LACNIC 45 · FORO TÉCNICO LATINOAMERICANO

EVPN & VXLAN el nuevo idioma del peering moderno

Cómo el control-plane BGP y los overlays L2 están reescribiendo la forma en que los IXPs e ISPs construyen sus redes.

SPEAKER
Ariel S. Weher

AFILIACIÓN
Ayuda.LA · LACNOG

DURACIÓN
25 minutos

Agenda

01	**Por qué necesitamos evolucionar el peering** El L2 clásico ya no escala con el peering moderno.	3 min
02	**Qué aportan EVPN y VXLAN** Data-plane, control-plane y la combinación.	4 min
03	**Arquitecturas típicas en IXPs/ISPs** Spine-leaf, multihoming, segmentación.	4 min
04	**Cómo se ve en la práctica** 4 escenarios reales con configs y outputs.	7 min
05	**Casos reales y un caso hipotético** Interconexión inter-IXP, multi-transporte y el IXP virtual.	4 min
06	**Recomendaciones prácticas** Buenas prácticas y cómo empezar.	3 min
07	**Conclusiones y preguntas** Lo que se llevan + Q&A.	2 min

MÓDULO 01

Por qué necesitamos evolucionar el peering

El peering creció. La red que lo sostiene, en muchos casos, no.

El contexto actual del peering

MÁS TRÁFICO, MÁS PARTICIPANTES

Crecimiento exponencial de bits y de peers conectados al mismo fabric. Lo que hace 10 años entraba en un par de switches hoy ya no.

MULTIVENDOR POR DEFAULT

Cisco, Juniper, Nokia, Arista, FRR, Cumulus. La interoperabilidad dejó de ser opcional —es una condición del negocio.

LÍMITES DEL L2 CLÁSICO

STP, flooding, broadcast, MAC learning por inundación. Funciona, pero no escala ni en peers ni en convergencia.

AUTOMATIZACIÓN Y RESILIENCIA

Configurar a mano ya no es viable. Necesitamos redes que un script entienda y que un fallo no rompa.

Los dolores reales del peering L2 hoy

ARP Y ND NO ESCALAN CON LOS PEERS

El descubrimiento crece de forma cuadrática con la cantidad de miembros. Cada peer nuevo es ruido para todos —y nadie tiene control sobre cuánto.

UN LOOP DE UN MIEMBRO LO PAGA EL IXP ENTERO

Broadcast storms, BUM excesivo, loops accidentales: el dominio de falla es la peering LAN completa. Mitigás con MAC ACLs y rate-limit, pero el problema conceptual sigue.

EL FABRIC NO RECUERDA QUIÉN ES QUIÉN

MAC spoofing, proxy-ARP no autorizado, MACs que reaparecen en otro puerto. Sin control plane explícito no hay tracking de movilidad ni policy nativa.

CUANDO ALGO SE ROMPE, TODO SE ENTERA

STP recalcula en segundos. BGP flapea porque el ARP del peer expiró mal. Una sola conexión al IXP es un SPOF garantizado.

A DÓNDE VAMOS

EVPN/VXLAN ataca los cuatro frentes de raíz: ARP suppression por leaf, BUM controlado por VNI, MAC duplicate detection nativo, y multi-homing all-active sin STP ni protocolo propietario. El control plane BGP reemplaza al flood-and-learn.

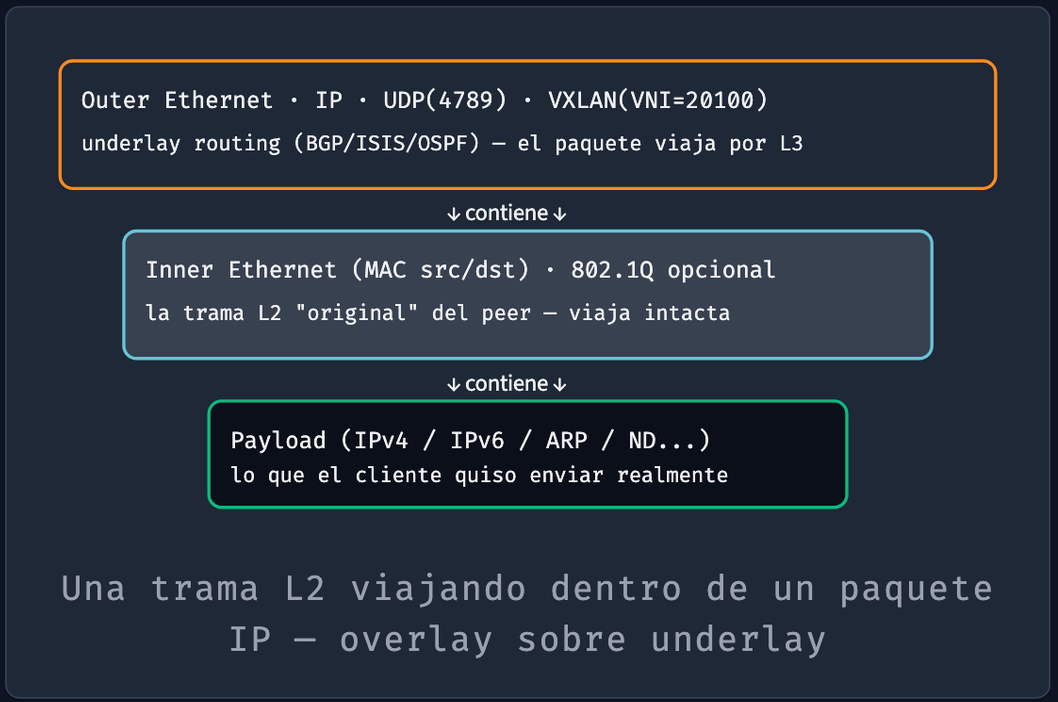
MÓDULO 02

Qué aportan EVPN y VXLAN

Dos piezas que se complementan: una mueve los paquetes, la otra decide cómo.

