ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS

INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

MAESTRÍA EN GERENCIA DE SISTEMAS

TRABAJO DE TITULACIÓN, PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGISTER EN GERENCIA DE SISTEMAS

TEMA: DESARROLLO DE UNA PROPUESTA DE ADMINISTRACIÓN Y APROVISIONAMIENTO DE CLIENTES EN EL DATACENTER DE TELCONET DE LA CIUDAD DE QUITO BASADA EN EL USO DE LA TECNOLOGÍA VXLAN

AUTOR: NARANJO ESPÍN, EDISON FABRICIO

DIRECTOR: ING. SALAZAR CHACÓN, GUSTAVO DAVID

SANGOLQUÍ

2018

CERTIFICADO DEL DIRECTOR

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, ***"DESARROLLO DE UNA PROPUESTA DE ADMINISTRACIÓN Y APROVISIONAMIENTO DE CLIENTES EN EL DATACENTER DE TELCONET DE LA CIUDAD DE QUITO BASADA EN EL USO DE LA TECNOLOGÍA VXLAN"*** realizado por el señor Naranjo Espín Edison Fabricio, ha sido revisado en su totalidad y analizado por el software antiplagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar al señor Naranjo Espín Edison Fabricio para que lo sustente públicamente.

Quito, 18 de Septiembre del 2017

Gustavo David Salazar Chacón

CC: 1716104797

DIRECTOR

AUTORÍA DE RESPONSABILIDAD

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

AUTORÍA DE RESPONSABILIDAD

Yo, Naranjo Espín Edison Fabricio, con cédula de identidad No 1712439601, declaro que este trabajo de titulación ***"DESARROLLO DE UNA PROPUESTA DE ADMINISTRACIÓN Y APROVISIONAMIENTO DE CLIENTES EN EL DATACENTER DE TELCONET DE LA CIUDAD DE QUITO BASADA EN EL USO DE LA TECNOLOGÍA VXLAN"*** ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas. Consecuentemente declaro que este trabajo es de mi autoría, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada.

Quito, 18 de Septiembre del 2017

Edison Fabricio Naranjo Espín

CC: 1712439601

AUTORIZACIÓN (PUBLICACIÓN BIBLIOTECA VIRTUAL)

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

AUTORIZACIÓN

Yo, Naranjo Espín Edison Fabricio, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución el presente trabajo de titulación ***"DESARROLLO DE UNA PROPUESTA DE ADMINISTRACIÓN Y APROVISIONAMIENTO DE CLIENTES EN EL DATACENTER DE TELCONET DE LA CIUDAD DE QUITO BASADA EN EL USO DE LA TECNOLOGÍA VXLAN"*** cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Quito, 18 de Septiembre del 2017

Edison Fabricio Naranjo Espín

CC: 1712439601

DEDICATORIA

A mi familia, por haber fomentado en mí el deseo de superación y el anhelo de triunfo en la vida. Gracias a ustedes, hoy puedo ver alcanzada mi meta, ya que siempre estuvieron impulsándome en los momentos más difíciles y porque el orgullo que sienten por mí fue lo que me hizo ir hasta el final. Va por ustedes, por lo que valen, porque admiro su fortaleza y por lo que han hecho de mí.

Ahora que he conseguido este objetivo al que lo puedo llamar nuestro porque siempre están conmigo, no sé que tan lejos o cerca puedo llegar, lo que si sé es que siempre retribuiré su confianza y sus buenos deseos.

Con mucho cariño

Edison F. Naranjo E.

AGRADECIMIENTO

En primer lugar deseo agradecer a Dios y a la Virgen Santísima por darme la vida, salud, protección y guía para poder alcanzar esta nueva meta.

A mi madre **Margoth Espín**, por su apoyo incondicional a cada momento, por sus consejos, sus valores, por la motivación constante que me ha permitido ser una persona de bien, pero más que nada, por su amor y su infinito deseo de bienestar para mí.

A mi padre **Luis Fernando Naranjo**, por su ejemplo de perseverancia, constancia y trabajo tesonero que lo caracterizan y que me ha transmitido siempre, por su amor y por el valor mostrado para salir adelante.

A mis hermanos **Katia** y **Fernando Naranjo** por su cariño, ayuda incondicional y porque nunca han dejado de creer en mí y cuan lejos puedo llegar.

A mis sobrinos **Verito** y **Fernandito** por ver en mí siempre un ejemplo a seguir.

Al área de Networking de Telconet S.A. por su valiosa ayuda para la elaboración de este trabajo.

Y finalmente agradezco al Ing. Gustavo Salazar por su apoyo y confianza para llegar a finalizar con éxito este proyecto de graduación.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

RESUMEN

Este trabajo fue desarrollado para analizar la estructura y operación de la tecnología Virtual Extensible LAN (VXLAN). Una simulación / emulación en un entorno virtualizado demostró las ventajas de VXLAN; y con base en los resultados obtenidos, esta investigación podría utilizarse como propuesta para la administración y aprovisionamiento de proveedores de servicios y clientes del centro de datos (DC) en el futuro.

Palabras Clave

- **VXLAN**
- **MP-BGP**
- **EVPN**
- **SDN**
- **MULTICAST**

ABSTRACT

This work was developed in order to analyze the structure and operation of Virtual Extensible LAN technology (VXLAN). A simulation/emulation in a virtualized environment demonstrated the advantages about VXLAN; and based on the results, this investigation could be used as a proposal for the administration and provisioning of Service Providers and Data Center (DC) clients in the future.

Keywords

- **VXLAN**
- **MP-BGP**
- **EVPN**
- **SDN**
- **MULTICAST**

CAPÍTULO I

GENERALIDADES

1.1. Antecedentes

Las TI están evolucionando hacia un modelo de consumo en la nube. Esta transición afecta la forma en que se están diseñando e implementando las aplicaciones, lo que conduce a una evolución en el diseño de la infraestructura de los data centers para satisfacer estos requerimientos. Como base de los data centers modernos, la red también debe tomar parte en esta evolución al mismo tiempo que existe un incremento en la virtualización de servidores y arquitecturas basadas en microservicios. En este nuevo paradigma se deben tratar los siguientes aspectos:

- **Flexibilidad** para permitir la movilidad del trabajo a través de cualquier sitio.

- **Resiliencia** para mantener los niveles de servicio aún en condiciones de falla.

- **Capacidad Multitenant** y una mejor segmentación de la carga de trabajo en la red.

- **Rendimiento** para proveer un ancho de banda adecuado y una latencia previsible, independiente de una escala para cargas de trabajo exigentes.

- **Escalabilidad** desde entornos pequeños hacia la nube manteniendo las características anteriores.

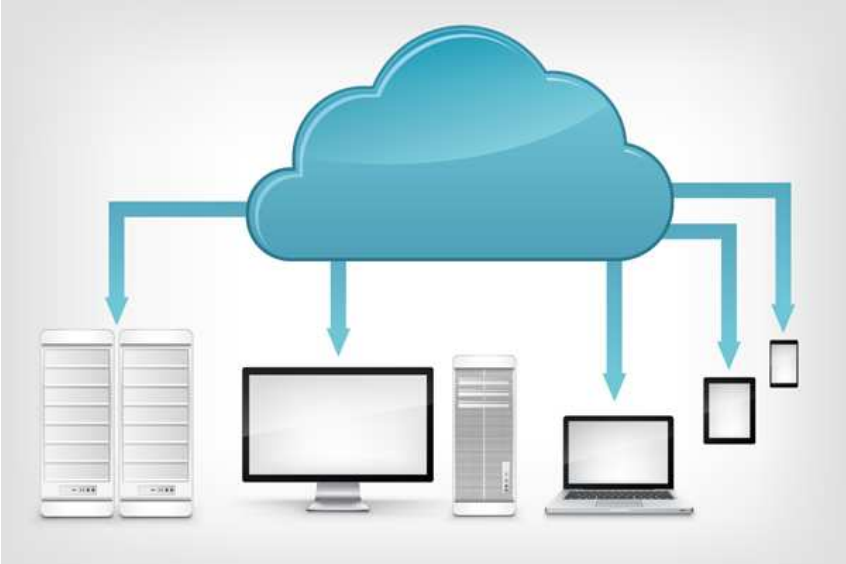


Figura 1. Sistemas basados en cloud

Como resultado de ésto, las redes de los data centers modernos están evolucionando en su diseño tradicional jerárquico a arquitecturas spine-leaf con hosts y servicios distribuidos a través de la red. Estas redes son capaces de soportar el incremento cada vez mayor de flujo de tráfico en las aplicaciones modernas. Cabe mencionar también que existen tecnologías de clustering y virtualización que requieren adyacencia a nivel de capa 2.

La evolución de la demanda de usuarios y requisitos de las aplicaciones sugieren un enfoque diferente que es simple y más ágil. La facilidad de aprovisionamiento y la velocidad constituyen métricas críticas de rendimiento de la infraestructura de red de los data centers que soportan ambiente físicos, virtualizados y de cloud; sin comprometer la escalabilidad o seguridad. Estos son los motores principales de la industria para buscar soluciones de red definidas por software (SDN).

Una opción alternativa es VXLAN Fabric con plano de control BGP EVPN, la cual provee una solución flexible, escalable y administrable que soporta la creciente demanda en ambientes basados en la nube.

1.2. Planteamiento del problema

La necesidad de redes virtuales propias a cada cliente dentro de las infraestructuras Cloud multitenant, suele implicar la utilización de técnicas de overlay que permitan la separación lógica de los datos que por ellas circulan, así como una gestión segmentada y la elasticidad de la topología de red, alcanzando de esta manera las ventajas de la virtualización de máquinas también para el ámbito de red.

Uno de los principales problemas al implementar una infraestructura de red, sea esta física o virtual, es el agotamiento de direcciones IPV4, lo que conlleva a la búsqueda y reutilización de redes disponibles en clientes cancelados y

dispositivos de red que ya no están en uso; y además, la falta de escalabilidad en la segmentación por medio de VLANs.

Al utilizar la tecnología VXLAN se busca solucionar el problema de escalabilidad en la asignación de VLANs así como también optimizar el direccionamiento lógico ya que se podrá conectar sitios remotos con el mismo bloque de direccionamiento IPV4.

1.3. Formulación del problema

Este trabajo de investigación tuvo origen con base en las siguientes inquietudes:

- Escalabilidad a nivel de VLAN.
- Conexión de redes remotas con el mismo bloque de direccionamiento IPV4.

Para contribuir a la solución de estos problemas, se realizará una simulación donde se indicará las ventajas de utilizar la tecnología VXLAN. El modelo generado y los resultados obtenidos se podrán utilizar posteriormente para cualquier implementación de VXLAN en infraestructuras que brinden servicios de Cloud Computing.

1.4. Justificación e importancia

Según Arizmendi (2014), los Data Centers Multitenant son aquellos que permiten proveer de un servicio y gestión aislados a varios clientes diferenciados, reservando a cada cliente un segmento de esa “partición”del Data Center. Gracias a las plataformas de virtualización, el uso de los Data Center Multitenant se ha incrementado. Actualmente no se encuentran solo en grandes proveedores de servicio, sino que también son utilizados internamente como parte de infraestructuras de red para asignar a cada departamento un tenant diferente y de esa manera proveer de los servicios IT de una manera controlada y segura. (Arizmendi, 2014)

Una forma tradicional de aprovisionar redes virtuales ha sido empleando el despliegue de arquitecturas overlay. El concepto de Overlay Network (redes superpuestas en español) es ampliamente utilizado en proveedores de servicio y empresas con redes de gran tamaño. La idea es implementar redes “encima“ de otras que ya han sido creadas a través de túneles entre los enlaces de los nodos overlay sobre una infraestructura de red establecida, llamada underlay, mediante la encapsulación de paquetes. (Colomés, 2015)

La justificación para realizar una simulación de la tecnología VXLAN consiste en dar a conocer las ventajas y limitaciones que brindan las redes overlay. Estas son algunas de las principales ventajas:

- Escalabilidad.
- Flexibilidad.

- Facilidad de reconfiguración de red.

- Separación lógica de redes de cada cliente.

1.5. Hipótesis

La propuesta de un modelo de implementación de la tecnología VXLAN con base en técnicas de simulación permitirá demostrar la conexión de dos redes separadas físicamente usando el mismo espacio de direccionamiento IPV4 y la misma etiqueta (tag) de VLAN en ambos sitios si así se requiriera.

1.6. Objetivos

1.6.1. Objetivo General

Desarrollar una propuesta de administración y aprovisionamiento de clientes para el DataCenter de Telconet de la ciudad de Quito empleando la tecnología VXLAN.

1.6.2. Objetivos Específicos

- Analizar la estructura y conceptos de la tecnología VXLAN.

- Simular el funcionamiento de la tecnología VXLAN en un entorno virtualizado.

- Documentar una propuesta de aprovisionamiento de clientes en el Data Center de Telconet de la ciudad de Quito basada en la tecnología VXLAN.

CAPÍTULO II

MARCO TEÓRICO

2.1. Introducción

Las redes overlay son una técnica utilizada en el estado del arte de los data centers modernos sobre una red por naturaleza estática mediante la virtualización. Antes de entrar en detalles de cómo trabajan las redes overlay, los desafíos que pueden enfrentar y la solución de sus problemas inherentes es necesario entender por qué las redes tradicionales son estáticas.

Cuando las redes fueron desarrolladas inicialmente, no existían aplicaciones que permitan la movilidad de un lugar a otro mientras éstas estaban en uso. Como resultado de ésto, los pioneros de TCP / IP usaron las direcciones IPV4 tanto para identificar al dispositivo como para su localización en la red. Esto era una cosa perfectamente razonable hacerlo con computadoras y aplicaciones que no permiten movilidad o con una movilidad muy limitada.

Hoy en día, los data centers modernos son implementados mediante máquinas virtuales (VM) o contenedores (dockers). La carga de trabajo de las aplicaciones virtualizadas puede ser insuficiente a través de múltiples ubicaciones. Las máquinas virtuales o contenedores pueden ser también móviles entre dis-

tintos hosts. Sus direcciones IPV4 ya no indican su localización. Debido a la estrecha relación entre la localización de los dispositivos finales y su identidad en el modelo de red tradicional, éstos requieren cambiar de dirección IPV4 cuando se mueven de una localización a otra. Ésto rompe el modelo de movilidad contínuo requerido por las aplicaciones virtualizadas.

Por lo tanto, el trabajo en red debe evolucionar desde el modelo estático a un modelo flexible para poder dar soporte a las comunicaciones entre las aplicaciones sin importar dónde se encuentren. Un enfoque es determinar la identidad de un dispositivo final desde su ubicación física en la red para que su localización puedan ser cambiada sin romper las comunicaciones entre dispositivos finales. Es aquí donde las redes overlay entran en escena.

2.2. Virtual Extensible LAN (VXLAN)

Virtual Extensible LAN (VXLAN) tal como se define en el RFC 7348 es una tecnología overlay diseñada para proporcionar conectividad de capa 2 y capa 3 sobre una red IP tradicional. Las redes IP proporcionan escalabilidad, balanceo de carga y recuperación predecible contra fallos. VXLAN logra esto a través del entunelamiento de tramas de capa 2 dentro de paquetes IP. VXLAN sólo requiere conectividad IP entre los dispositivos de borde que manejan VXLAN (VTEP), la cual es provista por un protocolo de enrutamiento.

El estándar VXLAN define el paquete ilustrado en la figura 2

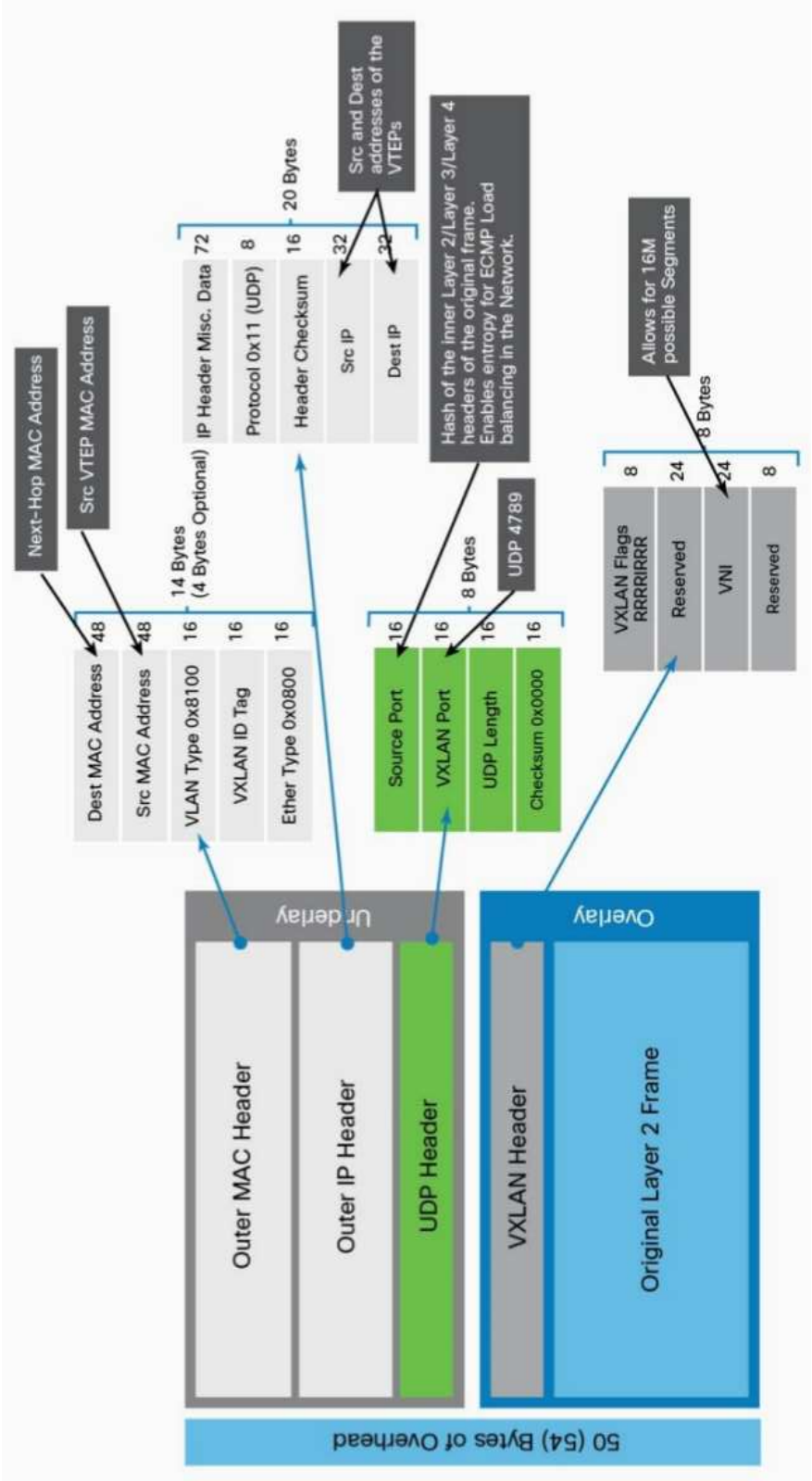


Figura 2. Jansen,D.(2017).VXLAN Frame Format Details.[Figura].Recuperado de Building Data Centers with VXLAN BGP EVPN

VXLAN usa una cabecera de 8 bytes que consiste en un identificador de 24 bits (VNID) y múltiples bits reservados. La cabecera VXLAN, a lo largo de la trama original Ethernet es colocada como una carga UDP. Los 24 bits del VNID se usa para identificar segmentos de capa 2 y mantener el aislamiento entre estos segmentos. Con los 24 bits del VNID, VXLAN puede soportar 2^{24} segmentos locales.

La terminología usada cuando se describen los componentes clave de la tecnología VXLAN es la siguiente:

- **VTEP.-** Elemento de hardware o software encargado de instanciar el entunelamiento VXLAN y llevar a cabo la encapsulación y desencapsulación de VXLAN. Conocido también como LEAF.

- **VNI.-** (Virtual Network Instance) Instancia de red lógica que provee servicios de capa 2 o de capa 3 y define un dominio de broadcast de capa 2.

- **VNID.-** (Virtual Network Identifier) Identificador de 24 bits que permite direccionar alrededor de 16 millones de redes lógicas.

- **Bridge-Domain.-** Conjunto de puertos físicos o lógicos que comparten el mismo dominio de broadcast.

- **SPINE.-** Dispositivo que interconecta los LEAF. No siempre require ser configurado como VTEP

Como se mencionó en el ítem 2.1, el uso de la tecnología VXLAN trae grandes beneficios a los Data Centers, entre los cuales se incluye además:

- **Multi-tenancy:** VXLAN intrínsecamente soporta multi-tenancy tanto para capa 2 (VNI de capa 2 separados, los que se encuentran representados lógicamente por bridge domains aislados) como para capa 3 (definiendo diferentes VRF para cada cliente).

- **Movilidad:** La capacidad overlay ofrecida por VXLAN brinda una extensión de servicios de capa 2 a los data centers proveyendo un despliegue flexible y movilidad a estaciones de trabajo tanto físicas como virtualizadas.

- **Incremento de la escalabilidad a nivel de capa 2:** El diseño empleando VLAN es limitado al uso de máximo 4096 segmentos de capa 2 debido a los 12 bits que corresponden al VLAN ID. VXLAN introduce un VNID de 24 bits que en teoría soporta más de 16 millones de segmentos de capa 2.

- **Soporte multi-path en capa 2:** Las redes de capa 2 tradicionales soportan un camino (path) activo debido a que el protocolo Spanning Tree (STP) forza a una topología libre de loops bloqueando caminos redundantes. VXLAN hace uso de una red subyacente de capa 3 (underlay network) para el uso varios caminos activos (multi-path).

2.2.1. Arquitectura Spine-Leaf

En este tipo de arquitectura cada leaf se interconecta con cada spine de manera redundante formando una topología donde la información viajará máximo dos saltos hasta alcanzar su destino. Es de alto rendimiento y muy usada en los esquemas de Data Center

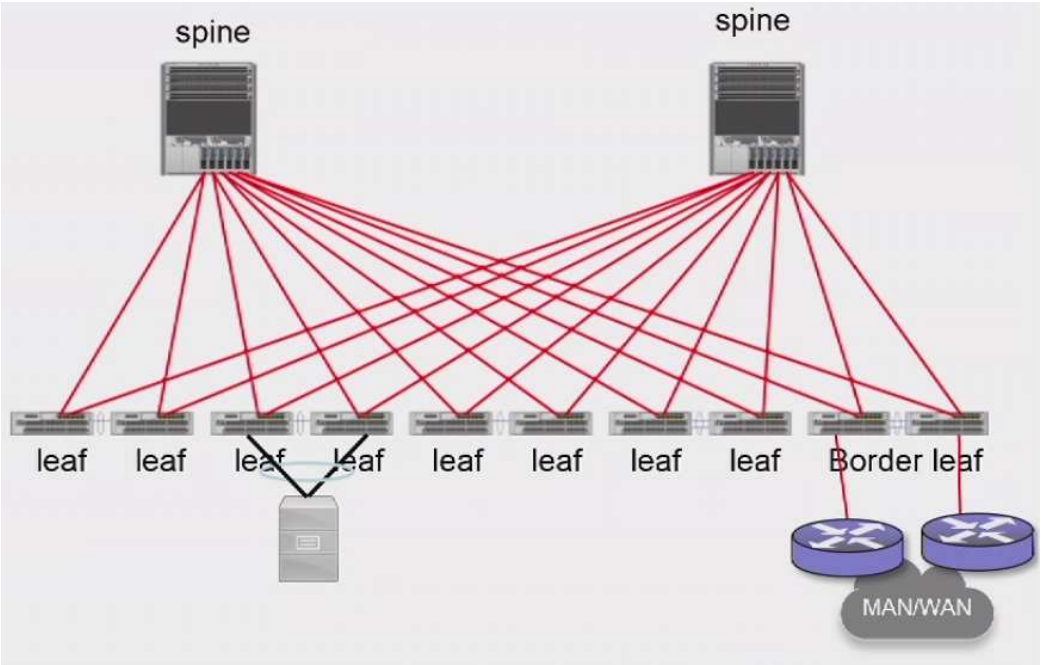


Figura 3. Arquitectura Spine-Leaf

2.2.2. Roles en la arquitectura Spine-Leaf

1. **SPINE**

   - Interconecta a los LEAFS.
   - Reenvía tráfico entre los LEAFS (tráfico EAST-WEST).
   - Route Reflector para EVPN

- Rendezvous-Point (RP) en la red underlay.
- Si no es un dispositivo de borde, no requiere ser configurado como VTEP.

