

ARISTA

なぜ今、EVPN VXLAN なのか？ -基礎から理解する 次世代ネットワークのコア技術-

アリスタネットワークスジャパン合同会社
神谷 尚秀

はじめに

- RFC8365 に準拠した EVPN/VXLAN は、現代の DC ファブリックにおける VPN サービス提供のデファクトスタンダードです。トポロジ内で、サーバやサービスの動的な配置・再配置を可能にします。最近では、大型から小型まで様々なハードウェア仕様のネットワーク機器が EVPN/VXLAN をサポートしており、通信事業者やエンタープライズ/キャンパスでの利用例も増えてきています。
- 本プログラムでは、EVPN/VXLAN の機能面を解説し、通信事業者やキャンパスでの活用事例を説明させて頂き、皆さんの今後のお仕事にお役立ていただければと思います。

全体の流れ

- EVPN / VXLAN 機能紹介
神谷 尚秀(アリスタネットワークスジャパン合同会社)
- 通信事業者の事例紹介
吉田 大介(株式会社長崎ケーブルメディア)
- キャンパスの事例紹介
池上 謙一(NTTドコモビジネス株式会社)
- 質疑応答

自己紹介

- 名前
 - 神□ 尚秀(かみたになおひで)
- 所属
 - アリスタネットワークスジャパン合同会社
システムズエンジニア
- 経歴
 - 2012-2023:
古河電気工業でルータのソフト開発、技術提案・サポートに従事
 - 2023-:
アリスタネットワークスで主にサービスプロバイダへの技術提案・サポートに従事
- その他活動
 - JANOG 運営委員
 - 「[show int インターネットの裏側解説](#)」YouTubeチャンネルのパーソナリティ