VXLAN en 2 minutos

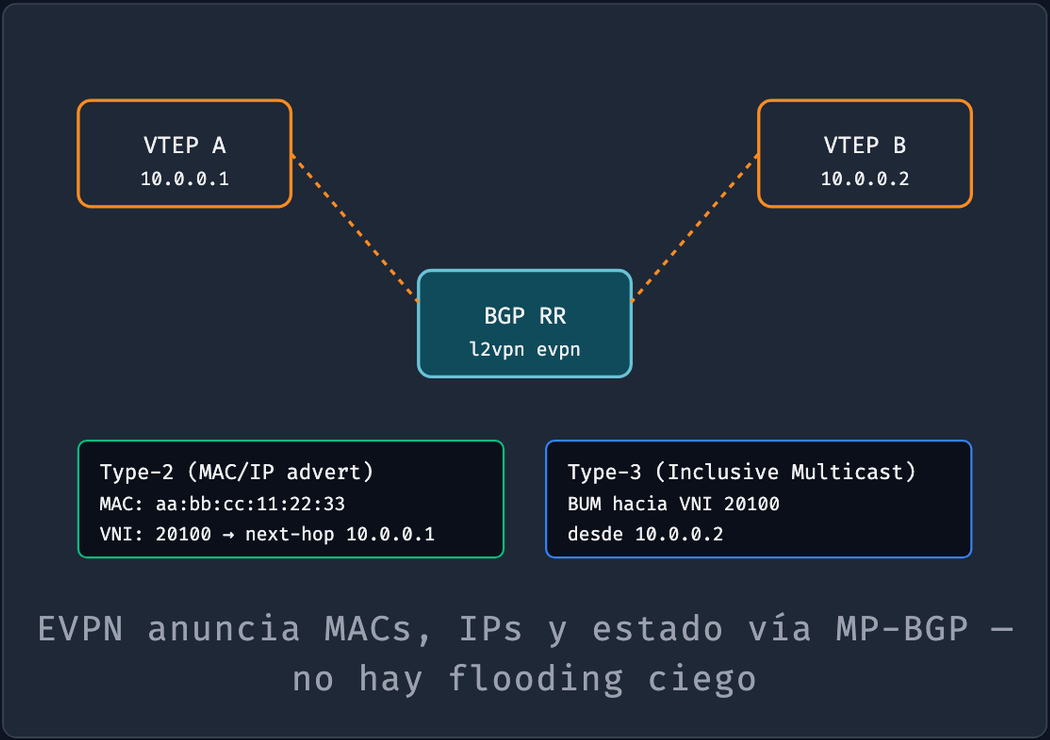
- Encapsula tramas L2 dentro de UDP/IP (puerto 4789).
- **VNI** de 24 bits → hasta **16M de segmentos** (vs. 4.094 VLANs).
- Extiende L2 sobre cualquier red L3 enrutada.
- Permite crecer sin rediseñar el underlay físico.
- Ideal para separar clientes, peers o servicios.



Una trama L2 viajando dentro de un paquete IP — overlay sobre underlay

EVPN en 2 minutos

- Control-plane basado en **MP-BGP** (familia `l2vpn evpn`).
- Reemplaza el aprendizaje por inundación: las MAC se anuncian, no se descubren.
- Distribuye MAC, IP, rutas, etiquetas y estado de los enlaces.
- Habilita **multihoming activo-activo** sin STP (ESI).
- Predecible, observable y debuggeable como cualquier BGP.



EVPN + VXLAN: la combinación que cambia las reglas

EVPN

Control-plane. Quién está, dónde está, cómo se llega.

VXLAN

Data-plane. Cómo viajan los bits.

JUNTOS

Overlay escalable sin STP, con segmentación limpia y base para automatización.

LA IDEA CLAVE

Separamos lo que cambia mucho (el L2 del peering, los clientes) de lo que cambia poco (el underlay IP). El underlay se vuelve aburrido y estable. El overlay, flexible y automatizable.