2. **LEAF**

- Dispositivo de borde en una interconexión VXLAN.
- Realiza la encapsulación y desencapsulación de paquetes VXLAN.
- Interconecta los dispositivos finales.

3. **LEAF DE BORDE**

- Dispositivo de borde en una interconexión VXLAN.
- Intercambia tráfico con redes externas y lo encapsula en paquetes VXLAN. (Tráfico NORTH-SOUTH)
- Intercambia información de protocolos de enrutamiento (IGP/EGP) con redes externas. (Tráfico NORTH-SOUTH)
- Se encuentra representado en la figura 4

4. **SPINE DE BORDE**

- Cumple las funciones de un SPINE y de un LEAF de borde.
- Provee conectividad con redes externas.
- Requiere ser configurado como VTEP.
- Se encuentra representado en la figura 5

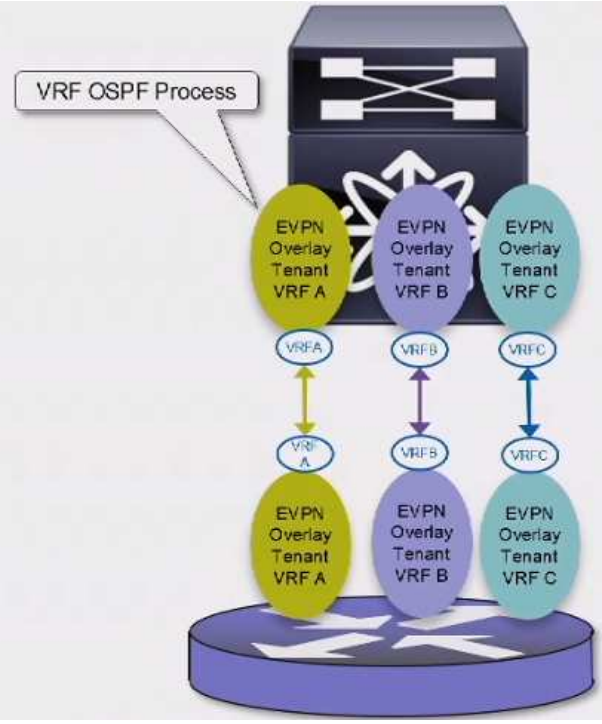


Figura 4. Interconexión de un Leaf de borde con un router externo

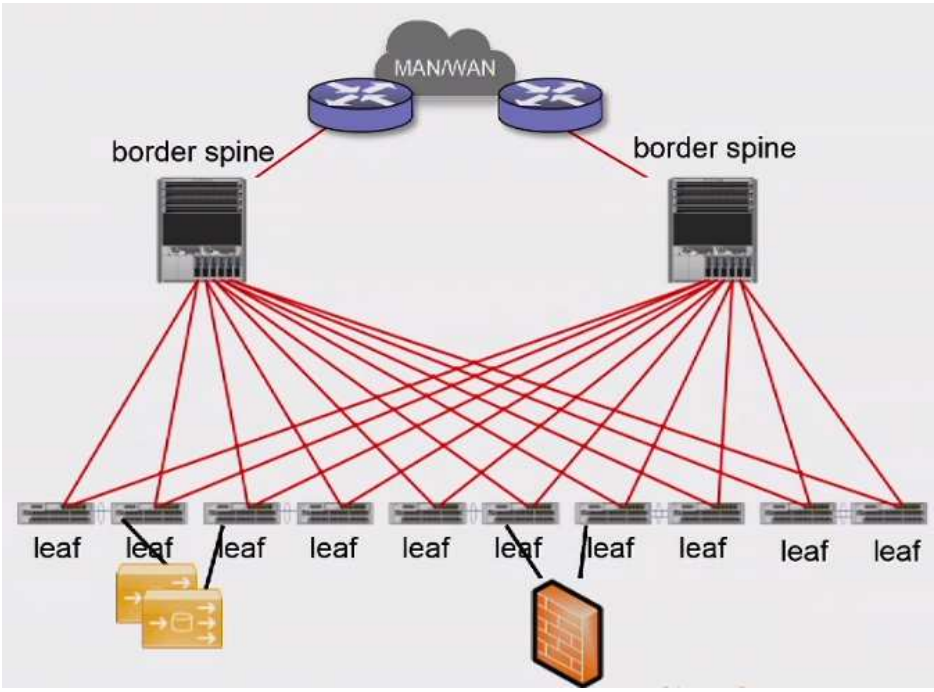


Figura 5. Spine de borde

2.2.3. VXLAN Tunnel Endpoint

VXLAN utiliza los dispositivos VXLAN Tunnel Endpoint (VTEP) para realizar la encapsulación y desencapsulación VXLAN. Cada VTEP tiene dos interfaces: una interfaz en el segmento LAN local para soportar la comunicación de punto final local mediante bridging y la otra interfaz para la red IP de transporte.

La interfaz que va a la red de transporte tiene una dirección IP única que identifica el dispositivo VTEP conocida como infraestructura VLAN. El dispositivo VTEP utiliza esta dirección IP para encapsular tramas Ethernet y transmite los paquetes encapsulados a la red de transporte a través de la interfaz. Un dispositivo VTEP también descubre otros VTEP remotos para sus segmentos VXLAN y aprende direcciones MAC remotas hacia mapeos VXLAN a través de ésta interfaz. Los componentes funcionales de los VTEP y la topología lógica que se crea para la conectividad de capa 2 a través de la red IP de transporte se muestra en la figura 6.

2.3. Funcionamiento de VXLAN

2.3.1. Plano de Datos

VXLAN requiere de una infraestructura de red (red underlay) para llevar a cabo el data plane forwarding.

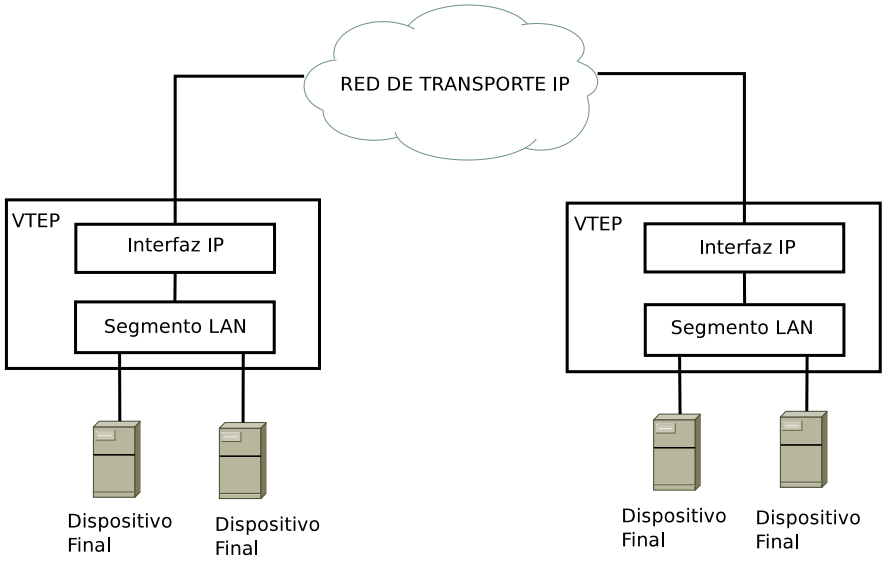


Figura 6. VXLAN Tunnel Endpoint

Éste último es requerido para proporcionar comunicación unicast entre los dispositivos finales conectados a la VXLAN Fabric. La figura 7 ilustra el data plane forwarding en una red que usa VXLAN.

Al mismo tiempo, la infraestructura de red puede usarse para enviar tráfico multidestino a dispositivos finales conectados a un dominio de broadcast común en capa 2 en la red overlay. Con frecuencia, este tráfico es conocido como **BUM**, el cual incluye tráfico de broadcast, unknown unicast y multicast.

Existen dos enfoques diferentes para permitir el tráfico BUM a través de VXLAN Fabric:

1. Hacer uso multicast en la red overlay (Protocolo independiente Multicast o PIM), para hacer uso de las capacidades de replicación nativa de los SPINES para distribuir tráfico a los dispositivos VTEP.

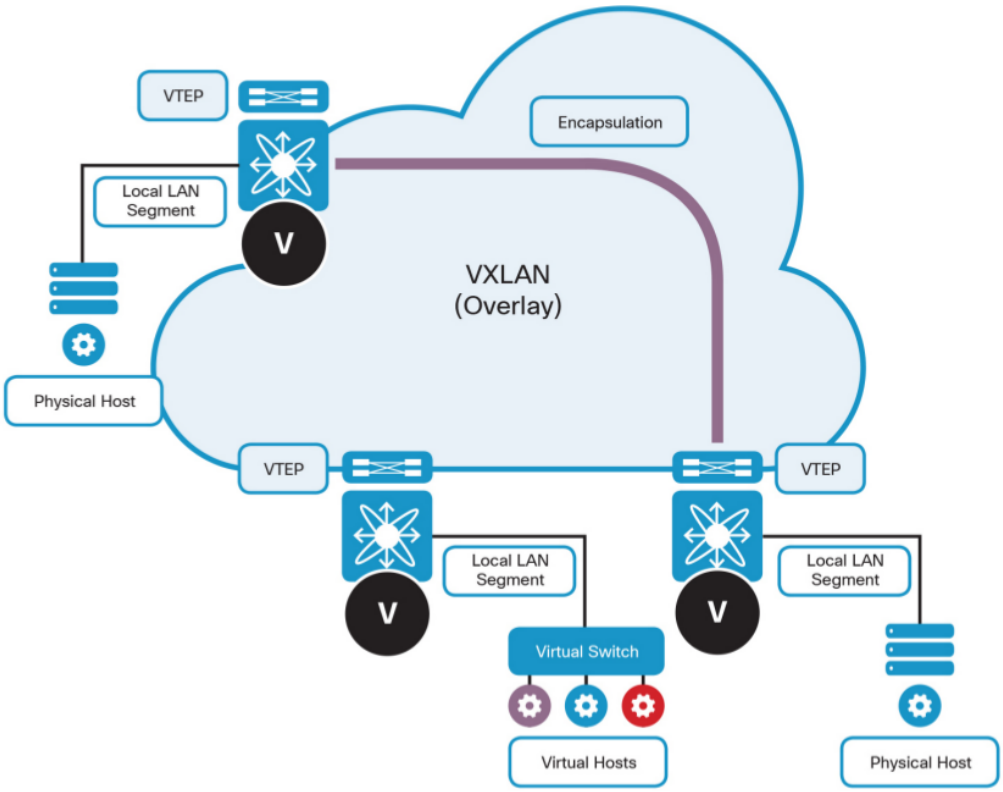


Figura 7. Jansen,D.(2017).VXLAN Overlay Network.[Figura].Recuperado de A Modern, Open and Scalable Fabric VXLAN EVPN

2. En escenarios donde no se puede implementar multicast, es posible hacer uso de las capacidades de replicación de recursos de los VTEP que crean múltiples copias de tramas tipo BUM que son enviadas a cada VTEP remoto. Este enfoque no es tan eficiente como usar multicast para la replicación de tráfico tipo BUM. VXLAN no cambia la semántica para el forwarding de capa 2 o capa 3 y permite al VTEP llevar a cabo funciones de routing y bridging mientras hace uso del entunelamiento VXLAN para el data plane forwarding.

Como tal, el VTEP ofrece un conjunto de funciones de gateway, que se detallan en la figura 8.

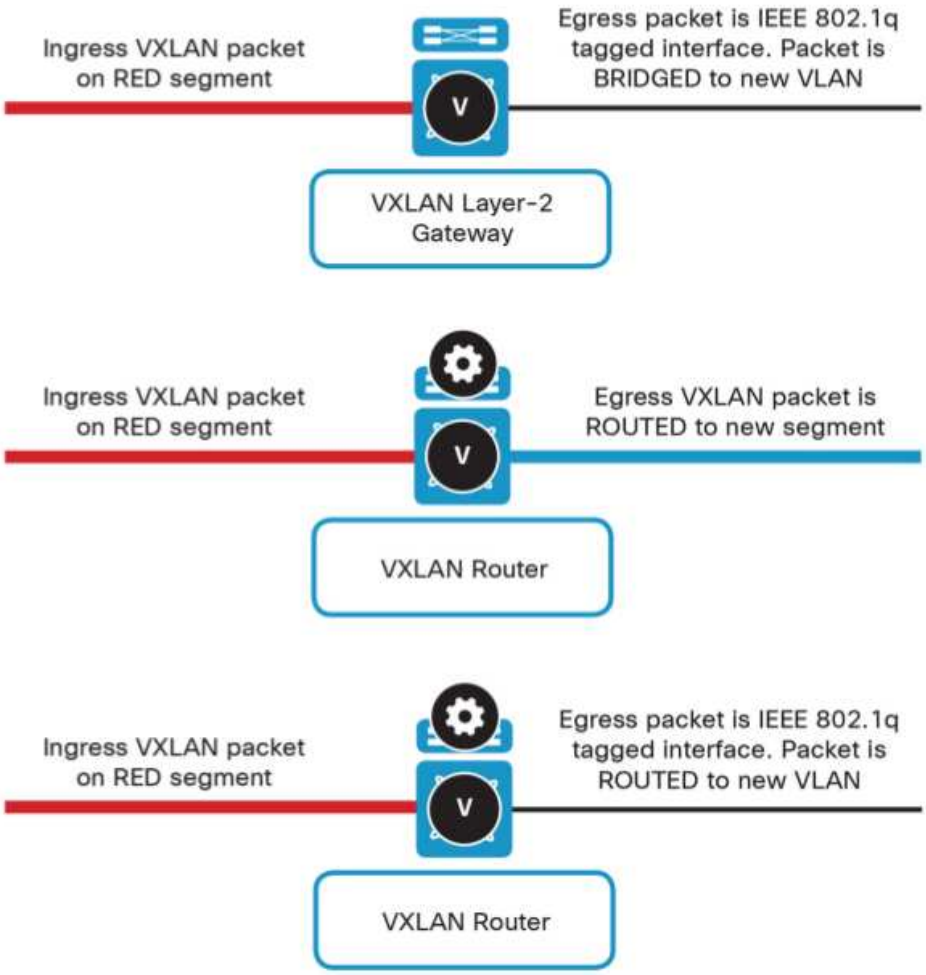


Figura 8. Jansen,D.(2017).VXLAN Gateway Functions.[Figura].Recuperado de A Modern, Open and Scalable Fabric VXLAN EVPN

- **Gateway de capa 2:** Puenteo de VXLAN a VLAN mapeando un segmento VNI a una VLAN usando un bridge-domain común.
- **Gateway de capa 3 (VXLAN Router):** Enrutamiento de VXLAN a VXLAN brindando conectividad de capa 3 entre dos VNI por lo que no se requiere ninguna función de decapsulación.
- **Gateway de capa 3 (VXLAN Router):** Enrutamiento de VXLAN a VLAN brindando conectividad de capa 3 entre un VNI y una VLAN.

2.3.2. Plano de Control

El plano de control, o método por el cual la disponibilidad y aprendizaje de VXLAN ocurre, se logró a través de lo que se conoce como comportamiento de inundación y aprendizaje. En palabras simples, inundación y aprendizaje es un método de análisis de datos en el que un VTEP que no conoce la localización de una determinada MAC de destino envía tramas a los grupos de multicast asociados a VXLAN. Multicast se usa para proporcionar un enfoque más administrable para el manejo de tráfico multidestino.

En lugar de aprender de la interfaz de origen asociada con la dirección MAC de origen, el host aprende la dirección IP de origen encapsulada del VTEP remoto. La metodología de inundación y aprendizaje se interesa tanto en el descubrimiento de VTEPs (entre pares) como al aprendizaje de la localización de dispositivos finales remotos.

Mientras que la metodología de inundación y aprendizaje presenta un nivel de seguridad muy bajo para la entrada de vendors que deseen implementar VXLAN, la desventaja más importante de ésta es la escalabilidad. La cantidad de tráfico multicast adicional introducido puede ser difícil de predecir y como tal ha sido una barrera para la adopción de algunos clientes empresariales.

En lo referente a problemas de escalabilidad, el concepto de plano de control para administrar el aprendizaje de direcciones MAC y descubrimiento de

VTEPs vecinos es deseable, y preferiblemente podría estar basado en protocolos que son generalmente bien entendidos. Multi-Protocol Gateway Protocol (MP-BGP) con Ethernet Virtual Private Network (EVPN) han sido planteados como el estándar IETF para el plano de control de VXLAN. Basado en el protocolo MP-BGP estándar, el plano de control MP-BGP EVPN brinda distribución de información de protocolo basado en el descubrimiento de VTEPs vecinos y conectividad de los dispositivos finales que permite mayor escalabilidad en el diseño de redes VXLAN overlay.

El plano de control MP-BGP EVPN introduce una serie de características que reduce la cantidad de tráfico inundado en la red overlay y habilita un óptimo envío de tráfico. Como información relevante en referencia al caso de uso de los Data Centers, EVPN provee información de conectividad para dispositivos finales a nivel de capa 2 y capa 3. Extendiendo este nivel de conectividad y añadiendo la capacidad de supresión ARP (arp suppression), se reduce la cantidad requerida de inundaciones de tráfico en la red. Un beneficio adicional del plano de control EVPN es que brinda descubrimiento de VTEPs vecinos y autenticación, mitigando el riesgo de VTEPs dudosos en la red VXLAN overlay.

Para entender la funcionalidad de MP-BGP EVPN, es necesario comprender el uso de MP-BGP en las redes MPLS. Una red tradicional MPLS tiene una estructura full mesh de routers BGP y route reflectors para balancear el intercambio de información y conectividad de una red L3VPN (o L2VPN en

el caso de VPLS[1]). La combinación de route distinguishers (RD) y direcciones VPNV4 aseguran la capacidad de identificar un ùnico objetivo, y rutas que pueden ser aprendidas selectivamente usando el filtrado a través de los route targets (RT). En el plano de control EVPN, técnicamente hay tres opciones de plano de datos:

- Multi-Protocol Label Switching (MPLS, draft-ietf-l2vpn-evpn)
- Provider Backbone Bridging (PBB,draft-ietf-l2vpn-pbb-evpn)
- Network Virtualization Overlay (NVO, draft-ietf-bess-evpn-overlay).

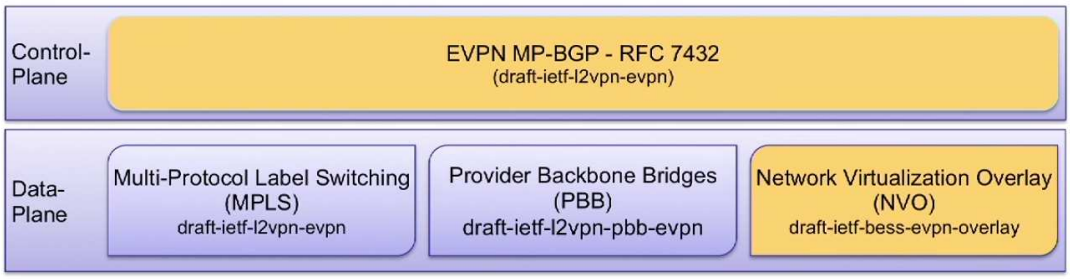


Figura 9. Bosquejo para EVPN IETF

2.4. Consideraciones de Networking para VXLAN

En las tradicionales redes de acceso de capa 2, el default gateway de capa 3 es comúnmente colocado en la capa de agregación. Generalmente, los switches de agregación hacen uso de protocolos de redundancia de primer

[1]Virtual Private LAN Services

salto tales como HSRP [2], VRRP [3] o GLBP [4] para proveer una dirección IP de default gateway redundante. Dependiendo de la configuración y del protocolo, pueden configurarlos como active/standby o active/active.

Con el crecimiento de la virtualización en los data centers, el diseño físico de la red y su representación lógica son notablemente diferentes. La virtualización fomenta la movilidad y ésto presenta ineficiencias ya que el default gateway fue declarado sobre la localización física de los recursos de red. El reenvío de tráfico continúa funcionando, sin embargo, la ineficiencia intrínseca creada por tráfico hair-pinning (NAT Loopback) no es óptimo.

2.4.1. Distributed Anycast Gateway

El uso de MP-BGP EVPN presenta la funcionalidad de distributed anycast gateway. La función de default gateway es distribuida a través de todos los nodos leaf en la infraestructura VXLAN. Con ésta funcionalidad se provee una mejor eficiencia y un alto ancho de banda transversal mientras se elimina la necesidad de un protocolo de redundancia. Además, el tráfico enrutado entre los dispositivos conectados al mismo leaf es localmente reenviado sin tener que enviarlo a los spines. Decrementando el conteo de saltos, se reduce también la latencia en la red.

---

[2] Hot Standby Router Protocol

[3] Virtual Router Redundancy Protocol

[4] Gateway Load Balancing Protocol

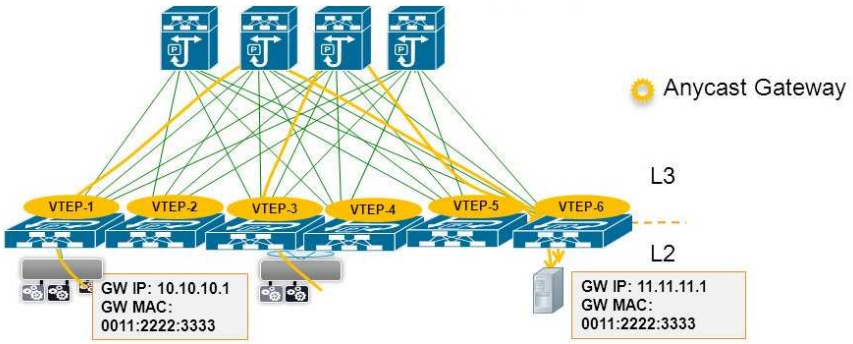


Figura 10. Distributed Anycast Gateway

2.4.2. Integrated Routing and Bridging

VXLAN presenta la funcionalidad de Integrated Routing and Bridging (IRB), la cual ofrece la capacidad de reenvío de tráfico a nivel de capa 2 y capa 3 directamente a un dispositivo leaf.

El bosquejo de EVPN define dos diferentes métodos de enrutamiento de tráfico entre redes VXLAN Overlay. El primer método se conoce como IRB asimétrico y el segundo método como IRB simétrico.

En el ***IRB asimétrico***, el VTEP de entrada lleva a cabo funciones de routing y bridging, mientras que el VTEP de salida solamente de bridging. Como resultado, el tráfico de retorno tomará un diferente VNI que el tráfico de origen. Esto requiere que los VNI de origen y destino se encuentren tanto en el VTEP de entrada como en el VTEP de salida. Esto conduce a una configuración más compleja ya que en todo los switches se debe instanciar todos los posibles VNI. Tal vez una consideración más urgente es la implicación de

escalabilidad de todos los dispositivos requeridos para obtener un número considerablemente mayor de dispositivos finales.