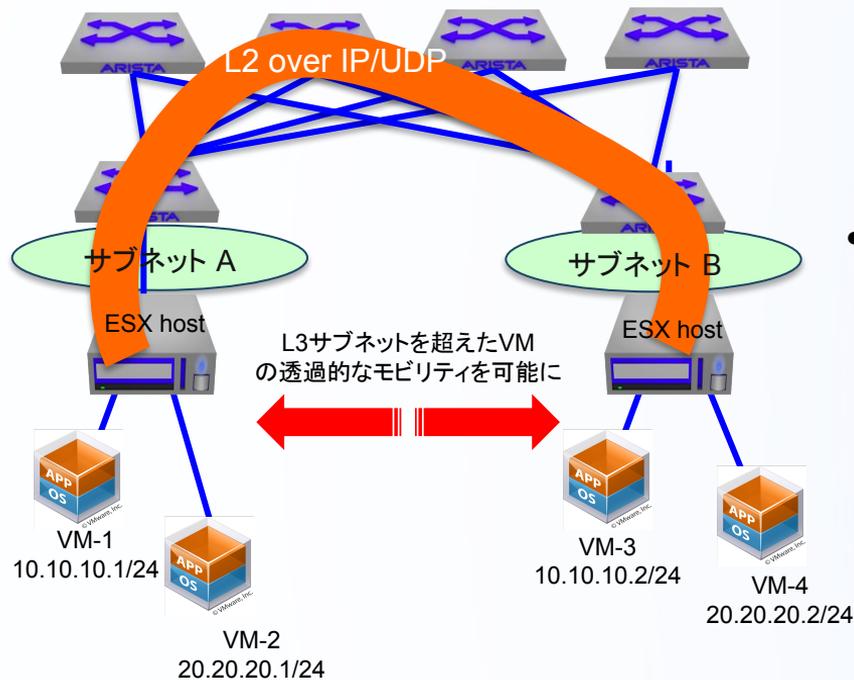


EVPN / VXLAN 機能紹介の流れ

- VXLAN 概要
- EVPN 概要
- EVPN Bridging
- EVPN Multihoming
- EVPN IRB
- まとめ

VXLAN 概要

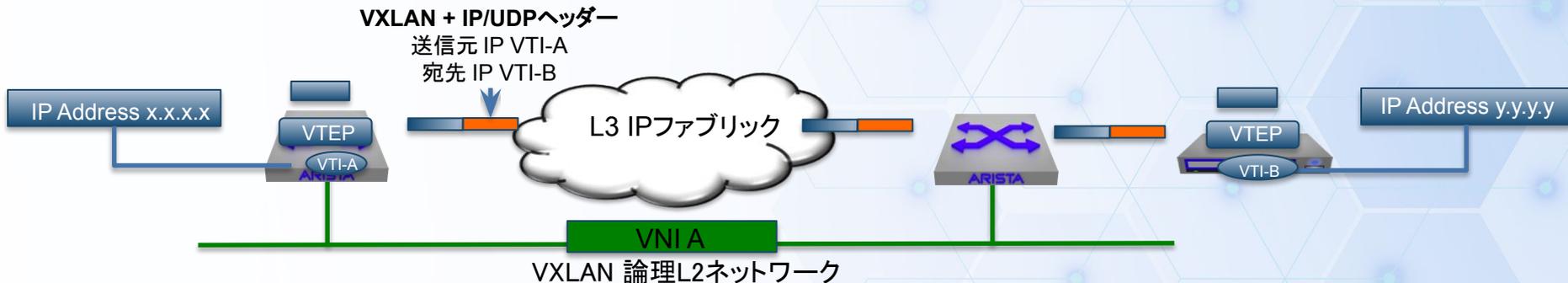
VXLAN 概要



- Virtual eXtensible LAN (VXLAN)
 - Arista/Vmware社などが中心となり仕様を策定
 - IETF RFC7348で広く認識された業界標準のVXLAN
- L3上で論理的なL2接続を実現
 - » オーバーレイ接続によって物理的なIPファブリックに対して透過的
 - » L3 IPファブリック上でL2の論理的な仮想接続を構成

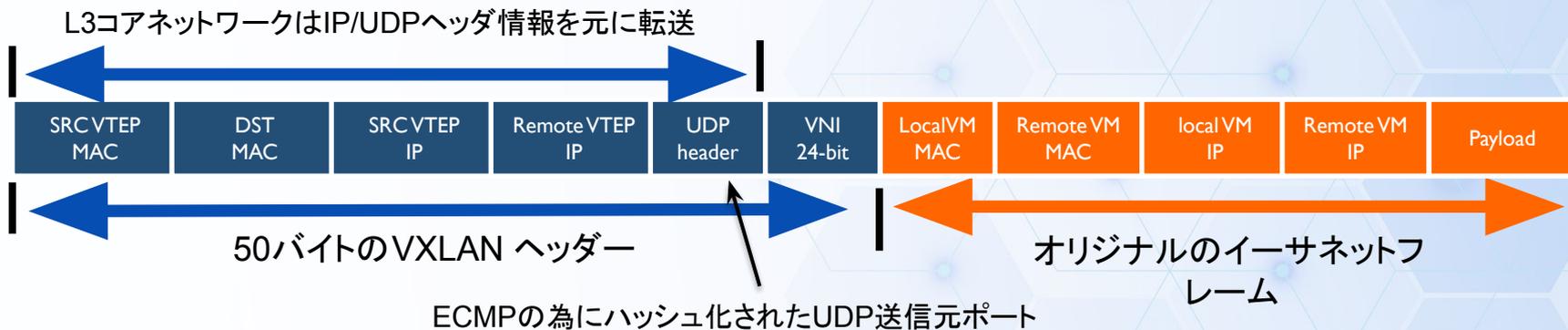
VXLAN 関連用語

- Virtual Tunnel End-Point (VTEP)
 - ホストをVXLANオーバーレイネットワークに接続する為の接続ポイント
 - VXLANヘッダーによるトンネリングの為のカプセル化とカプセル化解除を行う
- Virtual Tunnel Identifier (VTI)
 - カプセル化されたVXLANTラフィックのIPアドレス
- Virtual Network Identifier (VNI)
 - VXLANヘッダー内にある24ビットのフィールド
 - イーサネットフレームのレイヤ2セグメントを識別
- VXLAN ヘッダー
 - VTEPにてIP/UDP、そしてVXLANヘッダーが付与
 - ECMPのパス分散の為、送信元UDPポートは内部イーサネットフレームに基いたハッシュ