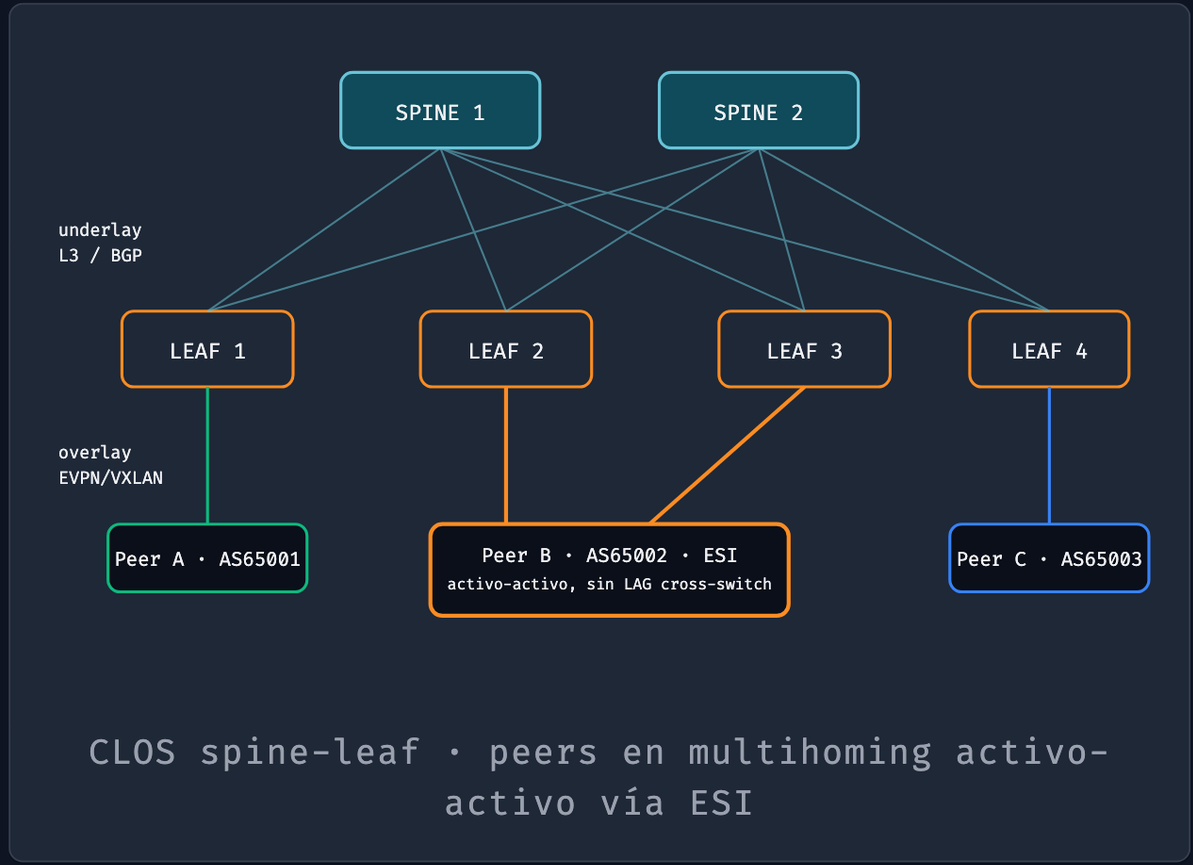
MÓDULO 03

Arquitecturas típicas en IXPs/ISPs

Spine-leaf, multihoming activo-activo y segmentación por VNI.

¿Cómo se podría aplicar en un IXP?

- **VNI/VRF por participante** — podríamos darle a cada uno segmentación limpia.
- **Broadcast controlado**: dejaría de haber tormentas al agregar un peer.
- **Multihoming** en puertos diferentes *sin* STP (ESI / All-Active).
- Mejor resiliencia ante loops y fallas.
- Menos cambios disruptivos para crecer.



CLOS spine-leaf · peers en multihoming activo-activo vía ESI

¿Y en un ISP?

Algunas formas en que *se podría* aprovechar el mismo modelo:

EDGE Y BNG MODERNOS

Migraciones de borde sin tocar el core. El overlay sigue al cliente, no al hardware.

SEGMENTACIÓN ESCALABLE

Cada cliente, cada servicio, cada VPN en su propio VNI. Sin pelearse con el espacio de VLAN.

IPV4 + IPV6 COEXISTIENDO

El overlay no discrimina. Dual stack sin rediseñar nada.

CONVERGENCIA MÁS RÁPIDA

BGP en el control-plane → reroute en sub-segundos contra reconvergencia STP.

LISTO PARA SDN Y TELEMETRÍA

gNMI, BMP, sFlow/IPFIX se integran nativamente. La red habla, no solo responde.

AUTOMATIZACIÓN DESDE DÍA CERO

Templates por VNI, integración con NetBox/Ansible. La red como código.

RD y RT · qué separa y qué une

RD — ROUTE DISTINGUISHER

Hace **única** cada ruta por su origen. Es `router-id : índice-VNI`. **No** decide el import.

RT — ROUTE TARGET

Es la **política** de import/export. Autoderivado `ASN : VNI` en FRR con `advertise-all-vni`.

EL RD SEPARA, EL RT UNE

Mismo VNI → RD *distinto* por VTEP, RT que define a qué tenant pertenece.

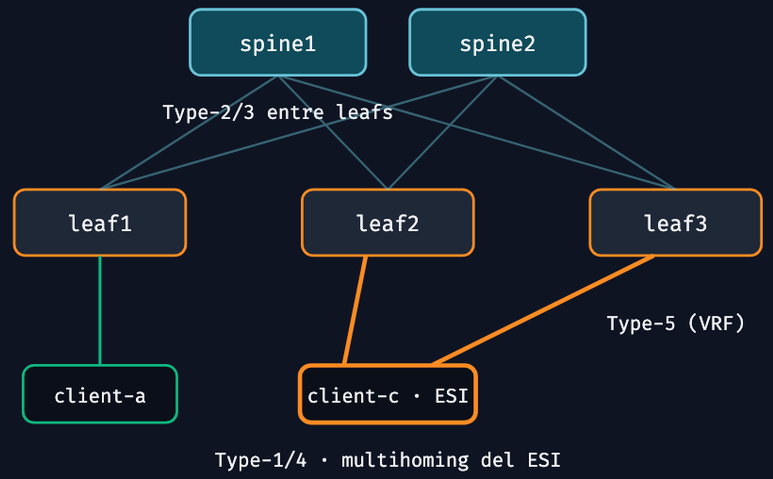
lab FRR 10.2.1 · show bgp l2vpn evpn vni

```
# mismo VNI 10100 (tenant-A) en los 3 VTEPs
       VNI   Type  RD              RT (imp=exp)
leaf1 10100 L2    10.0.0.11:2  65011:10100
leaf2 10100 L2    10.0.0.12:2  65012:10100
leaf3 10100 L2    10.0.0.13:2  65013:10100
# RD distinto por VTEP = ruta unica por origen
# RT difiere aca: fabric eBGP, 1 ASN/leaf
```

Los 5 route-types · el vocabulario de EVPN

BGP EVPN no anuncia «rutas» a secas: anuncia **tipos de noticia**, cada uno resuelve un problema puntual del fabric.

- **L2 / aprendizaje** → Type-2 (MACs) y Type-3 (BUM).
- **L3 / ruteo** → Type-2 con IP (hosts) y Type-5 (prefijos).
- **Multihoming** → Type-1 (convergencia) y Type-4 (DF election).



Todo lo que sigue es del lab · FRR 10.2.1 · `client-c` en LAG all-active a dos leaves

Type-2 · MAC/IP Advertisement

EL CABALLO DE BATALLA

Anuncia una MAC, opcionalmente con su IP, su VNI y el VTEP origen. El que más vas a ver.