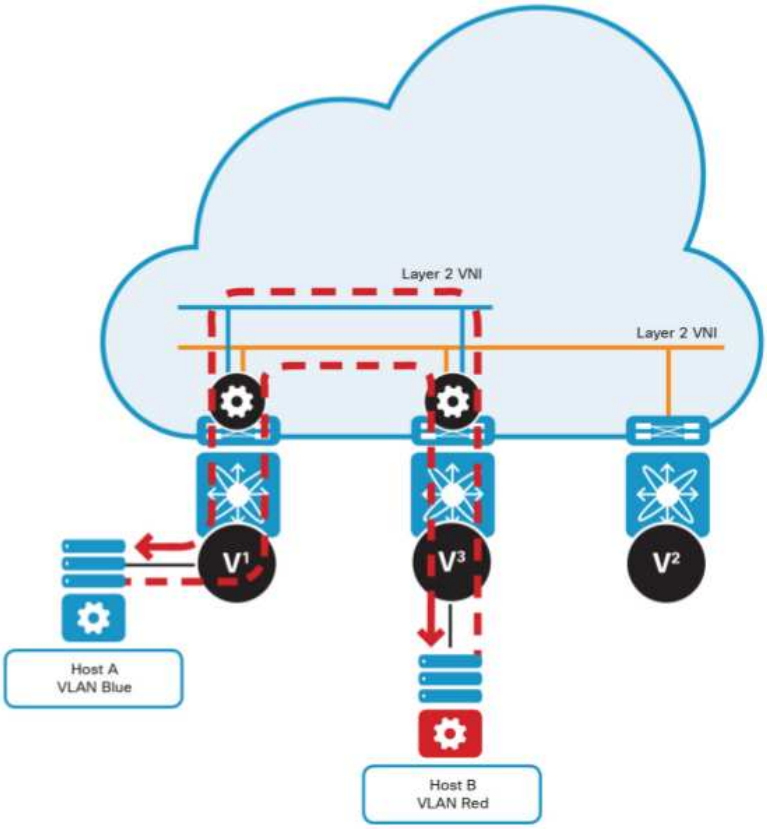


Figura 11. Jansen,D.(2017). Asymmetric IRB.[Figura]. Recuperado de A Modern, Open and Scalable Fabric VXLAN EVPN

En el ***IRB simétrico***, tanto el VTEP de entrada como el de salida proveen reenvío de tráfico en capa 2 y capa 3. Esto resulta en un comportamiento previsible de reenvío de tráfico y solamente los VNI de los dispositivos finales conectados localmente necesitan ser definidos en los VTEP (más el tráfico L3 VNI) que a su vez simplifica la configuración y reduce los requerimientos de escalabilidad a través de la optimización del uso de las tablas ARP y MAC.

Esto da como resultado una mejor escalabilidad en términos de un número total de VNI que una infraestructura VXLAN Fabric puede soportar.

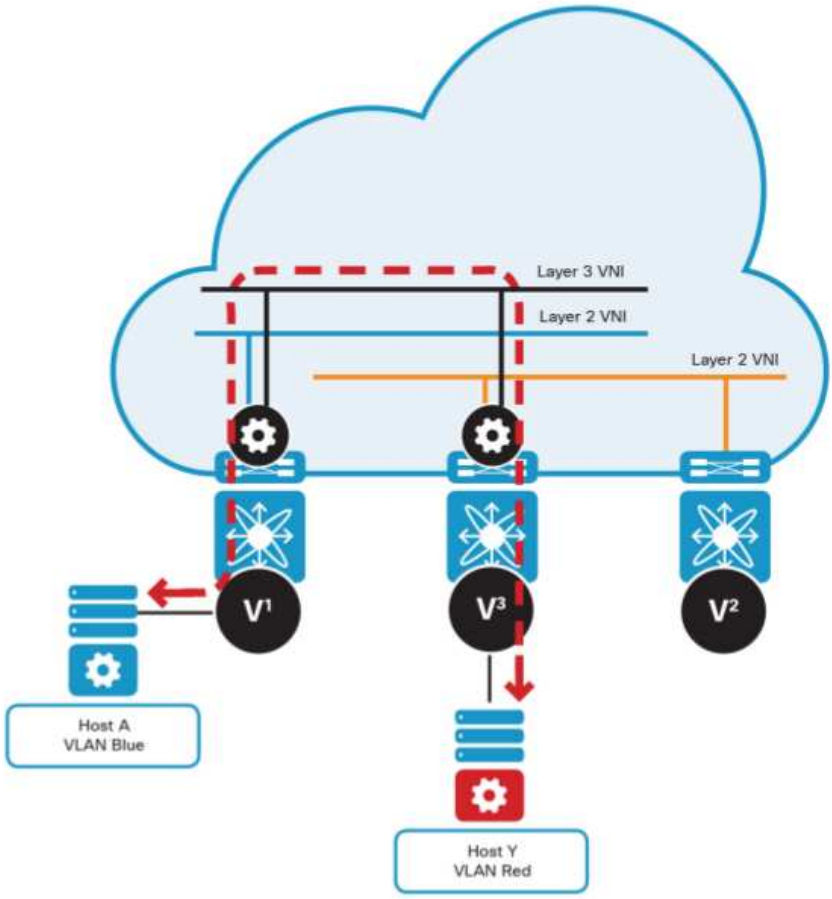


Figura 12. Jansen,D.(2017). Symmetric IRB.[Figura]. Recuperado de A Modern, Open and Scalable Fabric VXLAN EVPN

2.5. Proceso de encapsulación y desencapsulación en VXLAN

Se consideró el siguiente escenario (Ver figura 13):

Existen dos máquinas virtuales, VMA y VMB que están físicamente separadas por una red de transporte de capa 3. Éstas máquinas necesitan conectarse a la misma VLAN y utilizar el mismo segmento de red IPV4. En un esquema

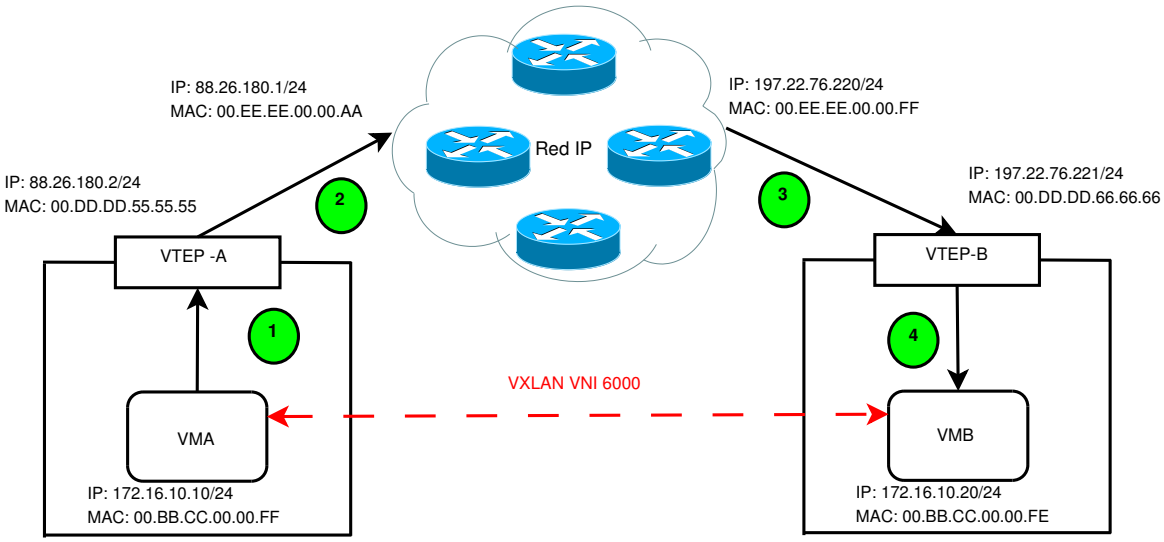


Figura 13. Topología para la demostración de encapsulación VXLAN

de Networking tradicional muy probablemente se declare la inviabilidad de este requerimiento.

Las dos máquinas virtuales pueden trabajar en la misma subred empleando VXLAN como red de transporte overlay, aún cuando ambos dominios de capa 2 están separados por múltiples dispositivos de capa 3. Veamos entonces, cómo es el proceso de encapsulación y desencapsulación en VXLAN.

1. VMA envía un paquete a la IP 172.16.10.20 (VMB), este paquete es encapsulado en una trama Ethernet estándar con los siguientes parámetros:

   - **IP de destino:** 172.16.10.20 (VMB)
   - **IP de origen :** 172.16.10.10 (VMA)
   - **MAC de destino:** 00:BB:CC:00:00:FE (VMB)

- **MAC de origen:** 00:BB:CC:00:00:FF (VMA)

Esta trama es enviada a su respectivo VTEP, en este caso VTEP-A y la llamaremos trama interna Ethernet (*inner ethernet*). Se asume que el paquete IPV4 contenido en esta trama transporta datos de una sesión TELNET entre ambas máquinas virtuales (es decir, el paquete IP transporta un segmento TCP + datos de aplicación de tipo TELNET).

A modo de ejemplo, la longitud de los datos de aplicación será de 90 bytes y del encabezado TCP de 32 bytes (se asume, además, que TCP utiliza sus campos opcionales ya que la longitud el mínima del encabezado TCP es de 20 bytes y la máxima puede llegar hasta 60). Entonces, VMA envía un segmento de 122 bytes de longitud transportado dentro de un paquete IP, el cual agrega 20 bytes extra que corresponden al tamaño del encabezado IPv4.

Hasta el momento tenemos un paquete IPV4 de 142 bytes siendo encapsulado dentro de una trama Ethernet estándar transportándose desde VMA hacia VTEP-A. La trama Ethernet interna tiene un encabezado normal de 14 bytes más un trailer de 4 bytes para FCS. En total suman 160 bytes en la trama Ethernet interna.

2. Una vez que el mensaje llega a VTEP-A comienza el proceso de encapsulación VXLAN. Este dispositivo añade un encabezado de 8 bytes que contiene un VNI o VXLAN Network Identifier con el valor 6000 (puede

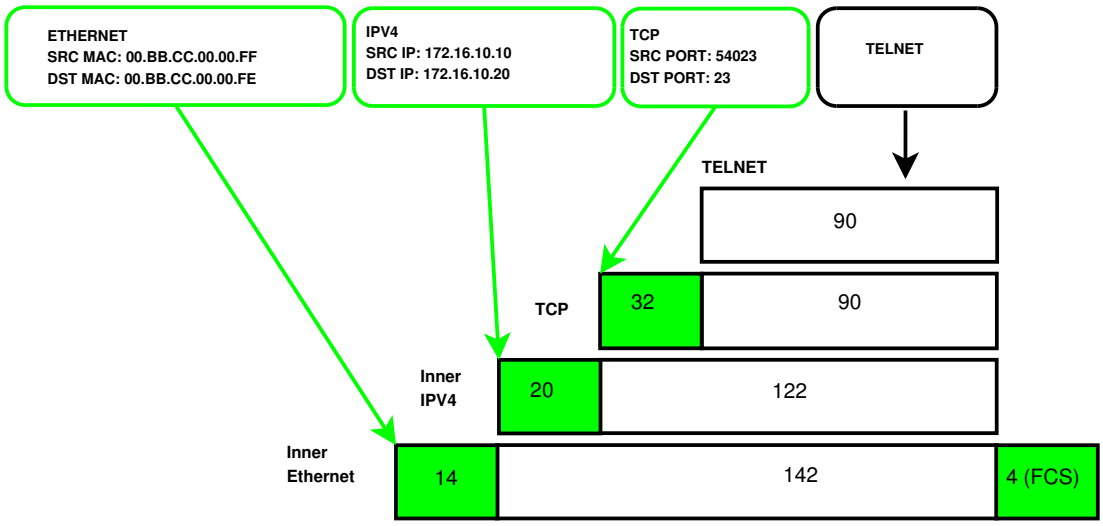


Figura 14. Trama Inner Ethernet

llegar a 2^{24}). Se descarta el FCS ya que no es necesario para el transporte de extremo a extremo. A la trama de 14 bytes + 142 (sin FCS) se le suman los 8 bytes de VXLAN.

VTEP-A transforma esa trama Ethernet interna, (incluyendo el encabezado VXLAN) en la carga de un paquete UDP. A este paquete UDP, al cual denominaremos "***UDP externo***", se le agrega un encabezado IPv4 normal y nuevamente se encapsula en una trama Ethernet externa (*outer ethernet*) (Ver figura 15)

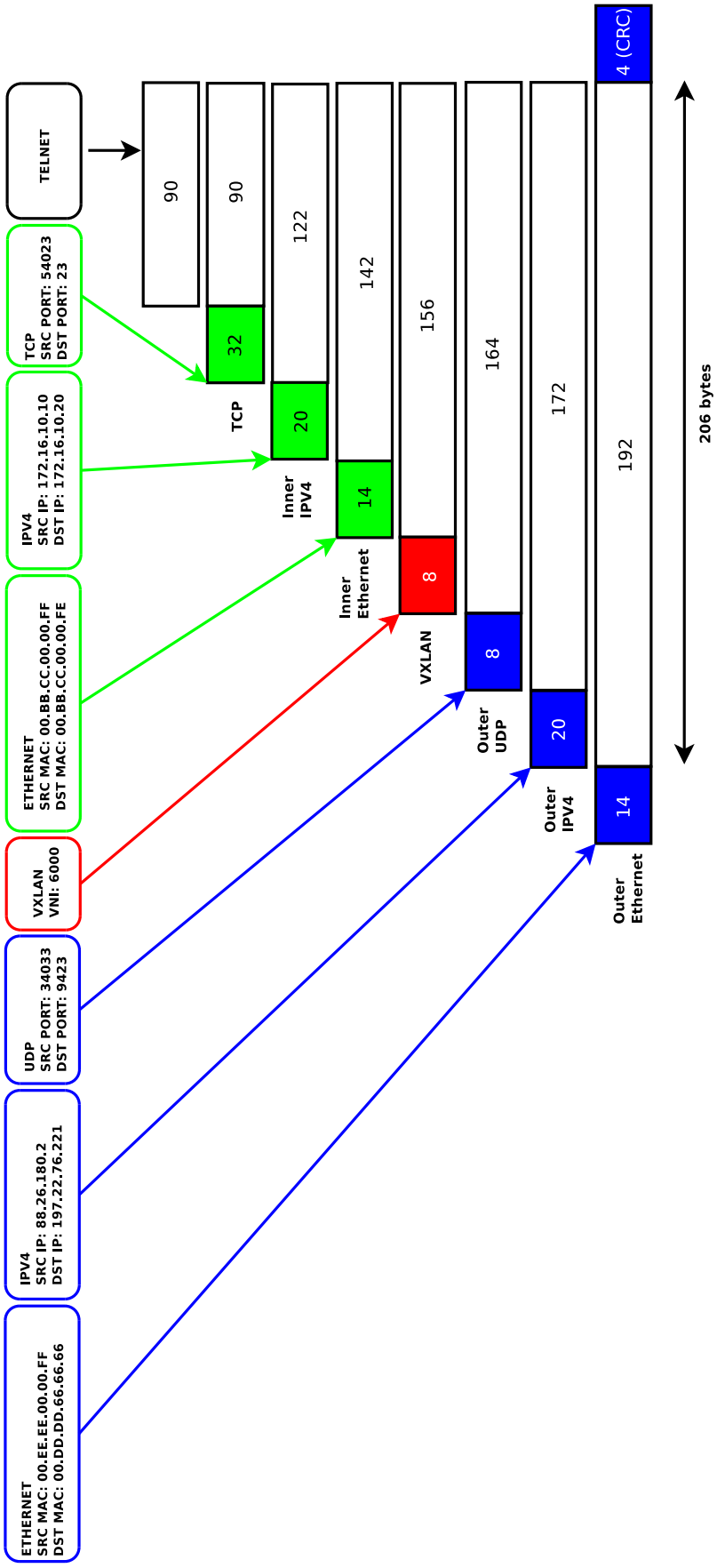


Figura 15. Trama Outer Ethernet

Haciendo una retroalimentación de lo anterior, hasta el momento tenemos lo siguiente:

- Se envía una trama Ethernet normal desde VMA hacia VTEP-A que contiene un paquete IP, que a su vez contiene un segmento TCP con datos de aplicación de tipo TELNET. Todo esto mide 14 +142 +4 (Encabezado Ethernet + Paquete IP completo + FCS).
- Al llegar a VTEP-A se descarta el FCS y se toma esta trama Ethernet (ahora solo de 14+142 bytes) y se le agrega un encabezado VXLAN de 8 bytes que contiene el valor de identificador de VXLAN ó VNI). Tenemos 8 + 156 = 164 bytes.
- Estos 164 bytes se almacenan como payload dentro de un mensaje UDP, el cual agrega 8 bytes más (el encabezado UDP siempre es de 8 bytes). Este es el paquete UDP externo u outer UDP.
- Se agrega además un encabezado IPv4 (20 bytes) con los valores de IP de origen y destino de los VTEPs.
- Como ese paquete IP necesita una trama para transportarse en su red local, se encapsula dentro de una trama Ethernet estándar (outer Ethernet) la cual agrega 14 bytes de encabezado + 4 bytes de trailer. Todo esto suma un total de 206 bytes.

Esta trama sale del VTEP-A hacia el siguiente router dentro de la red de backbone y es tratado como un paquete IP normal (debido al encabezado IP externo), enrutándose de acuerdo a su IP de destino. La red de backbone encaminará este paquete hacia VTEP-B.

La trama Ethernet externa que sale de VTEP-A lleva los siguientes parámetros:

- **MAC de destino:** 00:EE:EE:00:00:AA (Correspondiente al default gateway de VTEP-A)
- **MAC de origen:** 00:DD:DD:55:55:55 (VTEP-A)

Mientras que el paquete IPV4 externo, utilizado para llegar a VTEP-B contiene:

- **IP de destino:** 197.22.76.221 (VTEP-B)
- **IP de origen:** 88.26.180.2 (VTEP-A)

3. Antes de llegar a VTEP-B, verifiquemos el paquete IP externo y la trama que lo encapsula (Ver figura 16)

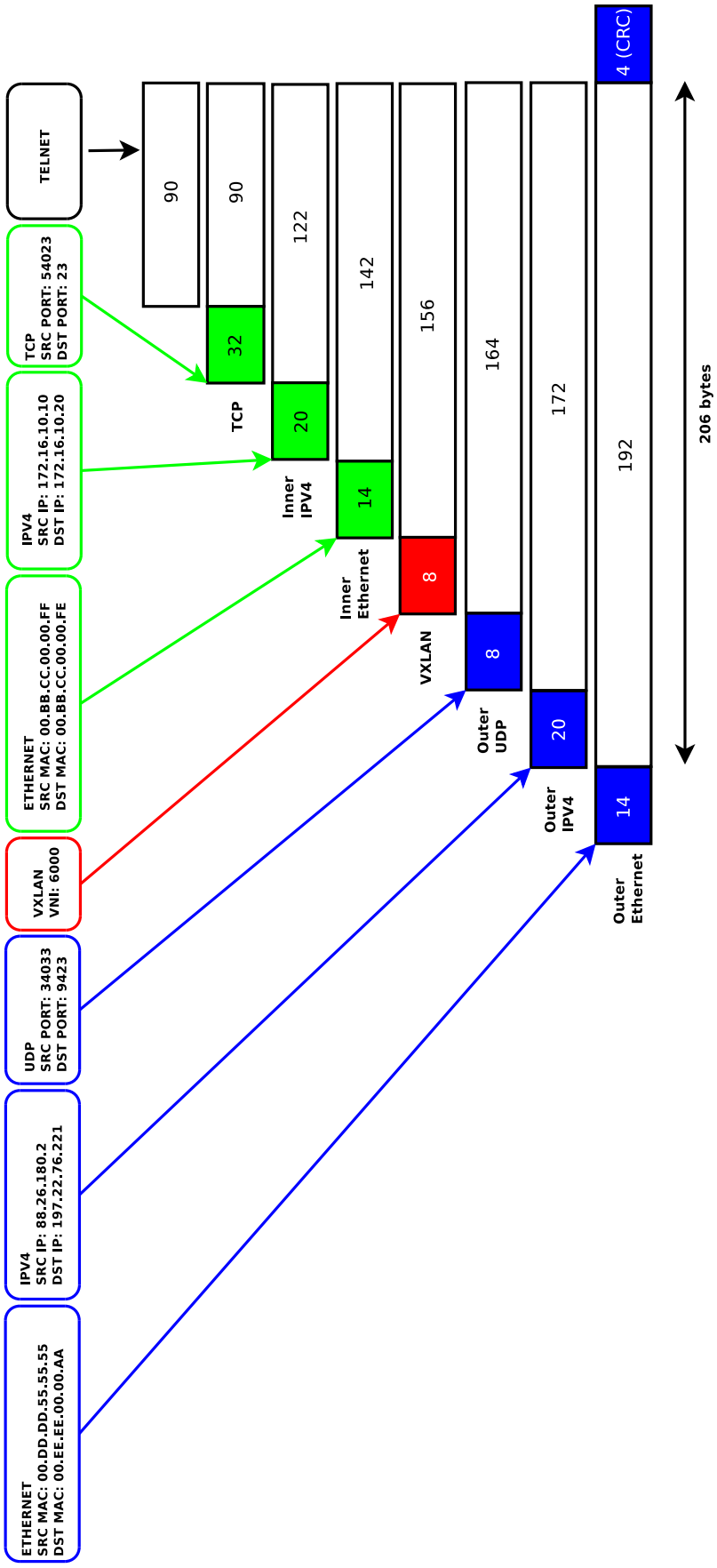


Figura 16. Trama Outer Ethernet antes de llegar a VTEP-B

Como se puede apreciar, en el paquete IP externo la información que cambia es la dirección MAC de origen y destino entre el último router y VTEP-B.

4. VTEP-2 recibe la trama externa y realiza la desencapsulación de la misma, retira el paquete IPv4 externo y también el encabezado UDP externo para verificar el VNI. Su configuración interna indica que el VNI 6000 solo puede conectarse a VMB. Ahora que VTEP-B eliminó todos los valores externos,solamente queda la trama interna Ethernet que se originó en VMA. Con esta información, la trama interna se mueve hacia VMB tal como si estuviese conectada en la misma LAN física.

   El contenido de la trama Ethernet interna es el mismo que en el punto 1. Se le agrega el FCS para respetar el estándar y realizar la verificación de errores.

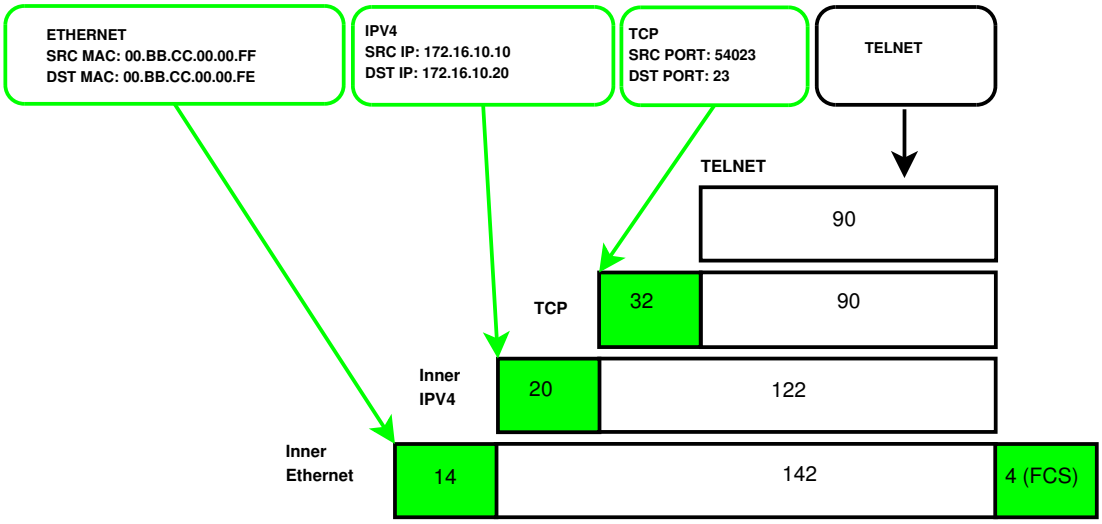


Figura 17. Trama desencapsulada luego de pasar por VTEP-B

Por último, la trama generada en VMA llega a VMB, el cual desencapsula nuevamente esta trama interna, retira el encabezado TCP y verifica los datos de aplicación. Para enviar información de VMB a VMA se realizará todo el proceso descrito anteriormente.

Según Colomés (2015), esta habilidad permite a la infraestructura gestionar redes multi-tenants al ser capaz de encapsular millones de VNIs dentro de un paquete UDP normal, soportando a la vez tanto direcciones IP como MACs sobrepuestas (overlapped) entre un cliente y otro (no dentro del mismo VNI, por supuesto).

CAPÍTULO III

ANÁLISIS DE BGP EVPN Y MEJORAS EN VXLAN

3.1. EVPN

Cuando se implementa una tecnología overlay , hay tres grandes tareas que deben cumplirse:

- Primero, debe haber un mecanismo de reenvío de paquetes.
- Segundo, debe haber un plano de control donde se pueda buscar la ubicación de un dispositivo o una aplicación y el resultado se utilice para encapsular el paquete para que pueda ser reenviado a su destino.
- Tercero, debe haber una manera de actualizar el plano de control de manera que sea siempre exacto. Al tener información errónea en el plano de control, los paquetes se enviarían a la ubicación incorrecta y probablemente sean descartados.

