac-trung-tam-du-lieu-dam-may>

5. <https://www.ciscopress.com/articles/article.asp?p=2999385&seqNum=3>