- **L2:** aprendizaje de MACs por control plane → adiós flood-and-learn.
- **L3:** binding MAC+IP → supresión de ARP/ND y ruteo simétrico.

leaf1 · route type macip · VNI 10100

```
# MAC sola (L2 puro)
[2]:[0]:[48]:[aa:c1:ab:f6:4b:93]
# MAC + IPv4  -> ARP suppression
[2]:[0]:[48]:[44:38:39:ff:00:64]:[32]:[100.64.0.1]
# MAC + IPv6  -> ND suppression
[2]:..:[128]:[fe80::28f8:39ff:fe6f:9af]
# MAC remota (origen leaf2)
RD 10.0.0.12:2  [2]:..:[aa:c1:ab:d8:e6:f7]
```

Type-3 · Inclusive Multicast (IMET)

EL TRÁFICO BUM

Broadcast, Unknown-unicast, Multicast. Cada VTEP que entra a un VNI dice «mandáme el BUM de este VNI a mi IP».

- Construye la lista de **ingress replication** (head-end).
- Es el «hola, estoy en este VNI» de cada VTEP.

leaf1 · route type multicast

```
# cada VTEP del VNI 10100 se anuncia
RD 10.0.0.11:2  [3]:[0]:[32]:[10.0.0.11]
RD 10.0.0.12:2  [3]:[0]:[32]:[10.0.0.12]
RD 10.0.0.13:2  [3]:[0]:[32]:[10.0.0.13]
# 3 VTEPs -> replican BUM entre si
```

Type-5 · IP Prefix Route

PREFIJOS, NO MACS

Anuncia **prefijos IP puros**, desacoplados de cualquier endpoint. El L3 hacia afuera del overlay.

- Subredes externas, sumarización, default, rutas a un borde/firewall.
- **vs Type-2:** Type-2 = host /32 *con* MAC; Type-5 = prefijo *sin* MAC.

CLAVE EN INTERCONEXIÓN

El IXP virtual del módulo 5 se apoya fuerte en Type-5.

leaf1 · route type prefix · L3VNI

```
# prefijos /24 puros, sin MAC asociada
RD 10.0.0.12:4  [5]:[0]:[24]:[100.64.0.0]
RD 10.0.0.13:8  [5]:[0]:[24]:[172.16.100.0]
# vs Type-2: [32]:[100.64.0.1] = host + MAC
```

Type-1 · Ethernet Auto-Discovery

EL ROUTE-TYPE DEL MULTIHOMING

Coordina a los VTEPs que comparten un mismo cliente (ESI). Dos trabajos:

- **Mass withdrawal** (per-ES): si se cae el enlace, *una* retirada por ESI → convergencia inmediata, sin retirar miles de MACs.
- **Aliasing** (per-EVI): «a este peer también llego yo» → habilita ECMP all-active.

SOLO APARECE CON ESI

En un fabric sin multihoming, Type-1 ni existe.

leaf1 · route type ead

```
# A-D per-EVI  (EthTag 0) -> aliasing / ECMP
[1]:[0]:[03:..:01]:[32]:[0.0.0.0]
     RT:65012:10100 ET:8
# A-D per-ES  (EthTag MAX-ET) -> mass withdrawal
[1]:[4294967295]:[03:..:01]:[32]:[0.0.0.0]
     ET:8 ESI-label-Rt:AA
```

Type-4 · Ethernet Segment Route

DESCUBRIMIENTO DE ES + DF ELECTION

Los VTEPs que comparten un ESI se descubren entre sí y eligen el **Designated Forwarder**.

- Sin DF, el BUM se entregaría *dos veces* al peer all-active.
- El DF decide quién reenvía BUM hacia ese segmento.

ESTO REEMPLAZA AL MLAG/VPC

Sin plano privado entre switches. Coordinan vía BGP. Multivendor.