La primera tarea, el reenvío de paquetes es algo que los dispositivos de networking siempre han realizado. Rendimiento, costo, confiabilidad y soportabilidad son consideraciones fundamentales para una red y deben aplicarse igualmente tanto a las redes underlay como a las redes overlay.

La segunda tarea, búsqueda de plano de control y encapsulación, es realmente un problema de rendimiento y capacidad. Si estas funciones son llevadas a cabo mediante un software, consumiría bastantes recursos de procesamiento y añadirían latencia al compararlos con soluciones de hardware.

El tercer componente es el medio por el cual las modificaciones del plano de control son actualizadas a través de los dispositivos de red. Esta actualización es realmente un desafío y concierne más a los administradores de un Data Center debido al potencial impacto que tendría para las aplicaciones debido a la pérdida de paquetes si el plano de control funciona mal.

3.1.1. Plano de Control de VXLAN

VXLAN es una tecnología overlay que no proporciona muchos de los mecanismos para escalabilidad y tolerancia a fallos que otras tecnologías de redes han desarrollado y han asumido. En una red VXLAN, cada switch crea una base de datos con los hosts directamente conectados. Un mecanismo es requerido para que otros switches aprendan acerca de estos hosts. En una red tradicional, no hay un mecanismo para distribuir esta información, el único plano de control disponible es el que se conoce como inundación y aprendizaje (flood and learn). Para que un host sea alcanzable, esta información era inundada a través de toda la red. Las redes Ethernet han operado con esta deficiencia por décadas.

Con la necesidad de incremento de redes escalables, el efecto de inundación y aprendizaje necesitaba ser mitigado. Para una red VXLAN, un plano de control debe ser capaz de distribuir conectividad a nivel de capa 2 y capa 3 a través de la red. Las primeras implementaciones de VXLAN carecían de la capacidad de transmitir información de conectividad a nivel de capa 2, por tanto, extensiones de Ethernet VPN fueron añadidas al multi protocolo BGP para transportar esta información.

3.1.2. Multiprotocolo BGP (MP-BGP)

MP-BGP es una extensión de BGP, basado en el RFC 4760. Este soporta distintos tipos de familias de direcciones tales como VPNV4, VPNV6, L2VPN EVPN. MVPN. Con respecto a la familia de direcciones L2VPN EVPN, ésta es la que permite usar VXLAN junto con EVPN. Una ventaja muy importante de MP-BGP es que permite transportar múltiples tipos de información a través de una vecindad BGP (BGP peering).

En una infraestructura VXLAN dentro de un Sistema Autónomo (AS) único se establecen vecindades BGP a través de internal BGP (iBGP). iBGP se emplea para intercambiar información entre las todas las vecindades BGP mediante un modelo síncrono a través de todo el AS. iBGP requiere de una estructura full mesh para el intercambio de información, lo cual es desventajoso debido a la cantidad de enlaces que se requieren, para cambiar este comportamiento se utiliza Route Reflectors para simplificar e incrementar

la escalabilidad de las vecindades de BGP. Al usar route reflectors, se reduce la cantidad de vecindades para un switch leaf.

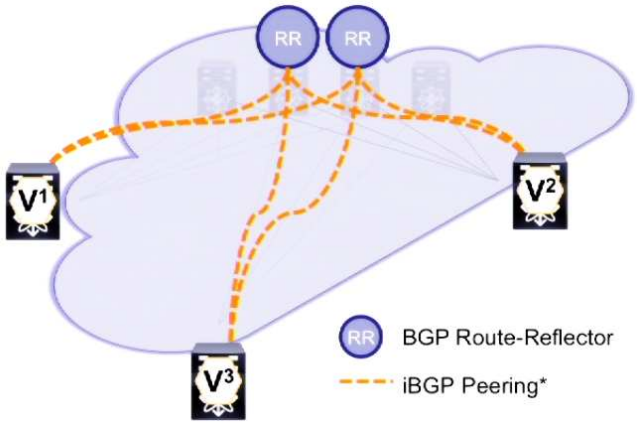


Figura 18. BGP Route Reflectors

Con BGP externo (eBGP) el establecimiento de vecindades BGP se logra entre diferentes sistemas autónomos (AS). Esto significa, por ejemplo, que un bgp speaker que se encuentra en el AS 65000 puede establecer vecindad con un bgp speaker que se encuentra en el AS 65003 (Ver figura 19). El intercambio de rutas recibidas por un BGP speaker es enviado a todos sus vecinos.

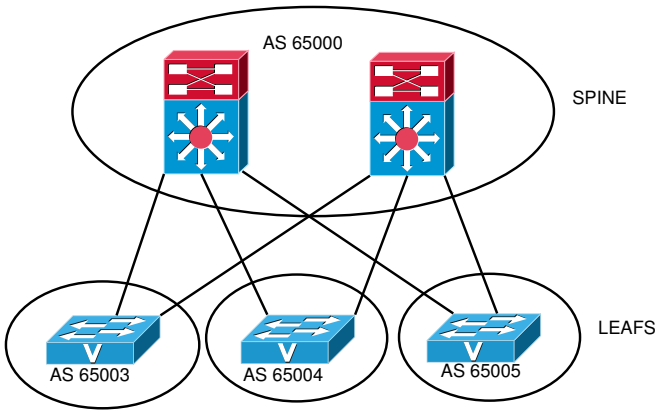


Figura 19. eBGP sin Route Reflectors

3.1.3. Route Distinguisher

El route distinguisher (RD) es un valor de 64 bits configurado para hacer único el direccionamiento de un prefijo VPN a lo largo de una red. Como se ilustra en la figura 20, el primer formato es el de tipo 0 donde dos bytes corresponden al sistema autónomo seguido de un número individual de 4 bytes.

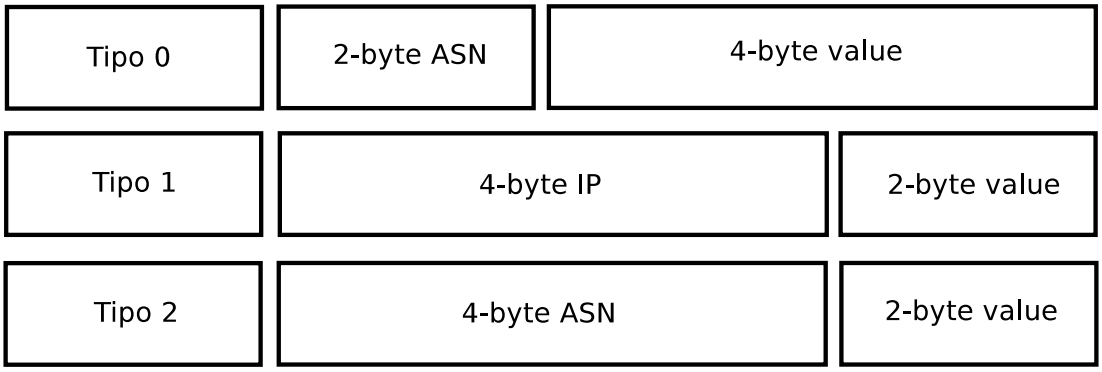


Figura 20. Formatos y tipos de route distinguisher

Para los tipos 1 y 2, la primera parte es similar en tamaño (4 bytes) pero su contenido es diferente: dirección IP en el tipo 1 y número de sistema autónomo en el tipo 2. La otra parte corresponde a un número individual de 2 bytes. Como buena prćctica para separación y eficiencia en el manejo de rutas, se usa un único RD por cada VRF [1].

En VXLAN se emplea el uso de RD automatizados, el formato que se utiliza es el de tipo uno, donde los primeros 4 bytes corresponden al Router ID (RID) y los otros 4 bytes a la VRF ID.

[1] Virtual Router and Forwarding

3.1.4. Route Target

Route Target (RT) brinda un conjunto de políticas respecto a cómo podemos seleccionar un prefijo para que sea importado o exportado en el plano de control. Constituye un campo de 8 bytes asociado a los parámetros de una vrf y representa un valor único que se puede adjuntar a un prefijo ya sea para importarlo o exportarlo.

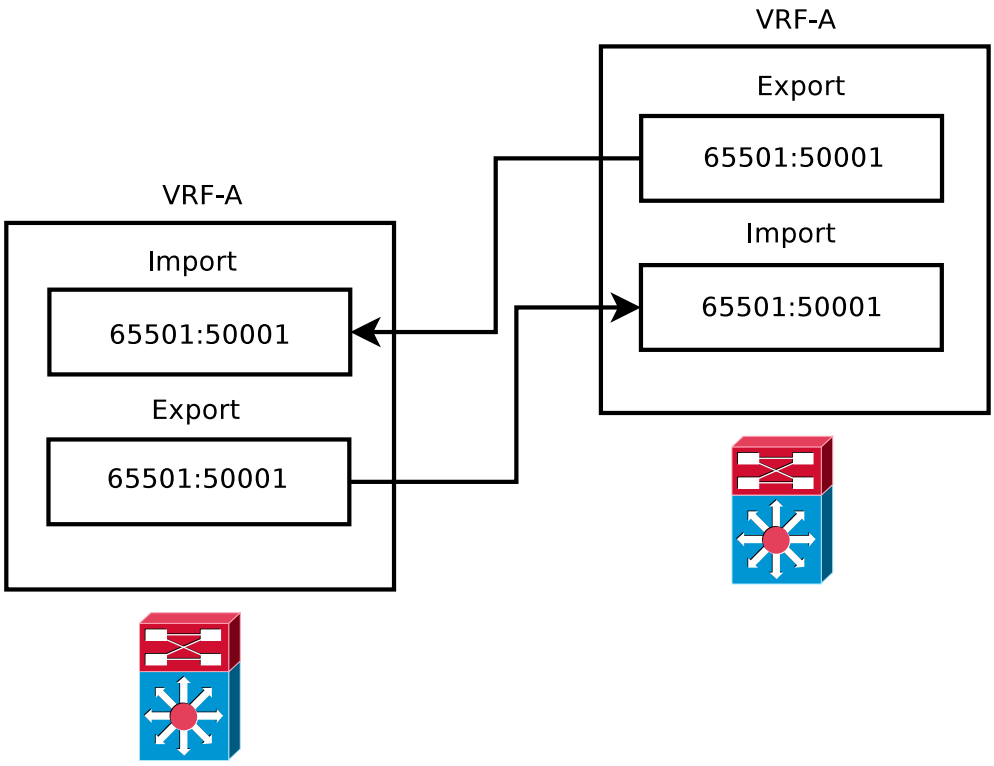


Figura 21. Route Target

3.1.5. Tipos de rutas

MP-BGP EVPN tiene diferentes definiciones para la información de conectividad de la capa de red (NLRI) como parte del RFC 7432. De igual

manera, los tipos de rutas MP-BGP EVPN tienen diferentes definiciones en el RFC 7432 para los tipos 1 a 4 y también para la de tipo 5 en draft-ietf-bess-evpn-prefix-advertisement (vea figura 22). Ruta tipo 1 (Ethernet auto-discovery [A-D] route) y ruta tipo 4 (Ethernet segment route) no están siendo usadas actualmente en la implementación EVPN de Cisco para VXLAN, pero las rutas tipo 2, 3, y 5 son bastante importantes.

RFC\Draft	Tipo de ruta	Descripción
RFC 7432	1	Ethernet auto-discovery (AD)
	2	MAC/IP Advertisement route
	3	Inclusive multicast Ethernet Tag route
	4	Ethernet segment route
draft-ietf-bess-evpn-prefix-advertisement	5	IP prefix route

Figura 22. Tipos de rutas para BGP EVPN

La ruta tipo 2 define el anuncio de rutas MAC/IP y es responsable de la distribución de la información de conectividad de direcciones MAC e IP en BGP EVPN. En la figura 23 se visualiza la estructura de la ruta Tipo 2.

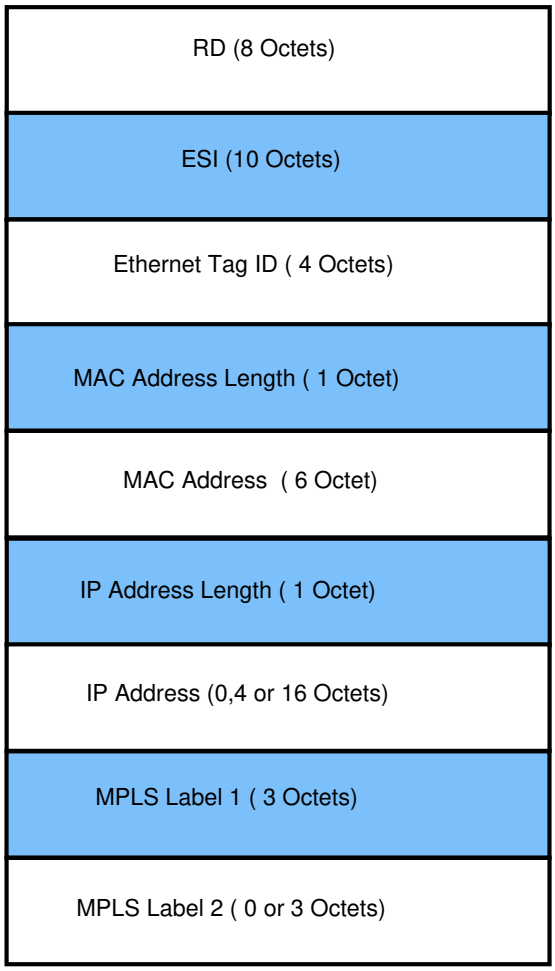


Figura 23. Ruta tipo 2 para BGP EVPN

La ruta tipo 3 es llamada ***inclusive multicast Ethernet tag route*** y es típicamente usada para crear las listas de distribución para el método de envío de datos denominado ingress replication. Este último provee una forma de replicar el tráfico multidestino a través de mensajes unicast. La ruta tipo 3 es inmediatamente generada y enviada a todos los VTEPs que participan en ingress replication tan pronto la VNI es configurada en el VTEP y esté operativa. Es diferente de la ruta tipo 2, la cual es únicamente enviada con información de IP/MAC, cuando los hosts finales se han aprendido. De

esta manera, cada VTEP está consciente de que todos los VTEPs remotos necesitan que se les envíe una copia de un paquete BUM en un determinado VNI. En la figura 24,se muestra la estructura de una ruta de tipo 3.

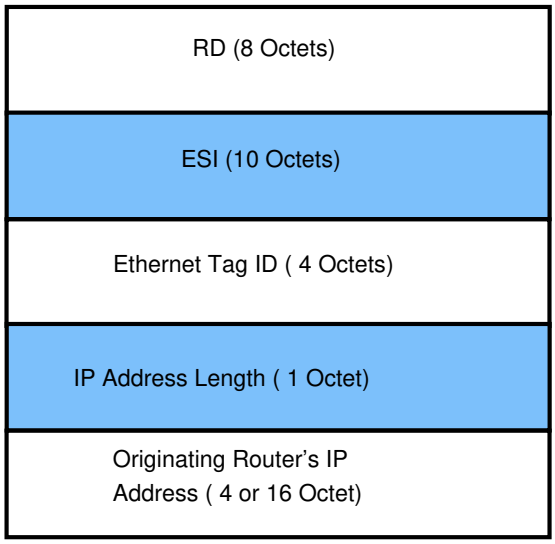


Figura 24. Ruta tipo 3 para BGP EVPN

El tercer tipo de ruta que se indica es la de tipo 5, denominada ***ip prefix route***. La ruta tipo 5 provee la capacidad de transportar información de prefijos IP dentro de EVPN permitiendo el transporte de los prefijos IPv4 e IPv6 con longitud variable (0 a 32 para IPv4 y 0 a 128 para IPv6). Los prefijos IP de rutas en la ruta tipo 5 no contiene información MAC de capa 2 dentro de su NLRI, y por lo tanto la ruta tipo 5 solamente incorpora la VNI de capa 3 necesaria para enrutamiento y multitenancy integrado. Además, las comunidades extendidas de la ruta tipo 5 llevan el route target, tipo de encapsulación, y la MAC del VTEP del siguiente salto en la red overlay. Con una ruta tipo 2, una dirección MAC es usada como identificativo de ruta , mientras que para una

ruta tipo 5, un prefijo.

IP es usado para la identificación de una ruta. Esto permite una separación clara para los dispositivos que manejan BGP EVPN, de este modo se evita cualquier procesamiento relacionado con direcciones MAC para rutas de prefijos IP anunciadas por EVPN (Ruta tipo 5).

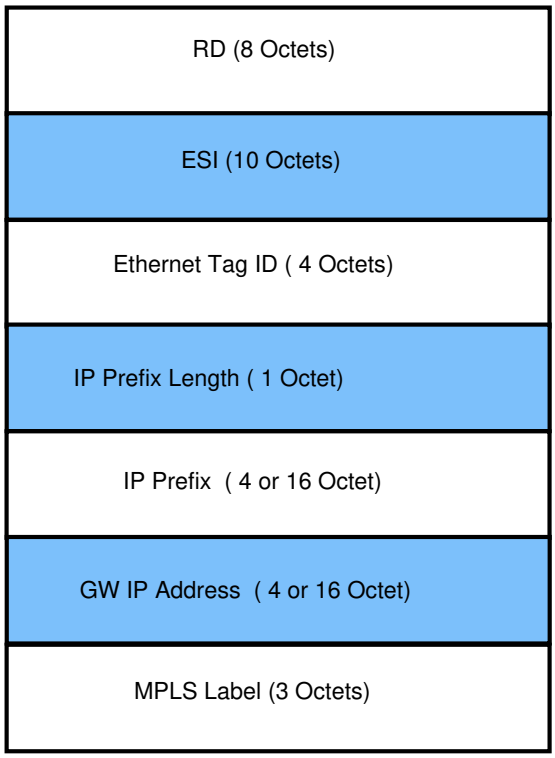


Figura 25. Ruta tipo 5 para BGP EVPN

3.2. Mejoras en VXLAN

VXLAN EVPN presenta varias mejoras, las cuales se encuentran sintetizadas en la figura 26.

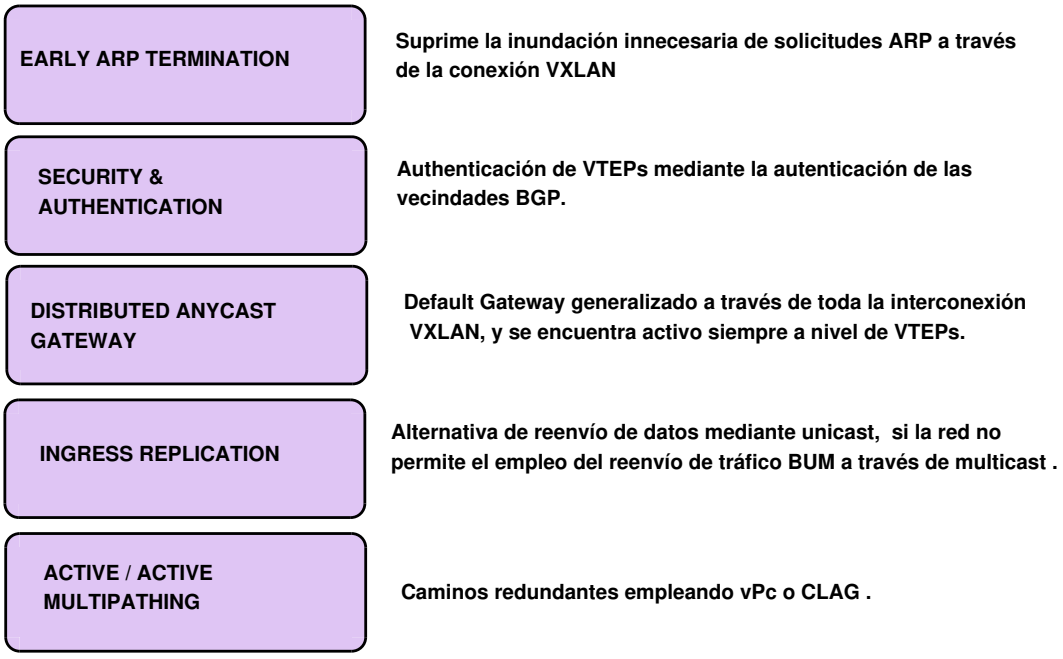


Figura 26. Mejoras en VXLAN

3.2.1. Seguridad Y Autenticación

- La autenticación para VXLAN se establece entre las vecindades BGP, a través de MD5 digest. Esto permitirá proteger las sesiones BGP y contra ataques de seguridad a TCP.

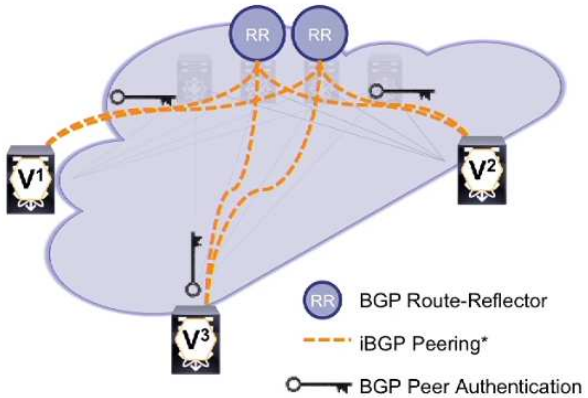


Figura 27. Autenticación BGP

- La autenticación asegura la integridad en la recepción de la información.
- Se puede filtrar prefijos de red mediante listas de control de acceso (ACL), prefix list y listas distribuidas.
- La encapsulación y desencapsulación VXLAN sólo sucede si el VTEP puede autenticarse mediante la sesión BGP.

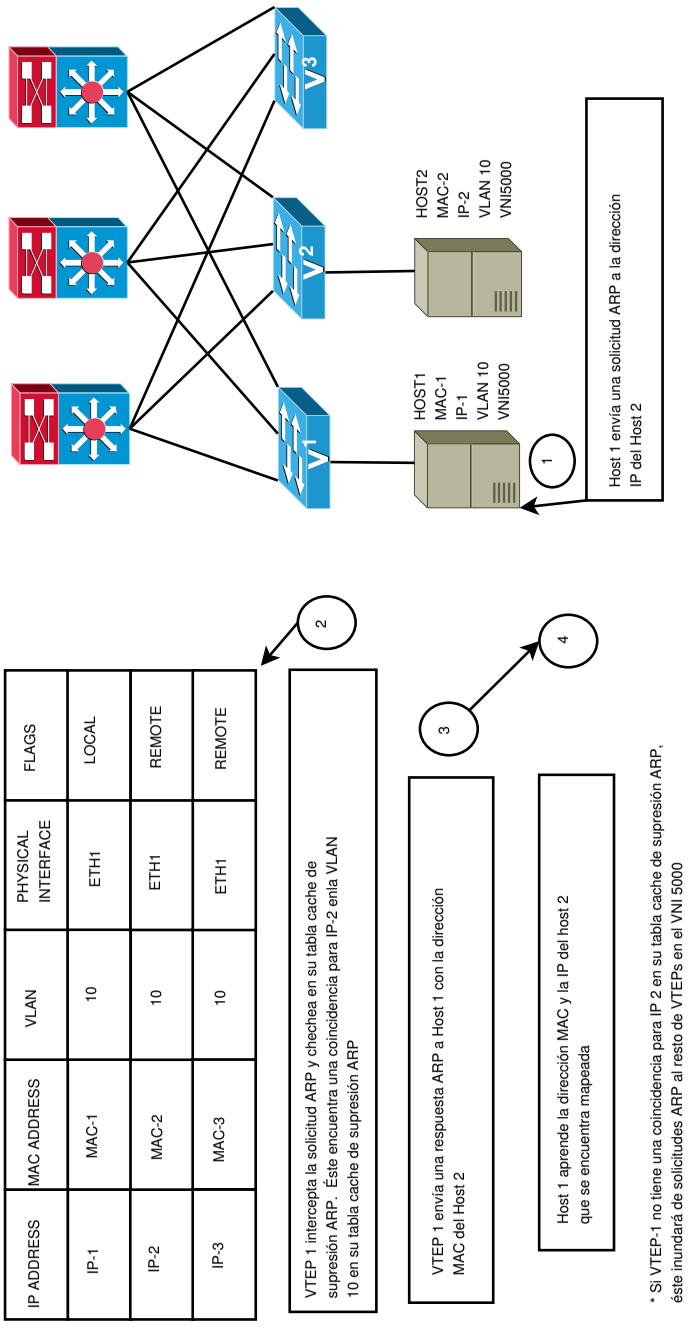
3.2.2. Supresión de ARP

La supresión de ARP es una mejora proporcionada por el plano de control MP-BGP EVPN para reducir la inundación de red causada por el tráfico de broadcast a partir de peticiones ARP.