leaf1 · route type es · show evpn es

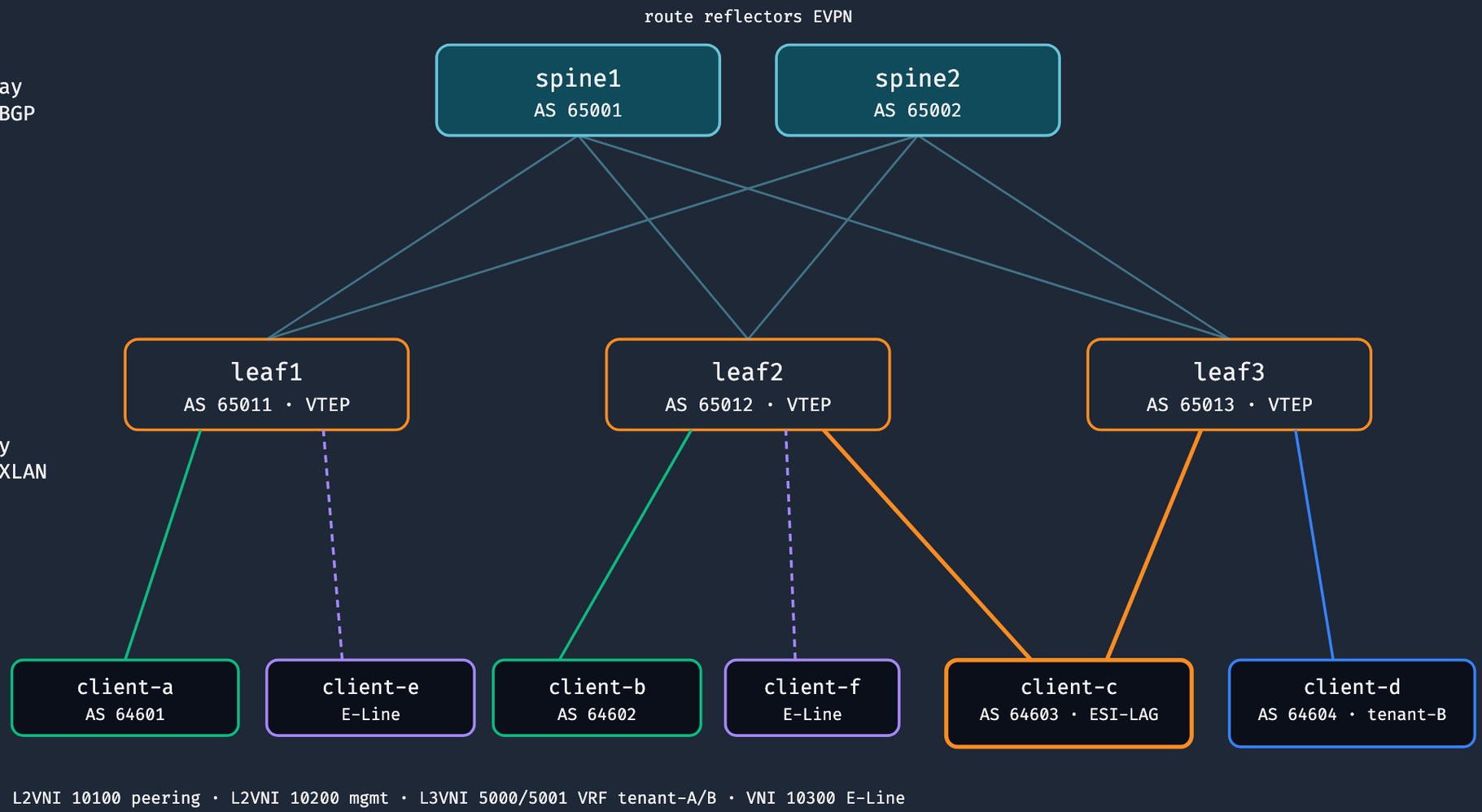
```
[4]:[03:02:..:0c:00:00:01]:[32]:[10.0.0.12] leaf2
[4]:[03:02:..:0c:00:00:01]:[32]:[10.0.0.13] leaf3
# ESI 03:.. = autoderivado de LACP
DF: leaf2 pref 50000 (DF) | leaf3 pref 40000
```

MÓDULO 04

Cómo se ve en la práctica

Cuatro escenarios reales con configs, outputs y route-types.

El lab que vamos a usar



Topología del lab · FRR 10.2.1 · client-c en ESI-LAG all-active a leaf2 + leaf3

Escenario 1 · IXP recibiendo un nuevo peer

leaf2 · FRR 10.2.1

```
# el puerto del peer entra al bridge del VNI
ip link set eth3 master br10100
# EVPN ya esta activo en todo el fabric:
router bgp 65012
 address-family l2vpn evpn
  advertise-all-vni
# no hay que tocar nada por-peer ni por-VNI
leaf2# show evpn mac vni 10100
aa:c1:ab:e7:ee:a1 local  eth3
aa:c1:ab:f6:4b:93 remote 10.0.0.11
```

Qué pasa cuando el peer manda su primera trama:

1. La MAC se aprende localmente en leaf2 (la ves como `local`).
2. leaf2 emite Type-2 (MAC/IP) anunciando esa MAC en el VNI 10100.
3. leaf2 ya tenía su Type-3 (IMET) — está en el árbol BUM del VNI.
4. El resto del fabric programa FIB y BUM solo.
5. Los demás peers lo ven en segundos, sin tocar nada.

ONBOARDING CERO-TOQUE

Con `advertise-all-vni` el VNI ya se anuncia. Sumar un peer = un puerto al bridge.

Escenario 2 · Multihoming activo-activo (ESI)

leaf2 · FRR 10.2.1

```
# bond LACP hacia el peer + 3 lineas de MH
interface bond-esi10
 evpn mh es-id 1
 evpn mh es-sys-mac 02:00:00:00:00:0c
 evpn mh es-df-pref 50000
leaf2# show evpn es
ESI                         Type ES-IF
03:02:..:0c:00:00:01        LR   bond-esi10
 DF status: df   DF preference: 50000
 VTEPs: 10.0.0.13 df_pref: 40000
```

Qué pasa cuando se cae `leaf2 ↔ peer`:

T+0ms

Falla del link

leaf2 detecta down del bond hacia el peer.

T+50ms

Mass Withdraw

leaf2 retira **Type-1 EAD-ES** por BGP.

T+150ms

Reconvergencia

El resto apunta todas las MACs del ESI a leaf3.

T<1s

Tráfico restablecido

Sin STP, sin LAG cross-chassis, sin trucos propietarios.

Escenario 3 · Troubleshooting con route-types EVPN

spine-1 · FRR

```
$ vtysh -c 'show bgp l2vpn evpn'
BGP table version is 47
Route Distinguisher: 10.0.0.2:100
*> [2]:[0]:[48]:[aa:bb:cc:11:22:33]
      10.0.0.2  ← MAC en leaf-2
*> [2]:[0]:[48]:[aa:bb:cc:11:22:33]:[32]:[192.0.2.10]
      10.0.0.2  ← MAC+IP, mismo host
*> [3]:[0]:[32]:[10.0.0.2] ← IMET (BUM)
*> [2]:[0]:[48]:[aa:bb:cc:11:22:33]
      10.0.0.5  ← ¡MAC duplicada!
```

Mapa mental de route-types:

- Type-1 EAD/ES → multihoming, anuncios per-EVI y per-ES.
- Type-2 MAC/IP → "quién está dónde". El 80% del troubleshooting.
- Type-3 IMET → "quién participa del BUM" del VNI.
- Type-4 ES discovery → "qué leafs comparten un ESI".
- Type-5 IP Prefix → ruteo L3 en el overlay (IRB).