Cuando la supresión de ARP está habilitada para un VNI específico, sus VTEPs mantienen una tabla cache de supresión de ARP para hosts IP conocidos y direcciones MAC asociadas en el segmento VNI.

Como se ilustra en la figura 28, cuando un host final en un VNI envía una solicitud ARP para otra dirección IP de host final, su VTEP local intercepta la solicitud ARP y comprueba la dirección IP resuelta por ARP en su tabla cache de supresión ARP. Si encuentra una coincidencia, el VTEP local envía una respuesta ARP en nombre del host final remoto. El host local a continuación, aprende la dirección MAC del host remoto en la respuesta ARP.

Si el VTEP local no tiene la dirección IP resuelta por ARP en su tabla de supresión ARP, inunda la solicitud ARP a los otros VTEP en el VNI.



IP ADDRESS	MAC ADDRESS	VLAN	PHYSICAL INTERFACE	FLAGS
IP-1	MAC-1	10	ETH1	LOCAL
IP-2	MAC-2	10	ETH1	REMOTE
IP-3	MAC-3	10	ETH1	REMOTE

Figura 28. Supresión de ARP

CAPÍTULO IV

CONCEPTOS DE FORWARDING Y SIMULACIÓN DE VXLAN EVPN

4.1. Multicast Forwarding

Al configurar VXLAN con multicast forwarding, los VNI definidos en un VTEP deben unirse al mismo grupo multicast. La configuración multicast debe admitir Any-Source Multicast (ASM).

Inicialmente, el VTEP sólo aprenderá las direcciones MAC de dispositivos conectados directamente a ellos.

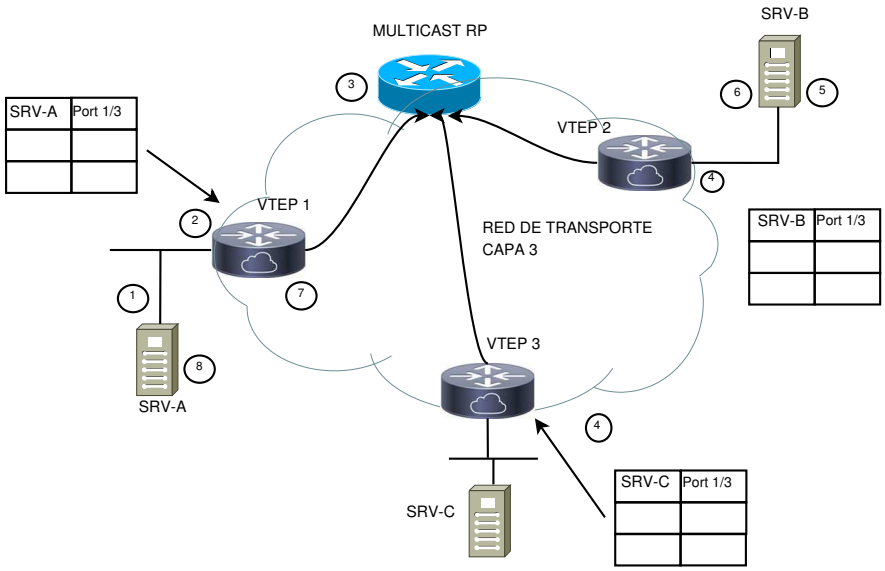


Figura 29. Aprendizaje de direcciones MAC - Equipos directamente conectados

Las direcciones MAC remotas se aprenden mediante una técnica de aprendizaje de direcciones MAC de la siguiente manera:

1. SRV-A quiere comunicarse con SRV-B. SRV-A generará una solicitud ARP tratando de descubrir la dirección MAC de SRV-B.

2. Cuando la solicitud ARP llegue a VTEP1 buscará su tabla local y si no encuentra una entrada, la encapsulará sobre VXLAN y la enviará por el grupo Multicast configurado para el VNI específico.

3. El RP recibe el paquete y enviará una copia a cada VTEP que se ha unido al grupo multicast.

4. Cada VTEP recibirá y desencapsulará el paquete VXLAN y aprenderá la dirección MAC de SRV-A apuntando a la dirección VTEP remota.

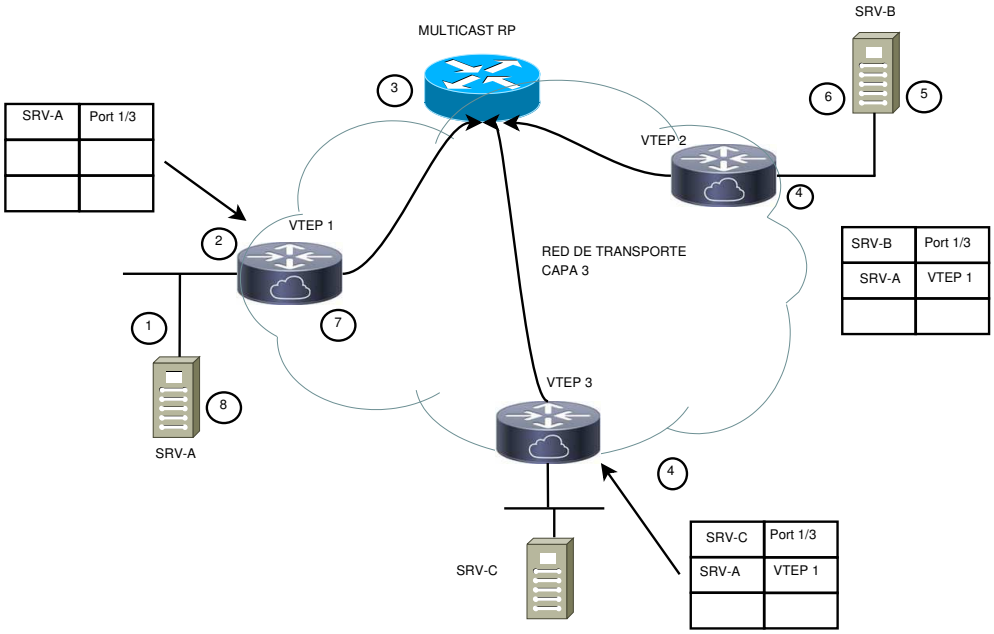


Figura 30. Aprendizaje de la MAC address de SRV-A apuntando al VTEP remoto

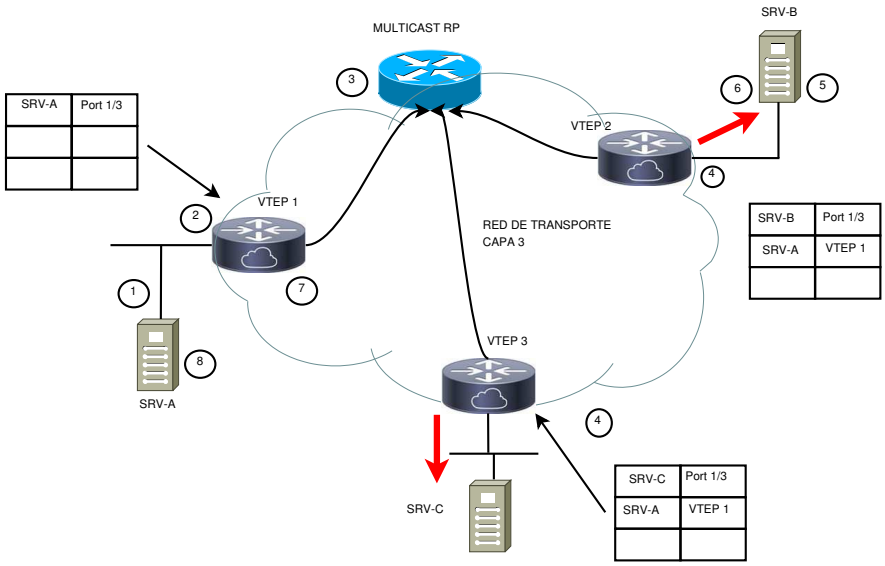


Figura 31. Envío de la solicitud ARP a sus destinos locales

5. Cada VTEP reenvía la solicitud ARP a sus destinos locales.

6. SRV-B genera la respuesta ARP. Cuando VTEP2 lo recibe, buscará en su tabla local y encontrará la información de que el tráfico destinado a SRV-A debe ser enviado a la dirección de VTEP1. VTEP2 encapsula la respuesta ARP con un encabezado VXLAN y lo envía como mensaje unicast a VTEP1.

7. VTEP1 recibirá y desencapsulará el paquete y lo entregará a SRV-A. Ver figura 32

8. Una vez que la información de la dirección MAC se aprende, los paquetes adicionales se enviarán a la dirección del VTEP correspondiente. Ver Figura 33

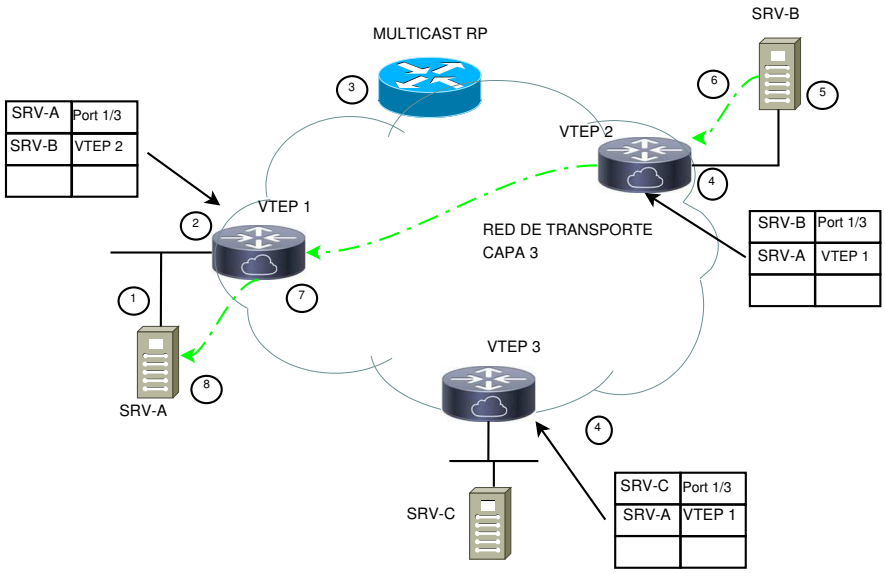


Figura 32. Envío de la respuesta ARP de SRV-B a SRV-A

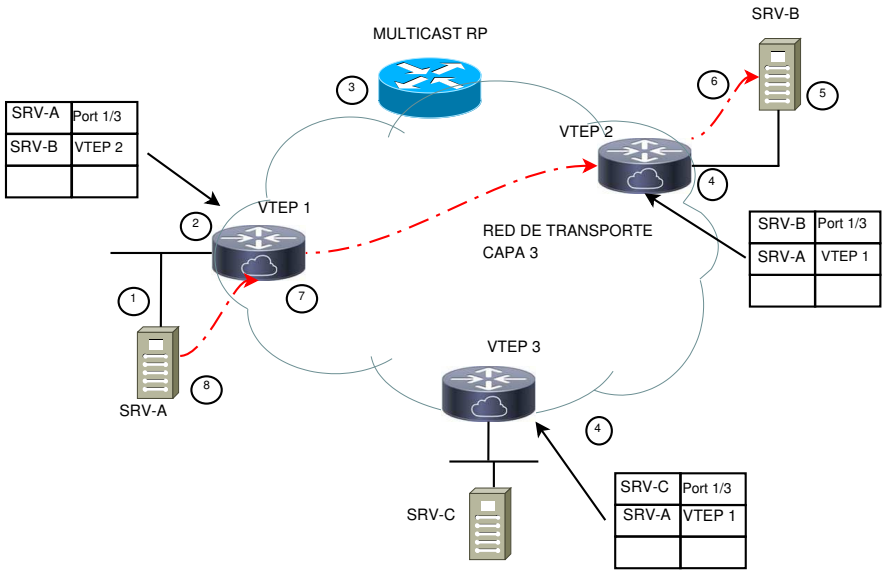


Figura 33. Envío de mensajes unicast de SRV-A a SRV-B

4.2. Unicast Forwarding

Unicast forwarding envía una copia del tráfico multidestino a todos los VTEPs participantes en la interconexión VXLAN.

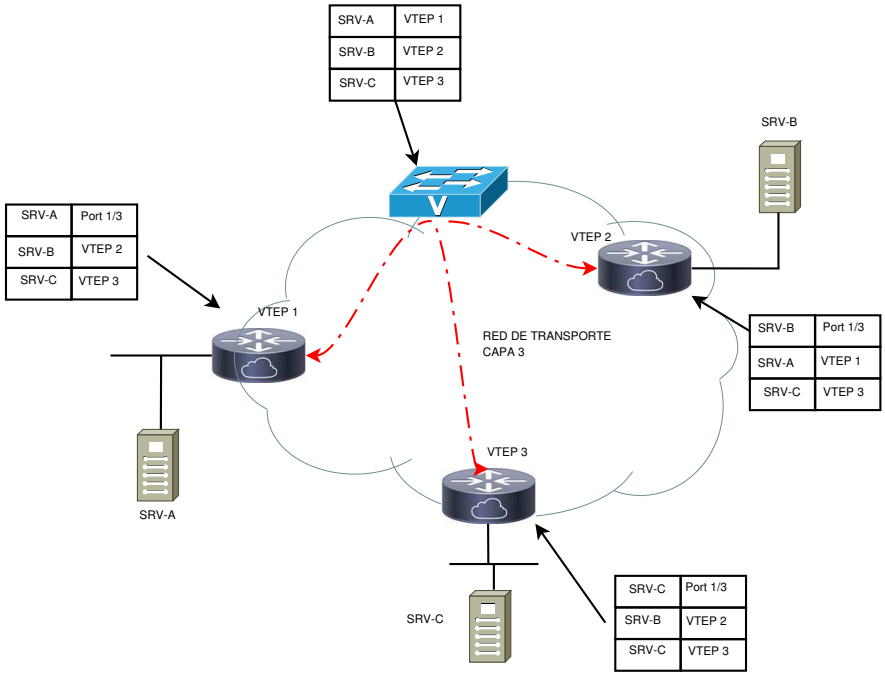


Figura 34. Envío de mensajes unicast

4.3. Simulación de VXLAN con dispositivos Cisco

En esta sección se describirá el proceso empleado para la simulación de VXLAN con dispositivos Cisco utilizando multicast forwarding. Los elementos utilizados fueron los siguientes:

- **Equipo:** Hp Pro Book 440
- **Procesador:** Core I7
- **Memoria RAM:** 16GB
- **Sistema Operativo:** Ubuntu 16.04 64 bits
- **Entorno de vitualización:** GNS3
- **Spines:** Router Cisco 7200

- **Leafs:** Routers CSR1KV
- **Clientes LAN:** VPCS
- **Máquina virtual equipos CSR1KV:** csr1000v-universalk9.03.16.05.S.155-3.S5-ext.ova
- **Imagen routers 7200:** c7200-advipservicesk9-mz.124-4.T1.bin

4.3.1. Topología

En la topología usada, cada LEAF se interconecta con cada SPINE. En los dispositivos LEAF hay una interfaz que se interconecta a los dispositivos de la LAN.

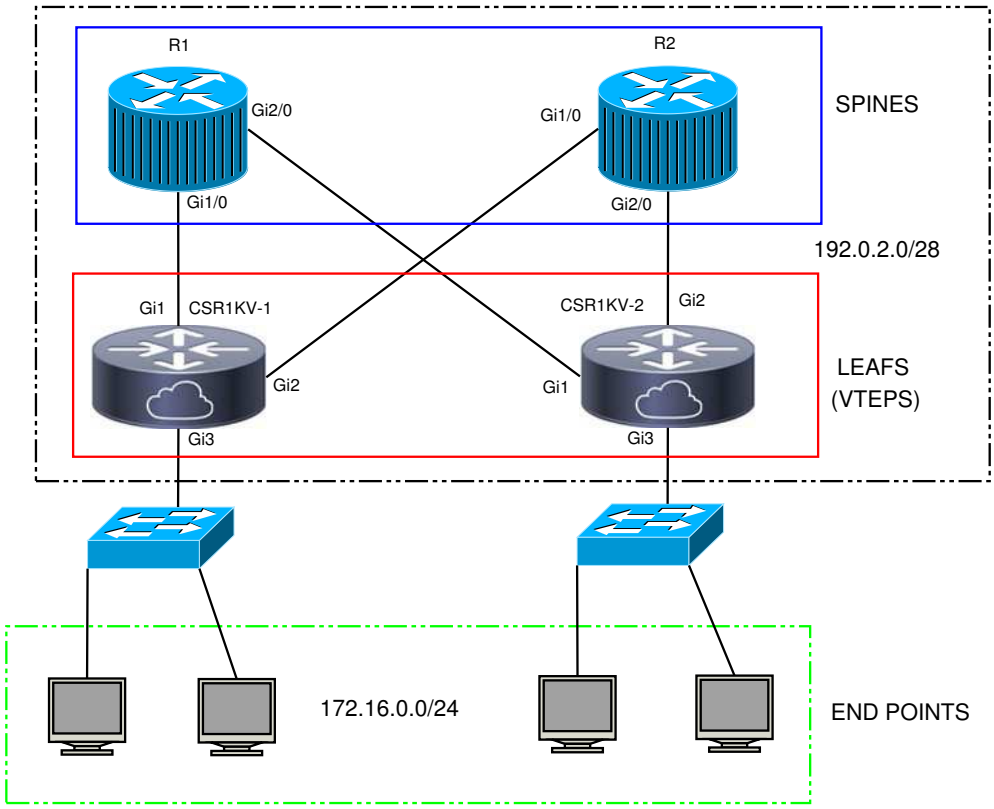


Figura 35. Topología para la simulación de VXLAN con Cisco

4.3.2. Direccionamiento IPV4

Dispositivo	Interfaz	IPV4	Máscara de Subred	Default Gateway
R1	Gi1/0	192.0.2.8	255.255.255.254	N/A
	Gi2/0	192.0.2.10	255.255.255.254	N/A
	loopback0	1.1.1.1	255.255.255.255	N/A
R2	Gi1/0	192.0.2.126	255.255.255.254	N/A
	Gi2/0	192.0.2.14	255.255.255.254	N/A
	loopback0	2.2.2.2	255.255.255.255	N/A

Figura 36. Direccionamiento IPV4 para los dispositivos SPINES

Dispositivo	Interfaz	IPV4	Máscara de Subred	Default Gateway
CSR1KV-1	Gi2	192.0.2.9	255.255.255.254	N/A
	Gi3	192.0.2.13	255.255.255.254	N/A
	loopback0	3.3.3.3	255.255.255.255	N/A
CSR1KV-2	Gi2	192.0.2.11	255.255.255.254	N/A
	Gi3	192.0.2.15	255.255.255.254	N/A
	loopback0	4.4.4.4	255.255.255.255	N/A

Figura 37. Direccionamiento IPV4 para los dispositivos LEAFS

Dispositivo	Interfaz	IPV4	Máscara de Subred	Default Gateway
PC1	eth0	172.16.0.1	255.255.255.0	N/A
PC2	eth0	172.16.0.2	255.255.255.0	N/A
PC3	eth0	172.16.0.3	255.255.255.0	N/A
PC4	eth0	172.16.0.4	255.255.255.0	N/A

Figura 38. Direccionamiento IPV4 para los dispositivos LAN

4.3.3. Configuraciones de enrutamiento

El protocolo de enrutamiento a ser utilizado será OSPF en una sola área y se las redes son punto a punto. Cabe indicar que en los dispositivos LEAFS no se anuncian rutas para el prefijo 172.16.0.0/24. Como ejemplo se indicarán las configuraciones de enrutamiento de los dispositivos R1 y CSR1KV-1

```
R1

--
router ospf 1
 network 1.1.1.1 0.0.0.0 area 0
 network 192.0.2.8 0.0.0.0 area 0
 network 192.0.2.10 0.0.0.0 area 0


CSR1KV-1
```

```
--------
router ospf 1
 network 3.3.3.3 0.0.0.0 area 0
 network 192.0.2.9 0.0.0.0 area 0
 network 192.0.2.13 0.0.0.0 area 0

CSR1KV-1#sh ip route 172.16.0.0
% Network not in table
```

4.3.4. Configuraciones de Multicast

- El enrutamiento multicast está deshabilitado de forma predeterminada en los router Cisco IOS, por lo que tenemos que habilitarlo.

```
R1, R2
(config)#ip multicast-routing
```

- Procedemos a configurar PIM (Protocol Independent Multicast) en las interfaces de R1 y R2

```
R1(config)#interface GigabitEthernet 0/1
R1(config-if)#ip pim sparse-mode
R1(config)#interface GigabitEthernet 0/2
R1(config-if)#ip pim sparse-mode
R1(config)#interface Loopback 0
```

```
R1(config-if)#ip pim sparse-mode
```

- Después se procede a configurar Multicast PIM Bootstrap (BSR). BSR es un protocolo que usamos para encontrar automáticamente el RP (Rendezvous Point) en nuestra red multicast. El grupo multicast está definido en la lista de acceso GROUP1-MCAST y permite el tráfico para el prefijo 239.0.0.0/8.

```
R1,R2
R1#show ip access-list GROUP1-MCAST
Standard IP access list GROUP1-MCAST
10 permit 239.0.0.0, wildcard bits 0.255.255.255
```

```
R1
!En esta configuracion se anuncia a R1 como RP
R1(config)#ip pim bsr-candidate Loopback0 0
```

- Se debe habilitar Multicast Bidirectional PIM. PIM bidireccional ha sido inventado para redes en las que tenemos muchas fuentes y receptores hablando entre sí.

```
R1,R2
```

```
R1(config)#ip pim rp-candidate Loopback0 group-list
           GROUP1-MCAST bidir
R1(config)#ip pim bidir-enable
```

- A continuación se indica la configuración multicast para los dispositivos CSR1KV. La interfaz que se interconecta los leafs con la LAN es la única que no tiene configuración multicast.

```
CSR1KV-1 #show run | i interface|ip pim
interface Loopback0
ip pim sparse-mode
interface GigabitEthernet2
ip pim sparse-mode
interface GigabitEthernet3
ip pim sparse-mode
interface GigabitEthernet4
##it's L2 only, hence no PIM configuration.
ip pim bidir-enable
ip multicast-routing distributed
```

4.3.5. Configuraciones para VXLAN

Para habilitar VXLAN en los dispositivos CSR1KV se deben configurar lo siguiente:

1. Interface NVE (Network Virtualization Endpoint)
2. Service Instance
3. Bridge-Domain

```
int nve 1
source-interface lo0
member vni 6010 mcast-group 239.0.60.10
!
interface GigabitEthernet4
service instance 60 ethernet
encapsulation untagged
exit
!
bridge-domain 1
member vni 6010
member GigabitEthernet4 service-instance 60
```

La vlan 60 (service-instance 60) se encuentra asociada al vni 6010 y el grupo multicast usado es el 239.0.60.10. Afortunadamente, la configuración es tan genérica que podemos copiarla y pegarla a todos los VTEP.