MAC MOBILITY / DUPLICADA

Si ves el mismo MAC en dos next-hops, EVPN tiene **MAC mobility** con sequence number. Si llega al límite → MAC *frozen*. El log te dice qué leaf empezó el problema.

Escenario 4 · Servicio L3 entre sedes (L3VNI / Type-5)

leaf1 · FRR 10.2.1

```
# tenant-A: routing L3 entre subredes
vrf tenant-A
 vni 5000
router bgp 65013 vrf tenant-A
 address-family ipv4 unicast
  redistribute connected
leaf1# show ip route vrf tenant-A 100.64.2.0/24
B>* 100.64.2.0/24 [20/0] vrf tenant-A
    via 10.0.0.12 onlink  (leaf2)
    via 10.0.0.13 onlink  (leaf3)  <- ECMP
```

Qué resuelve el L3VNI:

- Cada subred del tenant vive en un leaf; el **Type-5** las anuncia entre sedes.
- El prefijo se instala con **ECMP** hacia los dos leaves que lo originan.
- El tráfico inter-subnet se rutea en el fabric (routing simétrico), no se inunda.
- Para el ISP: **un VRF + un L3VNI por cliente**, sin VPLS ni pseudowires a mano.

L3 SOBRE EL MISMO FABRIC

El mismo control-plane que mueve el peering te da el routing entre sedes del cliente. Y todo es FRR — lo que corre en el lab.

MÓDULO 05

Casos reales y un caso hipotético

De lo que ya pasa hoy hasta el IXP virtual federado.

Qué podés hacer con EVPN en un IXP

Un abanico de posibilidades, de lo más conservador a lo más ambicioso. La mayoría de estos casos están demostrados en el lab; el resto son extensiones del mismo modelo.

FAMILIA 1 · EMPEZAR POR ADENTRO

Modernizar el fabric sin tocar lo que ven los miembros.

SOLO EL FABRIC INTERNO

EVPN/VXLAN puertas adentro; los miembros siguen viendo la misma peering LAN clásica. Cero impacto para ellos.

IP FABRIC SPINE-LEAF

Cada leaf es un VTEP; los spines solo enrutan el underlay IP. Topología CLOS predecible.

ELIMINAR STP DEL CORE

Underlay IP con ECMP y control-plane limpio (BGP+EVPN). Adiós spanning-tree en el core.

ESCALAR LAS MAC TABLES

Repartir el dominio L2 en VNIs; menos flood & learn en todo el fabric.

UNA VLAN DE PEERING = UN VNI

Mapear cada VLAN (v4, v6, RS…) a un VNI, aislando lógicamente cada tipo de tráfico.

Servicios a miembros e infraestructura propia

FAMILIA 2 · SERVICIOS A MIEMBROS

El IXP como proveedor de servicios, no solo de una LAN.

E-LINE / E-LAN POR MIEMBRO

Servicios L2 sobre VNIs dedicados, como una L2VPN sobre la red del IXP.

VPN/VRF POR MIEMBRO

Servicios L3 tipo Metro/ISP: cada VRF mapeada a uno o más VNIs.

VNIS PARA SERVICIOS ESPECIALES

VLANs dedicadas entre pocos miembros, testbeds, blackhole LAN, redes privadas de CDNs.

ANYCAST GATEWAY EN LOS LEAFS

Gateway L3 distribuido sin FHRP: la misma IP/MAC viva en todos los leafs.

FAMILIA 3 · INFRAESTRUCTURA PROPIA

El IXP usa EVPN para sus propias redes internas.

L3VNI PARA SERVICIOS DEL IXP

VRFs + L3VNI para management, RS, DNS, NTP, RPKI, portales de miembros.

AUTOMATIZACIÓN FUERTE

Plantillas BGP/EVPN/VXLAN; provisioning de VNIs, VRFs y puertos de miembros.

SEGURIDAD GRANULAR POR VNI

Filtros, policing, DHCP/ARP inspection, anti-MAC-spoofing a nivel del leaf.

El IXP distribuido entre varios PoPs

FAMILIA 4 · MULTI-SITE / DISTRIBUIDO

Estirar el IXP entre PoPs sobre transporte de terceros.

PEERING LAN ENTRE POPS

Extender la peering LAN entre sitios vía VXLAN sobre la red IP/MPLS de transporte.

MENOS BUM ENTRE SITIOS

ARP/ND suppression de EVPN evita saturar los enlaces inter-PoP con flooding.

MOVILIDAD ENTRE RACKS/POPS

Mover equipos entre racks o PoPs sin cambiar la lógica de la peering LAN para los miembros.

HONESTIDAD TÉCNICA

Multi-site es lo más potente *y* lo más delicado: estirar un dominio L2 entre sitios agranda el blast radius. Latencia, MTU end-to-end y el tamaño del dominio BUM importan. A veces conviene un PNI virtual puntual antes que estirar todo.

Casos típicos en interconexión moderna

INTER-IXP SOBRE TRANSPORTE DE TERCERO

Dos o más IXPs intercambian tráfico vía un carrier externo, con latencias, anchos de banda y caminos que no controlás.

→ EVPN-VPWS para PNI virtuales punto a punto. Si se estira VNI, atención a MTU end-to-end y BGP Link Bandwidth para BW asimétricos.

MIEMBRO MULTI-TRANSPORTE AL MISMO IXP

El miembro llega por dos uplinks distintos —fibra propia + carrier, o dos carriers con BW y latencia dispares.

→ EVPN-MH con ESI-LAG. Activo/activo si son comparables; activo/standby con DF preference si son asimétricos.

OPERADOR CONECTADO A MÚLTIPLES IXPS

Mismo operador, varios IXPs, decide políticamente por dónde sale cada prefijo.

→ EVPN dentro de cada IXP es transparente. Es problema de routing del cliente, no del fabric.

LO IMPORTANTE

EVPN no fabrica ancho de banda ni baja latencia. Lo que hace es darle al operador herramientas estándar para tratar transportes heterogéneos como ciudadanos de primera —y dejar de inventar cada solución a mano.

Un caso hipotético: el IXP virtual federado

¿Qué pasaría si varios IXPs, gestionados por organizaciones distintas, se federaran para ofrecer una **peering LAN virtual común**?

- Cada IXP mantiene su fabric, su gobernanza y su contrato con miembros.
- Los **border-leafs** de cada fabric hacen EVPN inter-AS entre sí (RFC 9136, options A/B/C).
- Una **VNI compartida** con **RT translation** en los bordes para controlar qué cruza.
- **MAC duplicate detection y anti-spoofing** negociados en el peering inter-IXP.
- Un miembro conectado a un IXP local ve a miembros de los otros IXPs como si estuvieran en la misma LAN.

LO QUE ESTO HABILITARÍA

Cobertura geográfica ampliada sin construir presencia física. Resiliencia cooperativa frente a fallos de un fabric. Política de peering negociada entre organizaciones independientes.

LO QUE HAY QUE RESOLVER

Confianza entre dominios. Quién filtra qué. Loops cross-domain. SLAs entre operadores. Gobernanza distribuida. Modelo económico del transporte inter-IXP.

Patrones de adopción que ya se ven

- **IXPs medianos** migrando a EVPN para manejar más peers sin rediseñar el core.
- **ISPs regionales** reemplazando STP y bordes saturados con fabrics EVPN/VXLAN.
- Operaciones **multivendor** conviviendo en producción —Arista, Nokia, Juniper, FRR, sin un solo proveedor dueño.
- EVPN para **servicios L2 empresariales** (E-Line, E-LAN, E-Tree) sobre el mismo fabric del peering.
- Coexistencia con redes legacy mientras se migra *peer por peer*, no de golpe.

PATRÓN COMÚN

Casi nadie migra "big bang". El patrón ganador es construir el fabric nuevo al costado, conectar los servicios menos críticos primero, y mover el resto cuando el equipo ya respira tranquilo con la operación.

LO BUENO

Hoy hay equipos con varios años de operación real en producción, no solo PoCs. La tecnología dejó de ser experimental.

MÓDULO 06

Recomendaciones prácticas

Riesgos, buenas prácticas y cómo empezar sin romper lo que ya funciona.

Riesgos y desafíos honestos

❌ LO QUE DUELE

- Hardware/firmware compatible no es universal.
- RT, IRB, MLAG, ESI: cada vendor lo implementa con matices.
- Control-plane mal diseñado = peor que el problema original.
- Complejidad inicial superior a VLANs simples.
- Formación del equipo: BGP no es opcional —es central.
- Percepción de que hay que rediseñar accounting, mirror y compliance.

✅ LO QUE MITIGA

- Validar la matriz de soporte antes de comprar.
- Lab con Containerlab antes de tocar producción.
- Diseñar el control-plane *primero*, el resto después.
- Empezar con un solo VNI y un solo tenant.
- Capacitar al equipo en BGP profundo —es la inversión clave.
- Tu stack operativo se preserva: sFlow/IPFIX en el edge, port mirror, ERSPAN-over-IP, taps. Counters per-VNI más limpios que per-VLAN.

Buenas prácticas

LEAF-SPINE SIEMPRE QUE SE PUEDA

Topología CLOS predecible, sin loops L2, con paths iguales y balanceo natural por ECMP.

UNDERLAY SIMPLE Y ABURRIDO

eBGP unnumbered (IPv6 link-local), sin direccionamiento p2p que mantener, sin features exóticas. El underlay no es donde innovás.

EVPN COMO ÚNICA FUENTE DE VERDAD

MAC e IP se aprenden en el control-plane, no por flooding. Si no está en EVPN, no existe.

DOCUMENTAR VNIS Y RT EN NETBOX

Un VNI sin documentar es un incidente esperando. Source of truth o nada.

AUTOMATIZACIÓN DESDE EL DÍA CERO

Ansible, Nornir, Terraform. Aunque sean 4 switches: si lo hacés a mano, lo vas a hacer mal a escala.

TELEMETRÍA DESDE EL PRINCIPIO

gNMI, BMP, sFlow/IPFIX. La red moderna habla; aprovechalo.

Endurecer el underlay y BGP

leaf1 · FRR 10.2.1

```
# todo esto corre en el lab
 neighbor FABRIC bfd profile FABRIC
 neighbor FABRIC ttl-security hops 1
 neighbor FABRIC maximum-prefix 5000
 no bgp default ipv4-unicast
leaf1# show bfd peers brief
fe80::..feab:fa2f .. up (150ms x3)
# pendiente sumar: TCP-AO y CoPP
```

DETECCIÓN DE FALLOS

BFD (ya activo) desacopla el fallo de los timers de BGP. Bajar intervalos + echo mode.

PLANO DE CONTROL

GTSM/TTL-security (RFC 5082): exige TTL=255, descarta spoofing remoto. **CoPP** rate-limita al CPU.

CONTENER EL BLAST RADIUS

maximum-prefix por sesion. **Filtros**: el underlay solo transporta loopbacks /32 de los VTEP.

AUTENTICAR LA SESION

TCP-AO (RFC 5925), MD5 como piso. En unnumbered el riesgo baja, pero se documenta igual.

EVPN NO ENDURECE EL UNDERLAY SOLO

El fabric da un control plane limpio; GTSM, prefix-limits, BFD y CoPP son trabajo de hardening aparte. El lab ya trae GTSM, prefix-limits y BFD.

Recomendaciones específicas para IXPs

- **VNI por peer** o por VLAN pública —depende del modelo de IXP.
- **Route-servers**: integrarlos como leafs o como miembros del fabric, no fuera de él.
- **Minimizar flooding** en el dominio EVPN (proxy-ARP/ND, suppression).
- **Troubleshooting** usando route-types (Type-1 a Type-5) como mapa mental.
- Considerar **telemetría** con sFlow/IPFIX por peer para visibilidad real.