4.3.6. Pruebas de conectividad y análisis a través de CLI

- **Establecimiento de los NVE peers**

Los nve peers constituyen los vecinos que tienen los VTEPS en el dominio VXLAN. El identificativo de cada peer es la dirección loopback0 y el VNI asociado en el 6010.

```
CSR1KV-1#sh nve peers
Interface  Peer-IP          VNI        Peer state
   nve1    4.4.4.4          6010       -

CSR1KV-2#sh nve peers
Interface  Peer-IP          VNI        Peer state
   nve1    3.3.3.3          6010       -
```

Figura 39. NVE peers

- **Pruebas de conectividad entre los dispositivos de la LAN**

```
PC1> ping 172.16.0.4

84 bytes from 172.16.0.4 icmp_seq=1 ttl=64 time=19.795 ms
84 bytes from 172.16.0.4 icmp_seq=2 ttl=64 time=15.467 ms
84 bytes from 172.16.0.4 icmp_seq=3 ttl=64 time=16.132 ms
84 bytes from 172.16.0.4 icmp_seq=4 ttl=64 time=15.944 ms
84 bytes from 172.16.0.4 icmp_seq=5 ttl=64 time=16.009 ms

PC4> ping 172.16.0.1

84 bytes from 172.16.0.1 icmp_seq=1 ttl=64 time=10.054 ms
84 bytes from 172.16.0.1 icmp_seq=2 ttl=64 time=17.272 ms
84 bytes from 172.16.0.1 icmp_seq=3 ttl=64 time=16.348 ms
84 bytes from 172.16.0.1 icmp_seq=4 ttl=64 time=16.246 ms
84 bytes from 172.16.0.1 icmp_seq=5 ttl=64 time=16.108 ms
```

Figura 40. Conectividad entre los dispositivos de LAN

Como se ve en la figura 40, existe conectividad a nivel de capa 3 entre los dispositivos de la LAN. Con esta prueba se demuestra la hipótesis planteda en el capítulo 1 ya que según el diagrama de la simulación PC1

y PC4 se encuentran en localizaciones remotas y su direccionamiento IPV4 pertenece al mismo segmento de red, en este caso 172.16.0.0/24.

- **Aprendizaje de direcciones MAC en el bridge domain**

```
CSR1KV-1#show bridge-domain 1
Bridge-domain 1 (2 ports in all)
State: UP                    Mac learning: Enabled
Aging-Timer: 300 second(s)
    GigabitEthernet4 service instance 60
    vni 6010
   AED MAC address    Policy  Tag       Age  Pseudoport
   0   0050.7966.6801 forward dynamic   283  nve1.VNI6010, VxLAN
                                           src: 3.3.3.3 dst: 4.4.4.4
   0   0050.7966.6800 forward dynamic   279  nve1.VNI6010, VxLAN
                                           src: 3.3.3.3 dst: 4.4.4.4
   0   0050.7966.6803 forward dynamic   272  GigabitEthernet4.EFP60
   0   0050.7966.6802 forward dynamic   283  GigabitEthernet4.EFP60


CSR1KV-2#show bridge-domain 1
Bridge-domain 1 (2 ports in all)
State: UP                    Mac learning: Enabled
Aging-Timer: 300 second(s)
    GigabitEthernet4 service instance 60
    vni 6010
   AED MAC address    Policy  Tag       Age  Pseudoport
   0   0050.7966.6801 forward dynamic   247  GigabitEthernet4.EFP60
   0   0050.7966.6800 forward dynamic   243  GigabitEthernet4.EFP60
   0   0050.7966.6803 forward dynamic   237  nve1.VNI6010, VxLAN
                                           src: 4.4.4.4 dst: 3.3.3.3
   0   0050.7966.6802 forward dynamic   247  nve1.VNI6010, VxLAN
                                           src: 4.4.4.4 dst: 3.3.3.3
```

Figura 41. Aprendizaje de direcciones MAC en el bridge-domain 1

- **Tráfico a través de la interfaz NVE**

Con el comando *show nve interfave nve 1 detail* se pueve verificar el tráfico que atraviesa por la interfaz nve. Este tráfico constituye los paquetes encapsulados mediante VXLAN que tienen por origen o destino los VTEPs.

```
CSR1KV-1#sh nve interface nve1 detail
Interface: nve1, State: Admin Up, Oper Up Encapsulation: Vxlan
source-interface: Loopback0 (primary:3.3.3.3 vrf:0)
   Pkts In   Bytes In   Pkts Out  Bytes Out
       117      11312        120      11530


CSR1KV-2#sh nve int nve1 detail
Interface: nve1, State: Admin Up, Oper Up Encapsulation: Vxlan
source-interface: Loopback0 (primary:4.4.4.4 vrf:0)
   Pkts In   Bytes In   Pkts Out  Bytes Out
       117      11314        117      11312
```

Figura 42. Tráfico a través de la interfaz NVE

4.3.7. Análisis de VXLAN Wireshark

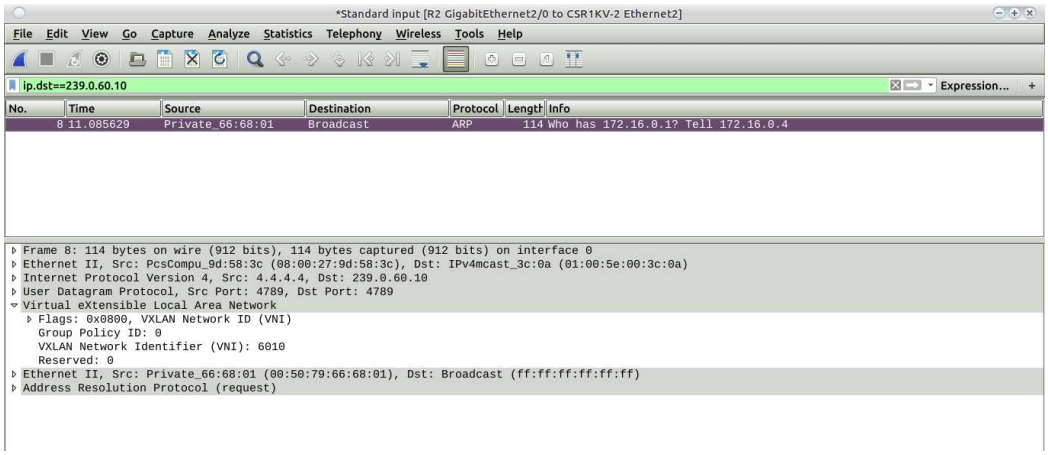


Figura 43. Análisis de VXLAN con Wireshark

La figura 43 representa la captura de un paquete con encapsulación VXLAN que sale desde el dispositivo CSR1KV-2 hacia R2. Como se trata de un reenvío de datos a través de multicast, la dirección IPV4 de destino corresponde a la dirección IPV4 del grupo multicast, en este caso 239.0.60.10.

4.4. Simulación de VXLAN EVPN con Cumulus Linux

En esta sección se describirá el proceso empleado para la simulación de VXLAN utilizando unicast forwarding. Los elemento utilizados fueron los siguientes:

- **Equipo:** Hp Pro Book 440
- **Procesador:** Core I7
- **Memoria RAM:** 16GB
- **Sistema Operativo:** Ubuntu 16.04 64 bits
- **Entorno de vitualización:** GNS3
- **Spines, Leafs y Clientes LAN:** Cumulus Linux
- **Máquina virtual equipos CSR1KV:** csr1000v-universalk9.03.16.05.S.155-3.S5-ext.ova

4.4.1. Topología

En la topología empleada (ver figura 44), cada LEAF se interconecta con cada SPINE. Los dispositivos SPINE perteneces al AS 65000, Los dispositivos LEAF pertenecen a los AS 65003 y 65004 y se interconectan entre sí mediante dos enlaces, y cada servidor se interconecta con los dispositivos LEAF a través de un enlace. Dichas conexiones permitirán generar la configuración Multi-chassis link aggregation (MCLAG).

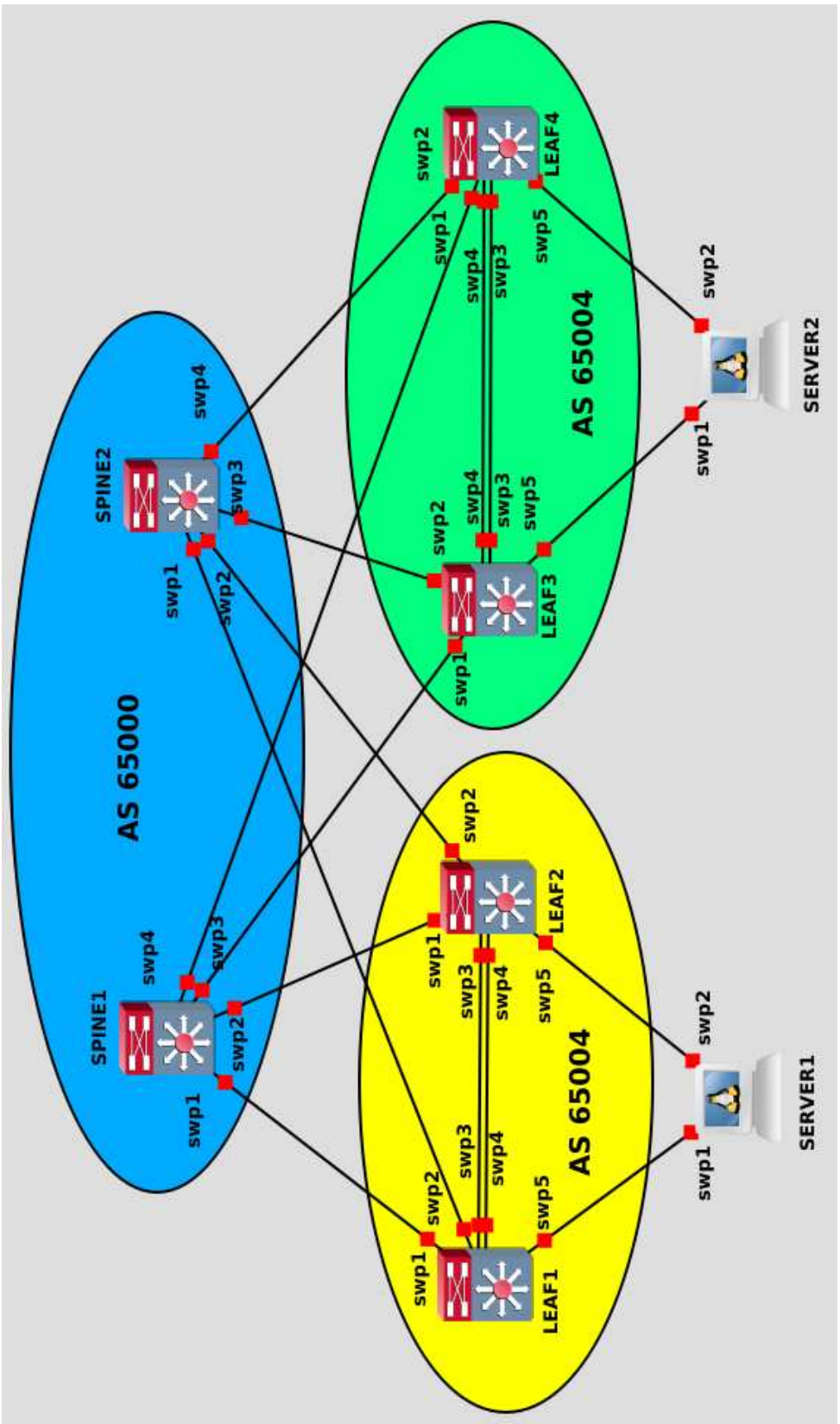


Figura 44. Topología para la simulación de VXLAN EVPN con Cumulus Linux

4.4.2. Direccionamiento

AS	Dispositivo
65000	SPINE1
	SPINE2
65003	LEAF1
	LEAF2
65004	LEAF3
	LEAF4

Figura 45. Sistemas Autónomos para la topología de simulación con Cumulus Linux

Dispositivo	Interfaz	IPV4	Máscara de Subred	Default Gateway
SPINE1	swp1	192.0.2.8	255.255.255.254	N/A
	swp2	192.0.2.10	255.255.255.254	N/A
	swp3	192.0.2.12	255.255.255.254	N/A
	swp4	192.0.2.14	255.255.255.254	N/A
	loopback	172.31.0.3	255.255.255.255	N/A
SPINE2	swp1	192.0.2.136	255.255.255.254	N/A
	swp2	192.0.2.138	255.255.255.254	N/A
	swp3	192.0.2.140	255.255.255.254	N/A
	swp4	192.0.2.142	255.255.255.254	N/A
	loopback	172.31.0.4	255.255.255.255	N/A

Figura 46. Direccionamiento IPV4 para los dispositivos SPINES

Dispositivo	Interfaz	IPV4	Máscara de Subred	Default Gateway
LEAF1	swp1	192.0.9.8	255.255.255.254	N/A
	swp2	192.0.2.137	255.255.255.254	N/A
	loopback	172.16.3.1	255.255.255.255	N/A
LEAF2	swp1	192.0.2.11	255.255.255.254	N/A
	swp2	192.0.2.139	255.255.255.254	N/A
	loopback	172.16.3.2	255.255.255.255	N/A
LEAF3	swp1	192.0.2.13	255.255.255.254	N/A
	swp2	192.0.2.141	255.255.255.254	N/A
	loopback	172.16.4.1	255.255.255.255	N/A
LEAF4	swp1	192.0.2.15	255.255.255.254	N/A
	swp2	192.0.2.143	255.255.255.254	N/A
	loopback	172.16.4.2	255.255.255.255	N/A

Figura 47. Direccionamiento IPV4 para los dispositivos LEAFS

4.4.3. Configuración de Quagga

En el archivo /etc/quagga/daemons, las opciones de zebra y bgpd deben estar activadas

Después, se debe habilitar el servicio quagga en todos los dispositivos, mediante los siguientes comandos:

```
cumulus@leaf1:~$ sudo systemctl enable quagga.service
cumulus@leaf1:~$ sudo systemctl start quagga.service
```

```
# The watchquagga daemon is always started. Per default in monitoring-only but
# that can be changed via /etc/quagga/debian.conf.
#
zebra=yes
bgpd=yes
ospfd=no
ospf6d=no
ripd=no
ripngd=no
isisd=no
pimd=no
ldpd=no
```

Figura 48. Configuración del archivo /etc/quagga/daemons

4.4.4. Configuración eBGP

Citaremos como ejemplo la configuración eBGP del dispositivo SPINE1, misma que se detalla en la figura 49. Se creó una prefix-list denominada PL_LO_CLOS con el fin de sólo permitir el tráfico de las redes 172.16.0.0/12 y 192.0.2.0/24. Para la red 192.0.2.0/24 se permitirán prefijos mayores o iguales a /31.

La configuración eBGP para los dispositivos LEAF es muy similar a la de los dispositivos SPINE. La diferencia radica en el peer group, el cual se denomina PEER_SPINE. La configuración eBGP de los dispostivos LEAF1 Y LEAF2 se detalla en la figura 50

4.4.5. Configuración MCLAG

Multi-Chassis Link Aggregation de enlaces, o MCLAG, permite generar una agregación de enlaces conectando dos puertos entre dos switches diferentes,

```
net add loopback lo ip address 172.31.0.3/32

net add bgp autonomous-system 65000
net add bgp router-id 172.31.0.3

net add routing prefix-list ipv4 PL_LO_CLOS seq 10 permit 172.16.0.0/12 ge 32 le 32
net add routing prefix-list ipv4 PL_LO_CLOS seq 20 permit 192.0.2.0/24 ge 31 le 31
net add bgp redistribute connected

net add bgp neighbor PEER_LEAF peer-group
net add bgp neighbor PEER_LEAF prefix-list PL_LO_CLOS out
net add bgp neighbor PEER_LEAF next-hop-self

net add bgp neighbor 192.0.2.9 remote-as 65003
net add bgp neighbor 192.0.2.9 description leaf1
net add bgp neighbor 192.0.2.9 peer-group PEER_LEAF

net add bgp neighbor 192.0.2.11 remote-as 65003
net add bgp neighbor 192.0.2.11 description leaf2
net add bgp neighbor 192.0.2.11 peer-group PEER_LEAF

net add bgp neighbor 192.0.2.13 remote-as 65004
net add bgp neighbor 192.0.2.13 description leaf3
net add bgp neighbor 192.0.2.13 peer-group PEER_LEAF

net add bgp neighbor 192.0.2.15 remote-as 65004
net add bgp neighbor 192.0.2.15 description leaf4
net add bgp neighbor 192.0.2.15 peer-group PEER_LEAF
```

Figura 49. Configuración eBGP para SPINE1

```
net add loopback lo ip address 172.16.3.1/32

net add bgp autonomous-system 65003
net add bgp router-id 172.16.3.1

net add routing prefix-list ipv4 PL_LO_CLOS seq 10 permit 172.16.0.0/12 ge 32 le 32
net add routing prefix-list ipv4 PL_LO_CLOS seq 20 permit 192.0.2.0/24 ge 31 le 31
net add bgp redistribute connected

net add bgp neighbor PEER_SPINE peer-group
net add bgp neighbor PEER_SPINE prefix-list PL_LO_CLOS out

net add bgp neighbor 192.0.2.8 remote-as 65000
net add bgp neighbor 192.0.2.8 description spine1
net add bgp neighbor 192.0.2.8 peer-group PEER_SPINE

net add bgp neighbor 192.0.2.136 remote-as 65000
net add bgp neighbor 192.0.2.136 description spine2
net add bgp neighbor 192.0.2.136 peer-group PEER_SPINE
```

Figura 50. Configuración eBGP para LEAF1

de tal manera que lógicamente funcionen como uno solo. Esto proporciona una mayor redundancia y mayor el rendimiento del sistema.

```
net add loopback lo ip address 172.16.3.2/32

net add bgp autonomous-system 65003
net add bgp router-id 172.16.3.2

net add routing prefix-list ipv4 PL_LO_CLOS seq 10 permit 172.16.0.0/12 ge 32 le 32
net add routing prefix-list ipv4 PL_LO_CLOS seq 20 permit 192.0.2.0/24 ge 31 le 31
net add bgp redistribute connected

net add bgp neighbor PEER_SPINE peer-group
net add bgp neighbor PEER_SPINE prefix-list PL_LO_CLOS out

net add bgp neighbor 192.0.2.10 remote-as 65000
net add bgp neighbor 192.0.2.10 description spine1
net add bgp neighbor 192.0.2.10 peer-group PEER_SPINE

net add bgp neighbor 192.0.2.138 remote-as 65000
net add bgp neighbor 192.0.2.138 description spine2
net add bgp neighbor 192.0.2.138 peer-group PEER_SPINE
```

Figura 51. Configuración eBGP para LEAF2

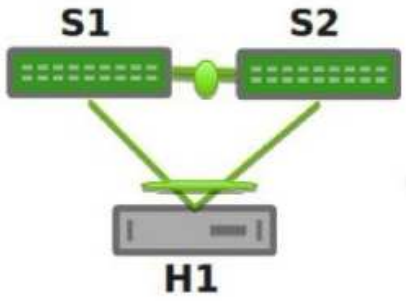


Figura 52. Esquema MCLAG

La configuración MCLAG se la realizó con los dispositivos LEAF. En la figura 53. Se indica las interfaces físicas y lógicas que se definieron para dicha configuración

La figura 54. muestra el direccionamiento de capa 2 y capa 3 utilizado en la configuración MCLAG

A continuación se muestra la configuración MCLAG para los dispositivos LEAF1 y LEAF2. Dicha configuración es igual en los dispositivos LEAF3 y LEAF4.

Dispositivo	Interfaz lógica	Interfaces físicas
LEAF1	bond0	swp3, swp4
LEAF4		
LEAF3		
LEAF4		

Figura 53. Interfaces de los equipos LEAF definidas para MCLAG

Dispositivo	Interfaz Lógica	IPV4	Dirección MAC MCLAG
LEAF1, LEAF3	bond0.4094	198.51.100.1/30	44:38:39:FF:40:94
LEAF2, LEAF4	bond0.4094	192.51.100.2/30	44:38:39:FF:40:94

Figura 54. Interfaces de los equipos LEAF definidas para MCLAG

```
net add bond bond0 bond slaves swp3-4
net add bond bond0 alias DEV=leaf2 IF=bond0
net add interface bond0.4094 alias MLAG DEDICATED
net add interface bond0.4094 ip address 198.51.100.1/30
net add interface bond0.4094 clag peer-ip 198.51.100.2
net add interface bond0.4094 clag sys-mac 44:38:39:FF:40:94
```

Figura 55. Configuración MCLAG para LEAF1

```
net add bond bond0 bond slaves swp3-4
net add bond bond0 alias DEV=leaf1 IF=bond0
net add interface bond0.4094 alias MLAG DEDICATED
net add interface bond0.4094 ip address 198.51.100.2/30
net add interface bond0.4094 clag peer-ip 198.51.100.1
net add interface bond0.4094 clag sys-mac 44:38:39:FF:40:94
```

Figura 56. Configuración MCLAG para LEAF2

4.4.6. Configuración MLAG Downlink

En esta configuración, con las interfaces que van hacia los servidores se forma una agregación de enlaces. A ésta la conoceremos como bond1.

```
net add bond bond1 bond slaves swp5
net add bond bond1 alias DEV=server1 IF=bond0
net add bond bond1 mtu 9000
net add bond bond1 clag id 1
```

Figura 57. Configuración MCLAG DOWNLINK para LEAF1 y LEAF2

```
net add bond bond1 bond slaves swp5
net add bond bond1 alias DEV=server2 IF=bond0
net add bond bond1 mtu 9000
net add bond bond1 clag id 1
```

Figura 58. Configuración MCLAG DOWNLINK para LEAF3 y LEAF4

Después se crea una interfaz tipo puente en cada uno de los dispositivos LEAF, cuyos miembros son las interfaces bon0 y bond1, definidas en la configuración MCLAG y MCLAG DOWNLINK, en las cuales se permitirá el paso de vlans.

```
net add bridge bridge ports bond0
net add bridge bridge ports bond1
net add bridge bridge vids 2-4093
```

Figura 59. Configuración de la interfaz bridge en los dispositivos LEAF

En los servidores se crea subinterfaces, las cuales representan las vlans a comunicar mediante el entunelamiento vxlan. En la figura 60, se indica el direccionamiento para las subinterfaces de las vlan 1, 100 y 200.

Dispositivo	Interfaz	IPV4	Máscara de Subred	Default Gateway
SERVER1	bond0	192.168.0.1	255.255.255.0	N/A
	bond0.100	192.168.100.1	255.255.255.0	N/A
	bond0.200	192.168.200.1	255.255.255.0	N/A
SERVER2	bond0	192.168.0.2	255.255.255.0	N/A
	bond0.100	192.168.100.2	255.255.255.0	N/A
	bond0.200	192.168.200.2	255.255.255.0	N/A

Figura 60. Direccionamiento IPV4 para los Servidores

```
-------------------------------------------------------------
|   SERVER1                       |     SERVER2               |
|   =======                       |     =======               |
|                                 |                           |
|   auto eth0                     |     auto eth0             |
|   iface eth0 inet dhcp          |     iface eth0 inet dhcp  |
|                                 |                           |
|   auto bond0                    |     auto bond0            |
|   iface bond0                   |     iface bond0           |
|   bond-slaves swp1 swp2         |     bond-slaves swp1 swp2 |
|   address 192.168.0.1/24        |     address 192.168.0.2/24|
|                                 |                           |
|   auto bond0.100                |     auto bond0.100        |
|   iface bond0.100               |     iface bond0.100       |
|   address 192.168.100.1/24      |     address 192.168.100.2/24 |
|                                 |                           |
|                                 |                           |
|   auto bond0.200                |     auto bond0.200        |
|   iface bond0.200               |     iface bond0.200       |
|   address 192.168.200.1/24      |     address 192.168.200.2/24 |
-------------------------------------------------------------
```

Figura 61. Configuración de las subinterfaces en los SERVERS

4.4.7. Configuración VXLAN

- Se debe configurar la ip virtual en cada dispositivo VTEP, correspondiente al identificativo MCLAG. Esta dirección ip virtual pertenece al rango de prefijos permitidos en la prefix list PL_LO_CLOS, por tanto va a ser anunciada en la tabla de enrutamiento.

```
LEAF1,LEAF2
===========

net add loopback lo clag vxlan-anycast-ip 172.16.3.100
net commit

LEAF3, LEAF4
============

net add loopback lo clag vxlan-anycast-ip 172.16.4.100
net commit
```

Figura 62. Configuración de la ip virtual en los dispositivos LEAF

- En cada dispositivo LEAF se debe configurar la interfaz VNI, misma que representa la instancia de red lógica y define un dominio de broadcast de capa 2. A cada VNI se de debe asignar una identificación (VNID) y se debe permitir su paso a través de la interfaz bridge, los VNID son para las vlan 1,100 y 200.

```
net add vxlan vxlan10001 vxlan id 10001
net add vxlan vxlan10001 bridge access 1

net add vxlan vxlan10100 vxlan id 10100
net add vxlan vxlan10100 bridge access 100

net add vxlan vxlan10200 vxlan id 10200
net add vxlan vxlan10200 bridge access 200

net commit
```

Figura 63. Configuración de las interfaces VNI en los dispositivos LEAF

- En los dispositivos LEAF se debe configurar el tunel VXLAN. En este caso será entre LEAF1 - LEAF2 y LEAF3 - LEAF4 donde LEAF1 y LEAF3 constituyen el origen del tunel VXLAN y LEAF2 y LEAF4 el destino del tunel VXLAN. En la figura 64 se muestra la configuración del tunel VXLAN en todos los dispositivos LEAF.

```
LEAF1
=====

net add vxlan vxlan10100 vxlan local-tunnelip 172.16.3.1
net add vxlan vxlan10200 vxlan local-tunnelip 172.16.3.1
net commit

LEAF2
=====
net add vxlan vxlan10100 vxlan local-tunnelip 172.16.3.2
net add vxlan vxlan10200 vxlan local-tunnelip 172.16.3.2
net commit

LEAF3
=====

net add vxlan vxlan10100 vxlan local-tunnelip 172.16.4.1
net add vxlan vxlan10200 vxlan local-tunnelip 172.16.4.1
net commit

LEAF4
=====

net add vxlan vxlan10100 vxlan local-tunnelip 172.16.4.2
net add vxlan vxlan10200 vxlan local-tunnelip 172.16.4.2
net commit
```

Figura 64. Configuración del túnel VXLAN en los dispositivos LEAF

4.4.8. Configuración EVPN Unicast

En los dispositivos tanto SPINE con LEAF en la address-family evpn dentro de cada proceso BGP se debe activar cada peer group correspondiente. En los dispositivos LEAF, adicional a lo anterior se debe anunciar todos los VNI,ésto permitirá conocer la información de los VNI en todo el dominio que utiliza VXLAN. En la figura 65 se muestra un ejemplo de ésta configuración en el dispositivo LEAF1 a través de CLI mediante el uso dela aplicación vtysh.

```
cumulus@leaf1:~$ sudo vtysh
Hello, this is Quagga (version 1.0.0+cl3eau13).
Copyright 1996-2005 Kunihiro Ishiguro, et al.
leaf3#
leaf3# configure terminal
leaf3(config)# router bgp 65004
leaf3(config-router)# address-family evpn
leaf3(config-router-af)# neighbor PEER_SPINE activate
leaf3(config-router-af)# advertise-all-vni
leaf3(config-router-af)# end
leaf3# wr
Note: this version of vtysh never writes vtysh.conf
Building Configuration...
Integrated configuration saved to /etc/quagga/Quagga.conf
[OK]
leaf3#
leaf3# exit
cumulus@leaf1:~$
```

Figura 65. Ejemplo de configuración de la address-family evpn mediante vtysh

La configuración para los dispositivos SPINE y LEAF se indica en la figura 66.

```
SPINE1, SPINE2
==============

router bgp 65000
 address-family evpn
  neighbor PEER_SPINE activate

LEAF1, LEAF2
============

router bgp 65003
 address-family evpn
  neighbor PEER_BB activate
  advertise-all-vni

LEAF3, LEAF4
============

router bgp 65004
 address-family evpn
  neighbor PEER_BB activate
  advertise-all-vni
```

Figura 66. Configuración de la address-family evpn

4.4.9. Deshabilitación del aprendizaje de direcciones MAC en los túneles VXLAN

Como último item de configuración, se debe deshabilitar el aprendizaje de direcciones MAC en las interfaces VNI. Para ésto, se debe agregar la opcion bridge-learning off. En la figura 67 se muestra un ejemplo en el dispositivo LEAF1

```
cumulus@leaf1:~$ diff -u /var/tmp/etc_network_interfaces /etc/network/interfaces
--- /var/tmp/etc_network_interfaces
+++ /etc/network/interfaces

auto vxlan10100
iface vxlan010100
bridge-access 100
+ bridge-learning off
mstpctl-bpduguard yes
mstpctl-portbpdufilter yes
vxlan-id 10100

auto vxlan010200
iface vxlan010200
bridge-access 200
+ bridge-learning off
mstpctl-bpduguard yes
mstpctl-portbpdufilter yes
vxlan-id 10200
```

Figura 67. Deshabilitación del aprendizaje de direcciones MAC

4.4.10. Resultados y pruebas de conectividad

- **Conectividad entre los servidores**

  Como se ve en la figura 68, existe conectividad a nivel de capa 3 entre los servidores. Con esta prueba nuevamente se demuestra la hipótesis planteada en el capítulo 1 debido a que éstos según la topología se encuentran en localizaciones remotas y el direccionamiento IPV4 para cada una de las vlans pertenece al mismo segmento de red.

```
cumulus@server1:~$ ping 192.168.0.2
PING 192.168.0.2 (192.168.0.2) 56(84) bytes of data.
64 bytes from 192.168.0.2: icmp_seq=1 ttl=64 time=7.44 ms
64 bytes from 192.168.0.2: icmp_seq=2 ttl=64 time=2.99 ms
64 bytes from 192.168.0.2: icmp_seq=3 ttl=64 time=2.92 ms
64 bytes from 192.168.0.2: icmp_seq=4 ttl=64 time=7.88 ms
64 bytes from 192.168.0.2: icmp_seq=5 ttl=64 time=3.06 ms
^C
--- 192.168.0.2 ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4006ms
rtt min/avg/max/mdev = 2.928/4.862/7.880/2.291 ms


cumulus@server2:~$ ping 192.168.0.1
PING 192.168.0.1 (192.168.0.1) 56(84) bytes of data.
64 bytes from 192.168.0.1: icmp_seq=1 ttl=64 time=3.95 ms
64 bytes from 192.168.0.1: icmp_seq=2 ttl=64 time=3.21 ms
64 bytes from 192.168.0.1: icmp_seq=3 ttl=64 time=2.94 ms
64 bytes from 192.168.0.1: icmp_seq=4 ttl=64 time=3.62 ms
64 bytes from 192.168.0.1: icmp_seq=5 ttl=64 time=3.09 ms
^C
--- 192.168.0.1 ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4006ms
rtt min/avg/max/mdev = 2.948/3.365/3.951/0.372 ms
```

Figura 68. Pruebas de conectividad entre server1 y server2

- **Tabla de direcciones MAC en los dispositivos LEAF**

  En la figura 69 se muestra el aprendizaje de direcciones MAC. Puede verse que la columna TunnelDest utiliza la dirección IP de loopback compartida (clag-id) del VTEP opuesto. El valor 00: 00: 00: 00: 00: 00 en la columna de direcciones MAC indica la replicación de tráfico BUM.

```
cumulus@leaf1:~$ net show bridge macs
VLAN     Master   Interface   MAC                TunnelDest    State      Flags          LastSeen
-------  -------  ----------  -----------------  ------------  ---------  -------------  --------
1        bridge   bond1       0a:00:27:3f:94:4c                                          00:00:13
1        bridge   bond1       0a:00:27:60:4f:90                                          00:00:20
1        bridge   bond1       08:00:27:3f:94:4c                                          00:00:50
1        bridge   vxlan10001  0a:00:27:76:45:44                offload                   00:06:06
1        bridge   vxlan10001  0a:00:27:82:6c:93                offload                   00:06:06
1        bridge   vxlan10001  08:00:27:76:45:44                offload                   00:01:34
untagged         vxlan10001  00:00:00:00:00:00  172.16.4.100  permanent  self           00:06:43
untagged         vxlan10001  0a:00:27:76:45:44  172.16.4.100               self, offload  00:06:06
untagged         vxlan10001  0a:00:27:82:6c:93  172.16.4.100               self, offload  00:06:06
untagged         vxlan10001  08:00:27:76:45:44  172.16.4.100               self, offload  00:01:34
untagged         vxlan10100  00:00:00:00:00:00  172.16.4.100  permanent  self           00:06:43
untagged         vxlan10200  00:00:00:00:00:00  172.16.4.100  permanent  self           00:06:43
untagged bridge  bond0       08:00:27:29:96:73                permanent                 00:07:05
untagged bridge  bond1       08:00:27:af:24:22                permanent                 00:07:05
untagged bridge  vxlan10001  8e:40:7e:c1:fd:60                permanent                 00:07:05
untagged bridge  vxlan10100  d6:3a:41:a2:3b:1d                permanent                 00:07:05
untagged bridge  vxlan10200  1e:87:3b:64:49:e3                permanent                 00:07:05
```

Figura 69. Aprendizaje de direcciones MAC en el dispositivo LEAF1

- **Tabla de enrutamiento**

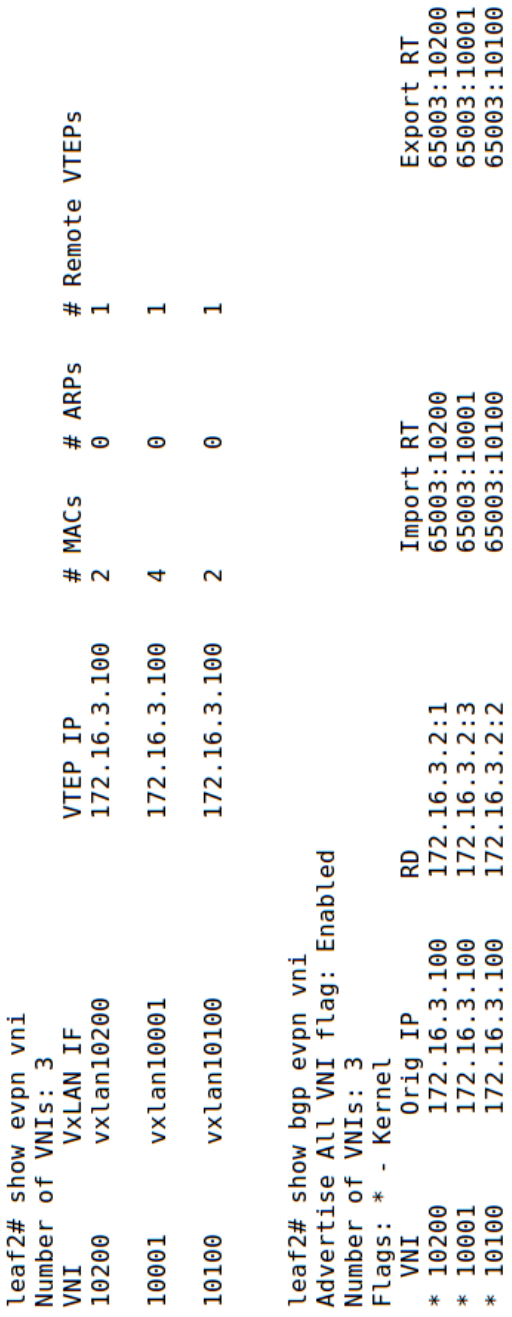
```
cumulus@spine1:~$ net show route

show ip route
=============
Codes: K - kernel route, C - connected, S - static, R - RIP,
       O - OSPF, I - IS-IS, B - BGP, P - PIM, T - Table, v - VNC,
       V - VPN,
       > - selected route, * - FIB route

B>* 172.16.3.1/32 [20/0] via 192.0.2.9, swp1, 00:08:41
B>* 172.16.3.2/32 [20/0] via 192.0.2.11, swp2, 00:08:35
B>* 172.16.3.100/32 [20/0] via 192.0.2.11, swp2, 00:08:31
  *                        via 192.0.2.9, swp1, 00:08:31
B>* 172.16.4.1/32 [20/0] via 192.0.2.13, swp3, 00:08:24
B>* 172.16.4.2/32 [20/0] via 192.0.2.15, swp4, 00:09:27
B>* 172.16.4.100/32 [20/0] via 192.0.2.13, swp3, 00:08:20
  *                        via 192.0.2.15, swp4, 00:08:20
C>* 172.31.0.3/32 is directly connected, lo
C>* 192.0.2.8/31 is directly connected, swp1
C>* 192.0.2.10/31 is directly connected, swp2
C>* 192.0.2.12/31 is directly connected, swp3
C>* 192.0.2.14/31 is directly connected, swp4
B>* 192.0.2.136/31 [20/0] via 192.0.2.9, swp1, 00:08:41
B>* 192.0.2.138/31 [20/0] via 192.0.2.11, swp2, 00:08:35
B>* 192.0.2.140/31 [20/0] via 192.0.2.13, swp3, 00:08:24
B>* 192.0.2.142/31 [20/0] via 192.0.2.15, swp4, 00:09:27
```

Figura 70. Tabla de enrutamiento del dispositivo SPINE1

La tabla de enrutamiento muestra las redes directamente conectadas denotadas por la letra C y las rutas aprendidas por BGP denotadas por la letra B. Para los prefijos 172.16.3.100 y 172.16.4.100 existen dos rutas, esto se debe a la redundancia generada mediante la configuración MCLAG entre los dispositivos LEAF.

- **Información de VNI y VTEPs**

En la figura 71 se puede visualizar la información de los VNI configurados para el entunelamiento VXLAN. Los route distinguisher y los route targe tanto import como export fueron generados automáticamente con base en lo indicado en el capítulo 3.



```
leaf2# show evpn vni
Number of VNIs: 3
VNI        VxLAN IF              VTEP IP         # MACs   # ARPs   # Remote VTEPs
10200      vxlan10200            172.16.3.100    2        0        1

10001      vxlan10001            172.16.3.100    4        0        1

10100      vxlan10100            172.16.3.100    2        0        1

leaf2# show bgp evpn vni
Advertise All VNI flag: Enabled
Number of VNIs: 3
Flags: * - Kernel
  VNI        Orig IP              RD                  Import RT           Export RT
* 10200      172.16.3.100         172.16.3.2:1        65003:10200         65003:10200
* 10001      172.16.3.100         172.16.3.2:3        65003:10001         65003:10001
* 10100      172.16.3.100         172.16.3.2:2        65003:10100         65003:10100
```

Figura 71. Información de VNI, RD y RT en LEAF2

- **Rutas EVPN**

En la figura 72 se muestran las rutas EVPN para el dispositivo LEAF2. Las rutas de tipo 2 proveen información de conectividad de los

dispositivos finales y las rutas de tipo 3 indican que se está empleando el reenvío de datos mediante unicast.

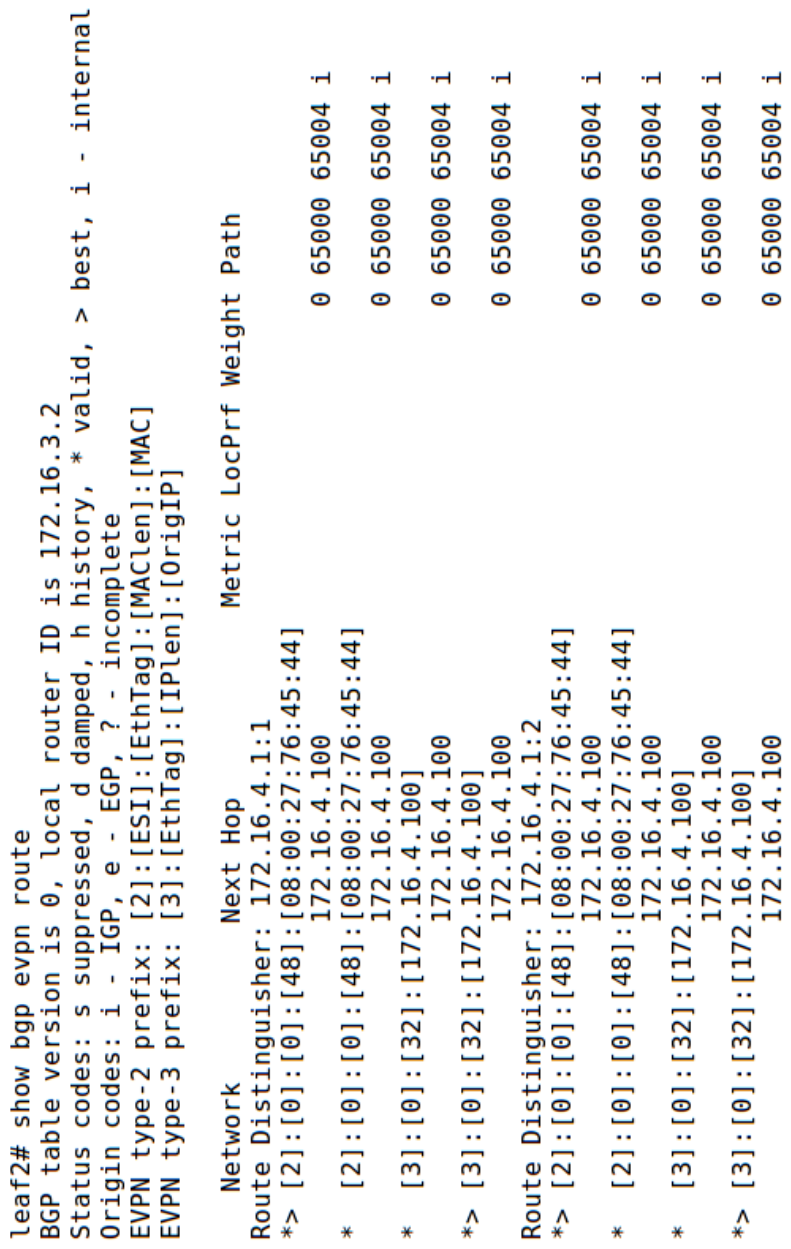


Figura 72. Rutas EVPN para LEAF2

CAPÍTULO V

PROPUESTA DE ADMINISTRACIÓN Y APROVISIONAMIENTO DE CLIENTES EMPLEANDO INTERCONEXIÓN VXLAN EN EL DATACENTER DE TELCONET S.A. DE LA CIUDAD DE QUITO

5.1. Introducción

En los diseños clásicos de redes jerárquicas, las capas de acceso y agregación proporcionan las funciones de Capa 2 y Capa 3 como elemento fundamental para la conectividad de los data centers. En los data centers más pequeños, este elemento principal proporcionaría suficiente escalabilidad para satisfacer los requisitos de conectividad y rendimiento. A medida de que los requisitos de escalabilidad crecen, este elemento principal es replicado con una capa de núcleo adicional para interconectarlos. Esto se refiere comúnmente a un Punto de Entrega (POD) y permite una escalabilidad modular consistente a medida que un data center crece.

Cuando se diseña una interconexión VXLAN, un único POD define una única interconexión VXLAN con una arquitectura spine-leaf escalable, tal como se indica en la figura 74.

Un VXLAN POD puede escalar a cientos de conmutadores y miles de puertos que cumplirán con la demanda de muchos entornos empresariales;

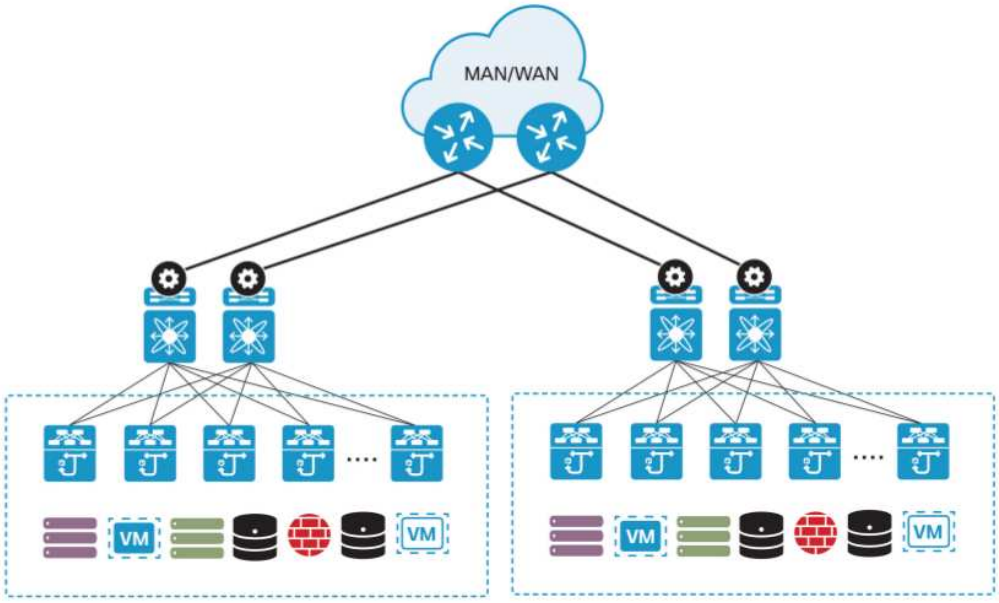


Figura 73. Diseño clásico de una red jerárquica

sin embargo, para satisfacer requisitos más complejos o de mayor escalabilidad, el VXLAN POD puede ser replicado en forma de un multi-POD. En un despliegue típico con múltiples ubicaciones de data centers, estas interconexiones VXLAN, ya sean exclusivas o basadas en POD, serán desarrolladas como un diseño VXLAN de varios sitios.

En este capítulo se describen las consideraciones de diseño para una futura implementación de un POD VXLAN en el Datacenter de Telconet de la ciudad de Quito basado en enfoques underlay y overlay. Además se indica un costo aproximado de la implementación y las ventajas que se obtendrán al realizar dicha inversión.

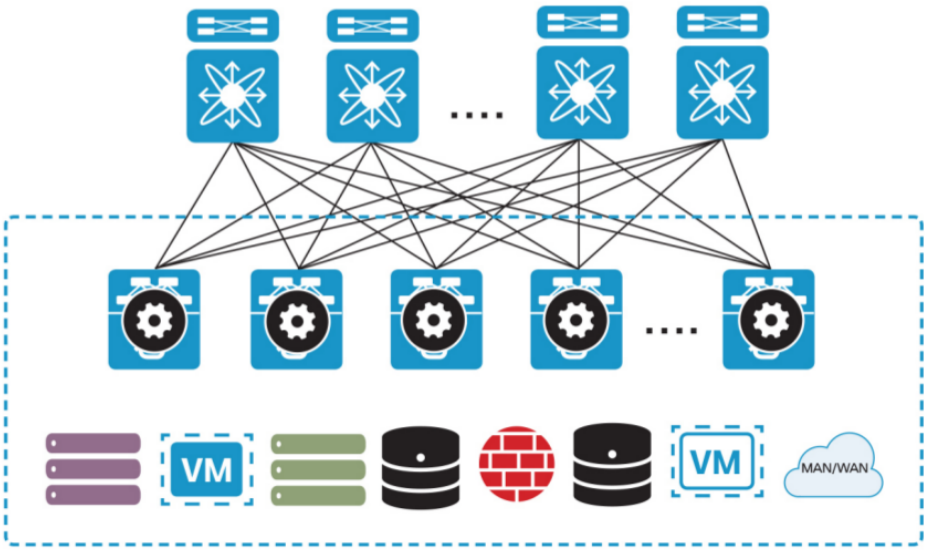


Figura 74. Interconexión VXLAN

5.2. Consideraciones Underlay

En la implementación de una interconexión VXLAN EVPN, es escencial contar con una red underlay que provea escalabilidad, disponibilidad y bases funcionales para soportar a la red overlay. En esta sección se analizan consideraciones importantes para el diseño de la red underlay.

5.2.1. Consideraciones para las interfaces ruteadas

- **Unidad máxima de transmisión (MTU)**

  Con el fin de mejorar el rendimiento de la red, se recomienda evitar la fragmentación y el reensamble de datos en los dispositivos que llevan a cabo la encapsulación y desencapsulación VXLAN. Por lo tanto se requiere aumentar la MTU en al menos 50 bytes (54 si está presente un encabezado 802.1Q en la trama encapsulada). Si el overlay usa una

MTU de 1500 bytes, la red de transporte debe estar configurado para alojar 1550 bytes (1554 bytes si se incluyen los encabezados 802.1Q) como mínimo. Se recomienda configurar también el Jumbo mtu en la red de transporte si las aplicaciones overlay utilizan tamaños de trama mayores a 1.500 bytes.

Con el fin de asegurarse de que los paquetes encapsulados con VXLAN pueden ser transportados exitosamente a través de la interconexión VXLAN , el aumento de MTU debe ser configurado en todas las interfaces de capa 3 que interconectan los nodos.