Checklist mínimo para IXP

- [x] Underlay con eBGP unnumbered + BFD
- [x] BGP EVPN con RR redundante
- [x] VNI documentado por peer/servicio
- [ ] ESI configurado para peers multihomed
- [ ] ARP/ND suppression habilitado
- [!] Plan de migración peer-by-peer

Recomendaciones específicas para ISPs

- **Migrar progresivamente** desde STP. Nunca de golpe.
- Aprovechar **multihoming activo-activo** desde el primer leaf.
- Integrar con herramientas de **automatización** existentes (NetBox, Ansible, Nornir).
- **Pruebas previas** en Containerlab con la misma topología que producción.
- Diseñar pensando en **IPv6 desde el inicio**, no como afterthought.

TRAMPA CLÁSICA DEL ISP

"Probamos en lab, funcionó, vamos a producción". El paso intermedio es *shadow traffic*: meter el fabric en paralelo con tráfico real espejado antes de cortar el legacy.

CAPACITACIÓN PRIMERO

Si el equipo no maneja BGP profundamente, EVPN/VXLAN se les va a sentir como magia negra. Inviertan en capacitación antes que en hardware.

¿Cómo empezar?

Semana 1

Laboratorio con Containerlab

Topología spine-leaf mínima (2 spines + 2 leafs) con FRR. En un par de horas tenés EVPN/VXLAN corriendo. El lab de esta charla: `enrut.ar/e4tTg`

Mes 1-2

Diseño, capacitación e interop

Definir VNIs/RT/numbering, formar al equipo en BGP/EVPN, y validar interop multivendor si aplica.

Mes 2-4

Primer servicio en paralelo

Un VNI, un tenant end-to-end. Shadow traffic junto al legacy. Medir convergencia y validar failover.

Mes 6-12+

Migración gradual

Servicio por servicio, con ventanas de cambio. El legacy convive hasta que cada pieza esté probada. Sin Big Bang.

Glosario rápido (1/2) · fundamentos y data plane

VTEP
VXLAN Tunnel Endpoint. El nodo que encapsula/desencapsula VXLAN.
Ej: cada `leaf` del lab es un VTEP, identificado por su loopback (10.0.0.11).

VNI
VXLAN Network Identifier. El segmento; 24 bits, ~16M valores.
Ej: VNI 10100 = la VLAN de peering del tenant-A en todo el fabric.

L3VNI
VNI asociado a un VRF para ruteo inter-subred (no L2).
Ej: VNI 5000 rutea entre las subredes del tenant-A (Type-5).

EVI
EVPN Instance. La instancia de servicio EVPN, suele mapear 1:1 con un VNI.
Ej: el EVI del VNI 10100 agrupa todas sus rutas Type-2.

ESI
Ethernet Segment Identifier. Marca el enlace/LAG de un cliente multihomed.
Ej: `03:02:..:0c:00:00:01`, autoderivado del LACP de client-c.

MULTIHOMING
Un cliente colgado de 2+ VTEPs a la vez, con redundancia activa.
Ej: client-c en LAG a leaf2 y leaf3 al mismo tiempo.

DF
Designated Forwarder. El VTEP que entrega el BUM al cliente multihomed (uno solo, para no duplicar).
Ej: leaf2 gana el DF con preferencia 50000.

BUM
Broadcast, Unknown-unicast, Multicast. El tráfico que hay que inundar.
Ej: un ARP request es BUM hasta que alguien responde.

INGRESS REPLICATION
Replicar el BUM haciendo copias unicast desde el VTEP de entrada.
Ej: leaf1 copia el broadcast a leaf2 y leaf3, uno por uno.

HER
Head-End Replication. Otro nombre de ingress replication.
Ej: sin multicast en el underlay, el BUM se resuelve con HER.

Glosario rápido (2/2) · control plane y avanzados

IMET
Inclusive Multicast Ethernet Tag (Type-3). El VTEP dice 'estoy en este VNI'.
Ej: los 3 leaves emiten IMET para armar el árbol BUM del VNI.

EAD-ES
Ethernet A-D per Ethernet Segment (Type-1). Retira un ES en una sola operación.
Ej: si cae el enlace, leaf2 manda 1 EAD-ES = mass withdrawal.

MAC MOBILITY
Mecanismo para seguir una MAC que se mueve de un VTEP a otro.
Ej: una MAC migra de leaf2 a leaf3; un número de secuencia mayor gana.

FIB
Forwarding Information Base. La tabla de reenvío programada en hardware.
Ej: el Type-2 recibido se instala en la FIB de cada VTEP.

PNI
Private Network Interconnect. Interconexión privada punto a punto entre dos partes.
Ej: dos operadores con un PNI virtual sobre el fabric, sin pasar por la LAN pública.

VPWS
Virtual Private Wire Service. Un circuito L2 punto a punto (E-Line).
Ej: unir dos sedes de un cliente como si fuera un cable directo.

GTSM
Generalized TTL Security Mechanism (RFC 5082). Exige TTL=255 en BGP.
Ej: el underlay es 1 salto: un paquete con TTL menor viene spoofeado, se descarta.

GNMI
gRPC Network Management Interface. Telemetría/config por streaming.
Ej: exportar contadores por VNI en vivo a Prometheus en vez de SNMP polling.

TCP-AO
TCP Authentication Option (RFC 5925). Autentica la sesión BGP.
Ej: reemplazo moderno de MD5 para proteger las sesiones del fabric.

¡Gracias!

EVPN y VXLAN no son el futuro del peering. Son el presente. La pregunta es cuándo lo adoptamos.

Email ariel@ayuda.la
Web ayuda.la
Comunidad nog.lat

EL LAB COMPLETO, ABIERTO

Cloná y reproducí todo lo que viste hoy:

enrut.ar/e4tTg