- **Direccionamiento de Capa 3**

  La conectividad entre los dispositivos de una interconexión VXLAN utiliza interfaces punto a punto direccionados con prefijos /30 ó /31. En la red underlay de un data center grande habrá varios enlaces ruteados, lo que conlleva a un alto consumo de direcciones IPV4.

  Citaremos el siguiente ejemplo:

  Una interconexión VXLAN requiere de 4 Spines y 6 Leafs. Para realizar el cálculo de la cantidad de direcciones IPV4 necesarias se realiza lo siguiente:

  - 4 Spines * 6 Leafs = *24 enlaces P2P*
  - 24 Enlaces * 2(/31) = *48 direcciones IPV4*
  - 10 direcciones loopback usadas como Router ID (RID)

- 10 direccones loopback usadas como direcciones IPV4 de VTEP

En total se utilizaría: $48 + 10 + 10 = 68$ direcciones IPv4

- **Direccionamiento IPV4 de interfaces Loopback**

  Como se indicó en el ite anterior, cada VTEP debe tener como mínimo 2 interfaces loopback. La primera interfaz loopback se utilizará como Router ID (RID) y la segunda representa la dirección de VTEP usada como orígen o destino para el tráfico encapsulado VXLAN.

5.2.2. Consideraciones de Enrutamiento

Al escoger un protocolo de enrutamiento para la red underlay hay varias opciones. Sin embargo, es importante considerar sus caracerísticas de convergencia. Especificamente, Open Short Path First (OSPF) e Intermediate System - Intermediate System (IS-IS) son dos tipos de protocolos IGP recomendados para interconexiones spine-leaf. Como el diseño spine-leaf provee múltiples caminos entre los leafs a través de los spines, los protocolos de estado de enlace calcuarán una topología consistente de múltiples caminos de igual costo a través de la red.

Al realizar esta selección, es imperativo considerar como las funciones de plano de control de la red overlay serán configuradas. Es recomendable usar un protocolo de enrutamiento distinto tanto para la red underlay como overlay, ya que al usar el mismo puede causar confusión.

- **Recomendaciones para el uso de OSPF**

  Al emplear OSPF en la red underlay, se recomienda que las redes sean de tipo *point-to-point*. Esto ayudará a evitar una sobregcarga innecesaria ya que no se ejecuta la elección del Designated Router (DR) y Backup Designated Router (BDR); así como el envío de LSA Tipo 2 a través de la red.

  Cabe indicar también que con redes tipo poit-to-point se reducirá el tiempo en el establecimiento de adyacencias entre los dispositivos spine y leaf.

- **Recomendaciones para el uso de IS-IS**

  Otro protocolo IGP de estado de enace es IS-IS. Este protocolo está ganando popularidad debido a una rápida convergencia en redes a gran escala aunque ha sido inicialmente desarrollada para ambientes de proveedores de servicio.

  IS-IS usa el protocolo de red sin conexión (CLNP) para la comunicación entre vecinos y no depende de IP. No hay cálculos en cambios de enlace y el cálculo SPF sólo se producen cuando hay un cambio de topología, lo cual ayuda con una con más rápida convergencia y estabilidad en la red. No se requiere cambios significativos en IS-IS para lograr una eficiente y rápida convergencia en la red underlay.

5.2.3. Recomendaciones para IP Multicast

IP Multicast provee un mecanismo eficiente para la distribución de tráfico multi-destino en una interconexión underlay. Para implementar multicast en una red underlay un protocolo de entutamiento PIM (Protocolo Independent Multicast) debe ser habilitado y debe ser consistente a través de todos los dispositivos de la red underlay. Los dos protocolos PIM más comunes son Sparse-Mode (PIM-ASM) y Bidireccional (PIM-Bidir). Esto implica la implementación de rendevouz points (RP).

Normalmente, los RP son implementados en los nodos SPINE, dada la posición central que tienen estos dispositivos en la interconexión VXLAN.
Es importante recordar que los VTEP representan el origen y destino del tráfico Multicast utilzado para transportar el tráfico BUM entre los puntos finales conectados a estos dispositivos.

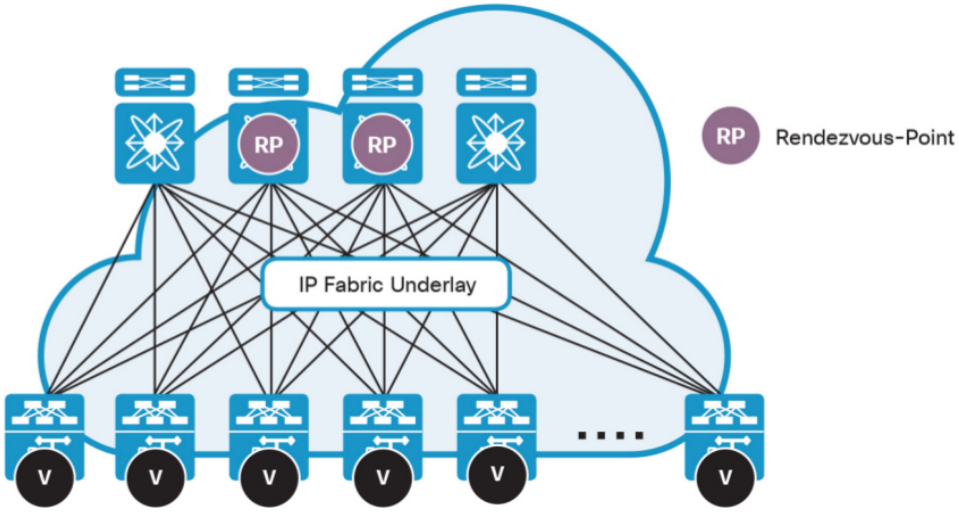


Figura 75. Ubicación de los RP en una VXLAN Fabric

5.2.4. Unicast Forwarding

Unicast Forwarding puede ser empleado como una alternativa a Multicast Forwarding para transportar tráfico BUM a través de la interconexión VXLAN.

Cuando se implementa unicast forwarding es importante considerar la escalabilidad total de la red y la cantidad esperada de tráfico multidestino. Esto es debido a que los VTEP usan una lista de direcciones de otros VTEP para enviar tráfico BUM como tráfico unicast, creando múltiples copiasdel mismo tipo de tráfico. Vale la pena indicar que la implementación del plano de control mediante MP-BGP habilita la lista de los VTEP conectados a la misma interconexión VXLAN. Estas direcciones IP son intercambiadas entre los VTEP a través del plano de control BGP EVPN.

5.3. Consideraciones Overlay

Luego de implementar una base sólida para VXLAN en la red underlay, las consideraciones overlay son también importantes para proporcionar la flexibilidad y funcionalidad requeridas.

5.3.1. Plano de Control VXLAN EVPN

EVPN es el plano de control para VXLAN y provee un método eficiente para el aprendizaje y distribución de rutas en la red VXLAN overlay. La información de enrutamiento incluye rutas MAC de capa 2, rutas IP de host a nivel de capa

3 y rutas IP de subredes a nivel de capa 3. El plano de control EVPN también introduce soporte multi-tenancy así como descubrimiento de VTEPs vecinos, seguridad y mecanismos de autenticación.

- **MP-BGP EVPN**

  EVPN usa MP-BGP como protocolo de enrutamiento para distribuir información de conectividad para la red VXLAN overlay incluyendo direcciones MAC de dispositivos finales, direcciones IP de dispositivos finales e información de conectividad de subredes.

- **Virtual Routing and Forwarding (VRF)**

  VRF define un dominio de enrutamiento de capa 3 para cada cliente en una interconexión VXLAN. En redes VXLAN EVPN, cada vrf de un cliente tiene un VNI de capa 3 usado como un backbone virtual para enrutar dentro de la vrf.

- **Route Distinguisher (RD)**

  Es el identificador de una vrf ya que cada vrf tiene un único RD en la red. Cuando un anuncio EVPN es enviado a los vecinos, el RD de la vrf al cual pertenece la ruta es antepuesto a la ruta original para hacerlo único dentro de la red. Esto permite que diferentes vrf utilicen direcciones ip superpuestas para que diferentes clientes puedan tener una verdadera autonomía en la administración de direcciones IP. El RD puede ser definido automáticamente para simplicar la configuración.

- **Route Target (RT)**

  Route Target es un atributo extendido en las actualizaciones de rutas EVPN usado para controlar la distribución de rutas en una red multi-clientes. Cada VTEP tiene un RT import y un RT export para cada vrf. Cuando un VTEP anuncia rutas EVPN, este añade su export RT a la actualización de la ruta. Las rutas serán recibidas por otros VTEP en la red. Éstos comparan en valor RT transportado con su valor de RT import. Si los dos valores coinciden, la ruta será aceptada y anunciada en la tabla de enrutamiento El RT puede ser definido automáticamente para simplicar la configuración.

- **Ubicación de los Route Reflectors**

  La ubicación de los route reflectors se recomienda implementarla en los dispositivos SPINE. Es así que dos SPINES tendrán configuración de BGP route reflector y todos los VTEP serán configurados como BGP route reflector clients. El route reflector reflejará las rutas EVPN para los VTEP.

5.4. Esquema de implementación

Luego de conversaciones realizadas con la Gerencia y Jefatura del área de Networking de Telconet S.A., se determinó que la implementación de la tecnología VXLAN será netamente para clientes del Data Center.

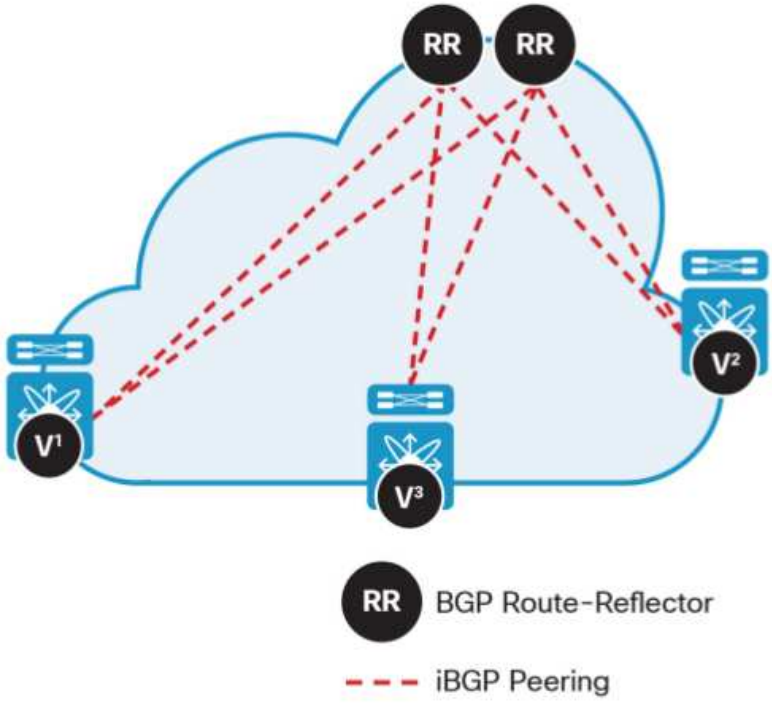


Figura 76. Ubicación de los RR en una VXLAN Fabric

Los elementos a usarse en dicha implementación del Datacenter de Quito son:

- Dos switches NEXUS 9500 (SPINES)
- Dos switches NEXUS 9300 (LEAFS)
- La interconexión entre los SPINES y LEAFS se realizará mediante enlaces de 40G.

En la figura 77 se indica el esquema de implementación multi-pod propuesto tanto para el datacenter de Quito como para el de Guayaquil. La interconexión de los Data Centers así como la conectividad externa se realizará a través de

la red MPLS de Telconet S.A.

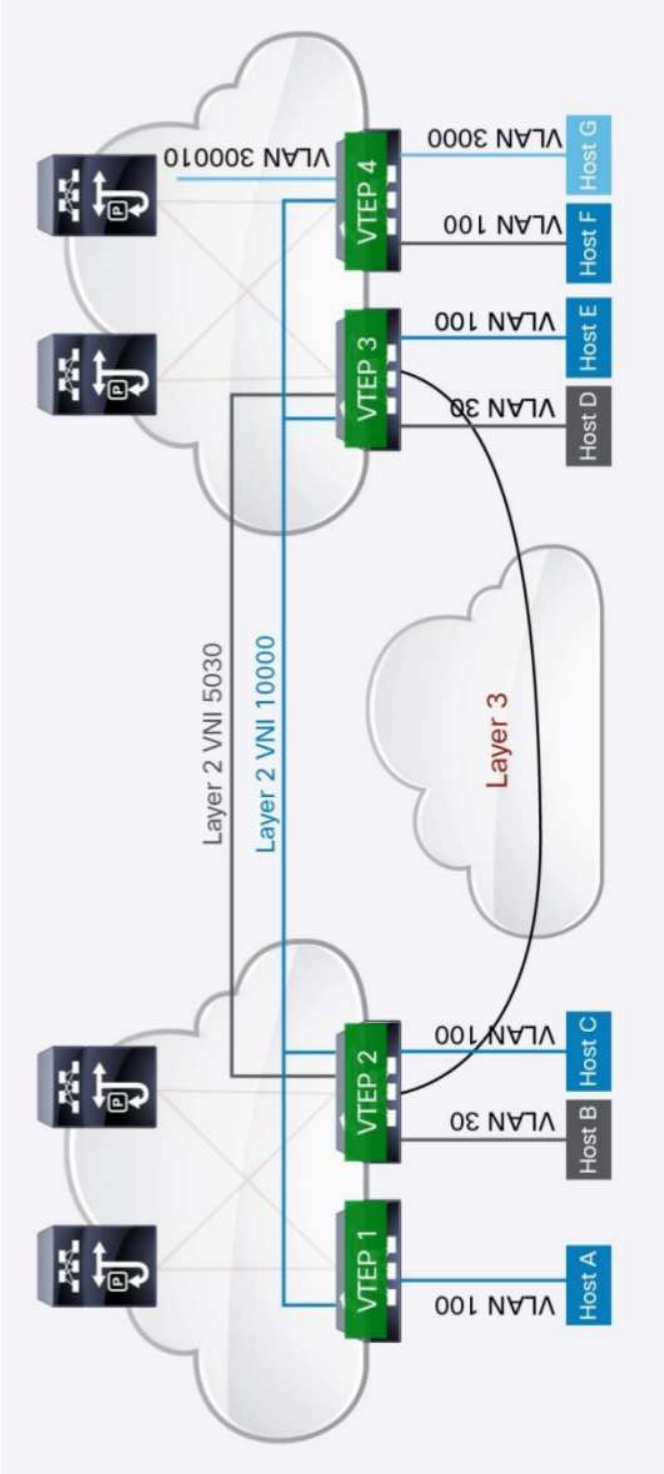


Figura 77. Esquema de implementación de VXLAN para Telconet S.A.

No se consideró implementar VXLAN en el resto de la red corporativa debido a que varios dispositivos no soportan la address-family evpn; por tanto, no se podría implementar el plano de control.

El costo aproximado para la implementación de cada POD empleando equipamiento Cisco es USD 250000; es decir que para la interconexión de los dos Data Centers la inversión a realizar será de USD 500000. Cabe indicar que al momento de emitir esta información, ningún otro vendor aparte de Cisco ha emitido una cotización formal para esta implementación; pero debido al monto de la misma se requerirá al menos dos cotizaciones más de otros vendors para definir la adquisición respectiva.

5.5. Consideraciones para el aprovisionamiento de clientes

En la actualidad, el esquema de aprovisionamiento para los clientes del Data Center de Quito es a través de VLANs. De acuerdo con información obtenida del área de TI de Telconet S.A., el rango de VLANs utilizadas es del 50 %, es decir que aproximadamente sobrarían 2000 VLANs.

Considerando la demanda de servicios mediante virtualización y cloud computing, es recomendable migrar progresivamente al esquema de implementación mediante VXLAN.

De acuerdo con la teoría revisada en el capítulo 1, se dispone de 2^{24} VNID, lo cual elevaría el nivel de escalabilidad para aprovisionamiento de clientes.

Como VXLAN se basa en entunelamiento, las redes LAN de los clientes pueden utilizar direccionamiento privado para interconexión, esto ayudaría en gran medida a la optimización en el uso de direccionamiento IPV4 público. Queda a consideración del área de networking de Telconet S.A. determinar el método más adecuado de administración y asignación de los VNID para los clientes del Data Center.

Otra consideración adicional es el empleo de una controladora ACI (Application Centric Infrastructure) para el aprovisionamiento de servicios a través de VXLAN. El propósito del uso de esta controladora es automatizar el proceso de administración de clientes en el Data Center de Quito y minimizar los errores a causa del factor humano ya que las configuraciones de VXLAN a través de CLI representan una gran carga administrativa para los operadores de los dispositivos SPINES y LEAFS.

Entre las ventajas del uso de la controladora ACI tenemos las siguientes:

- Punto único de aprovisionamiento a través de interfaz gráfica de usuario (GUI).
- Conectividad para cargas de trabajo físicas y virtuales con visibilidad completa en el tráfico de la máquina virtual.
- Compatibilidad de hypervisors e integración sin la necesidad de agregar software al hypervisor.

- Facilidad (y velocidad) de despliegue.
- Simplicidad de la automatización.
- Multitenancy.
- Capacidad para crear plantillas de configuración portátiles
- Seguridad basada en hardware.
- Eliminación de las inundaciones de tráfio en la interconexión
- Facilidad de amapear arquitecturas de aplicaciones a la configuración de red
- Capacidad para insertar y automatizar cortafuegos, balanceadores de carga y otros servicios L4-7.
- Proceso de configuración intuitivo y fácil.

El acceso a esta controladora estará basado en privilegios de usuario, es decir, de acuerdo a la función de cada área se asignarán permisos de lectura y/o escritura para la administración de la interconexión VXLAN.

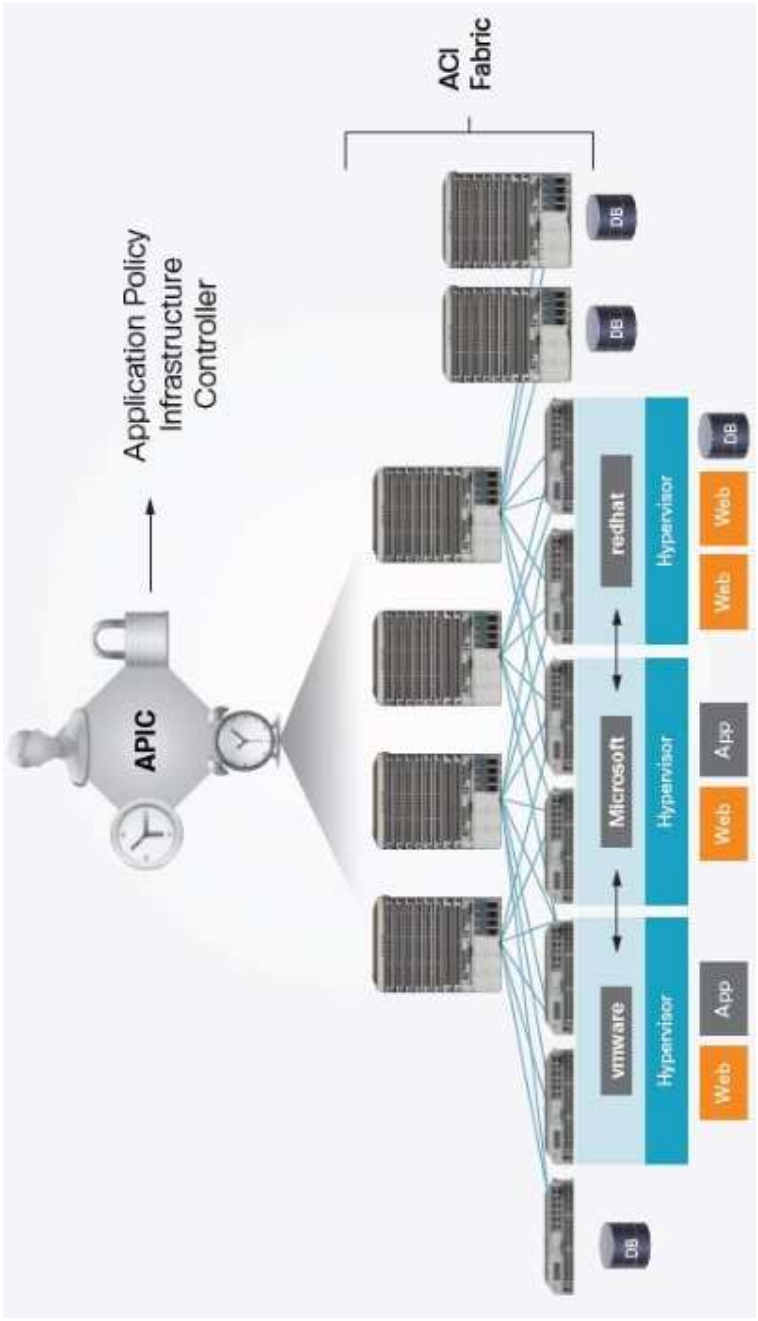


Figura 78. [ACI Fabric]. Recuperado de http://adaptingit.com/aci-101-fabric-discovery/

CONCLUSIONES

- Virtual Extensible LAN (VXLAN) es una tecnología que permitirá obtener grandes niveles de escalabilidad en comparación con las VLAN tradicionales. Además con las mejoras desarrolladas para su implementación se minimizará el envío innecesario de tráfico broadcast logrando así un mejor rendimiento en las interconexiones de red.

- Al realizar la simulación de VXLAN en un entorno virtualizado se demostró la hipótesis planteada en el presente trabajo. Se pudo establecer conectividad entre equipos finales separados remotamente empleando el mismo segmento de red; lo que en redes estáticas tradicionales es bastante complicado obtener. Ésto permitirá usar direccionamiento privado (RFC 1918) para interconectar clientes y de esta manera contribuir con la optimización en el uso de direccionamiento público IPV4.

- La futura implementación de VXLAN en Telconet S.A. estará orientada netamente a clientes administrados en el Data Center de la ciudad de Quito con réplica en Guayaquil. Si bien la inversión para la implementación de esta tecnología es alta, ésto le permitirá tener tecnología de punta, mantener ventaja competitiva frente a otros

proveedores de servicio y ofrecer mayor diversidad de servicios a los clientes.

LÍNEAS DE TRABAJO FUTURO

- Extender el alcance de esta investigaciòn usando direccionamiento IPV6 en la implementación del entunelamiento VXLAN, protocolos de enrutamiento y uso del plano de control mediante MP-BGP.
- Analizar el comportamiento de interconexión de redes con Generic Protocol Encapsulation (VXLAN-GPE), mismo que constituye la nueva evolución de VXLAN y su posible implementación con segment route como alternativa a MPLS.

ANEXO 1

ECUADOR TECHNICAL CHAPTERS MEETING 2017

Del presente trabajo, se elaboró una publicación para ser expuesta en el evento Ecuador Technical Chapters Meeting `http://sites.ieee.org/etcm-2017/`, llevado a cabo del 16 al 20 de Octubre del 2017.

El tema de la publicación fue: ´´Underlay and Overlay Networks: The approach to solve addressing and segmentation problems in the new networking era´´ y sus autores son Edison Naranjo y Gustavo Salazar.

BIBLIOGRAFÍA

Arizmendi, L. (2014). *Ideaas para Data Centers.* Obtenido de http://luisarimendi.blogspot.com

Colomes, P. (2015). *Sdn: Qué es vxlan y cómo funciona.* Obtenido de http://www.redescisco.net/sitio/2015/11/24/sdn-que-es-vxlan-y-como-funciona/

Jansen, D. (2017). *Building data centers with vxlan bgp evpn.* (1ra ed.,Vol. 1). San José, California: CiscoPress.

Bolton, D. (1975). *El empleo de la simulación en la administración educacional.* México, DF: Paidos.

Rodríguez, M., & Quesada, L. (2009). *La simulación computarizada como herramienta didáctica de amplias posibilidades.* Obtenido de http://www.rcim.sld.cu/revista 18/articulos pdf/simulacioncomputarizada.pdf

Jansen, D., & Kratigger, L. (2017). *A Modern, Open and Scalable Fabric VXLAN EVPN.* San José, California: CiscoPress.