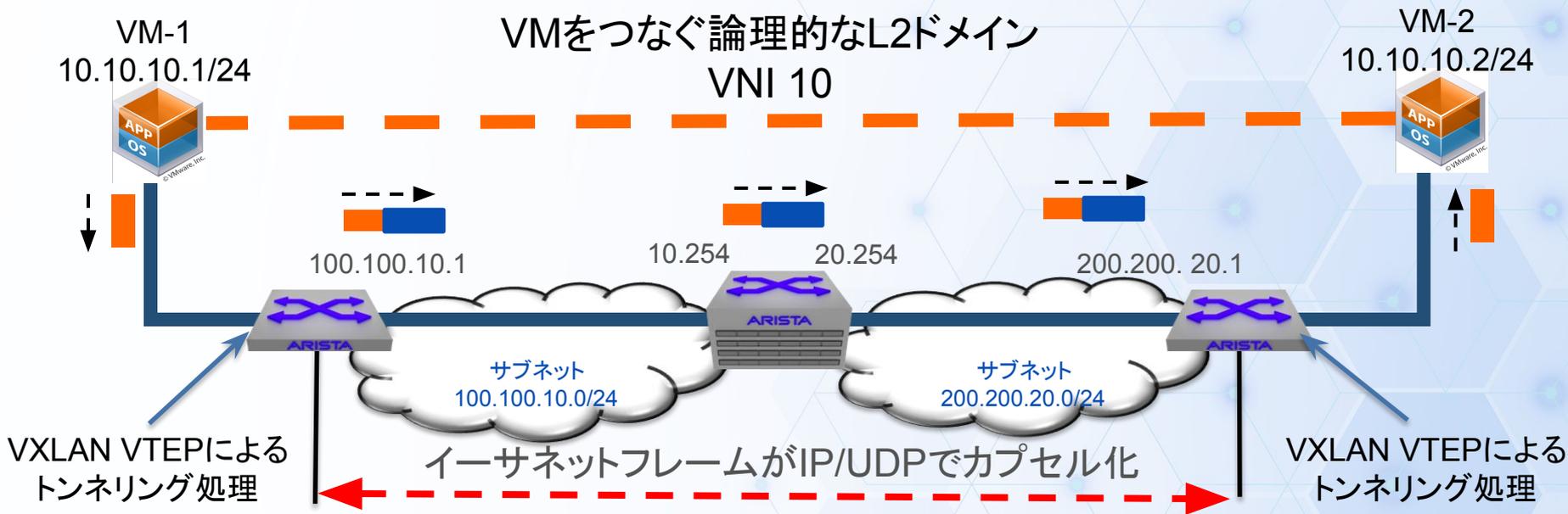


VXLAN フレームフォーマット

- イーサネットヘッダーは、ローカルVTEPのMACアドレスと宛先VTEPへの経路上のL3中間ノードのMACアドレス (14バイト+4バイト (802.1q)のオプション)
- IPヘッダーは送信元及び宛先VTEPのアドレス (20バイト)
- UDPヘッダーの送信元ポートはオリジナルのイーサネットヘッダーに基づいたハッシュ値、宛先ポートはIANA定義値 (8バイト) *VXLAN非対応のコアネットワークにおいてもECMPによるロードバランスを可能とする設計
- VXLANヘッダーは24ビットのVNIIによって1600万超のL2ドメインを実現 (8バイト)

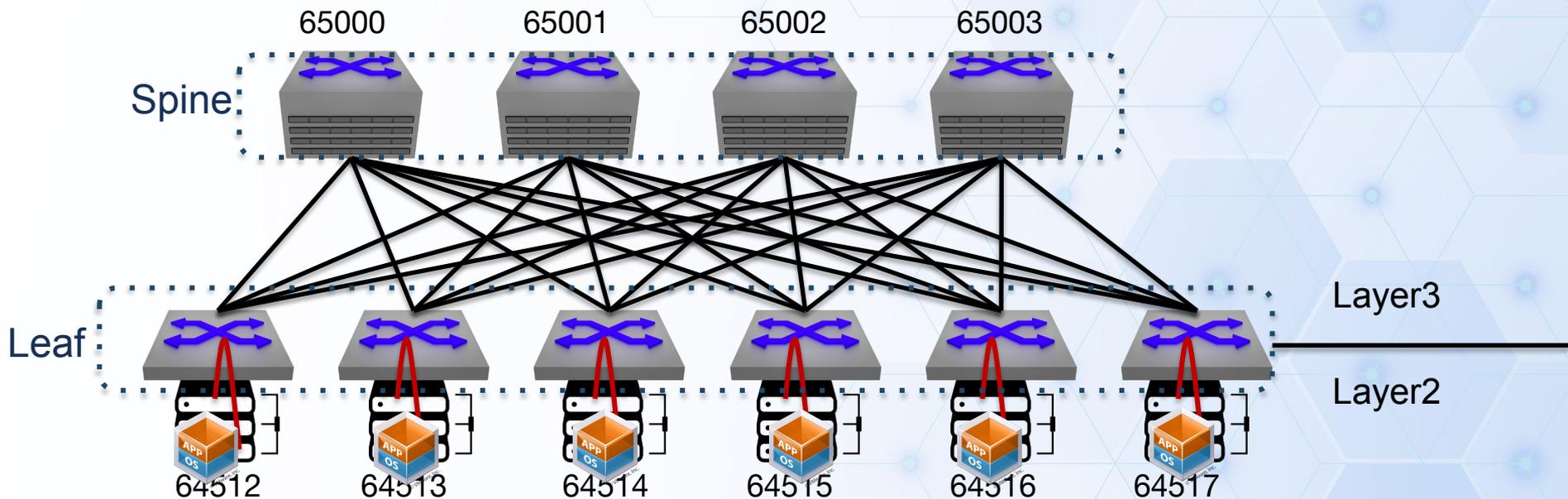


VXLAN オペレーション



VTEPにてカプセル化とカプセル化解除を行うことで
既存のIPネットワーク上を透過的に転送可能

VXLAN + IP Closデザイン



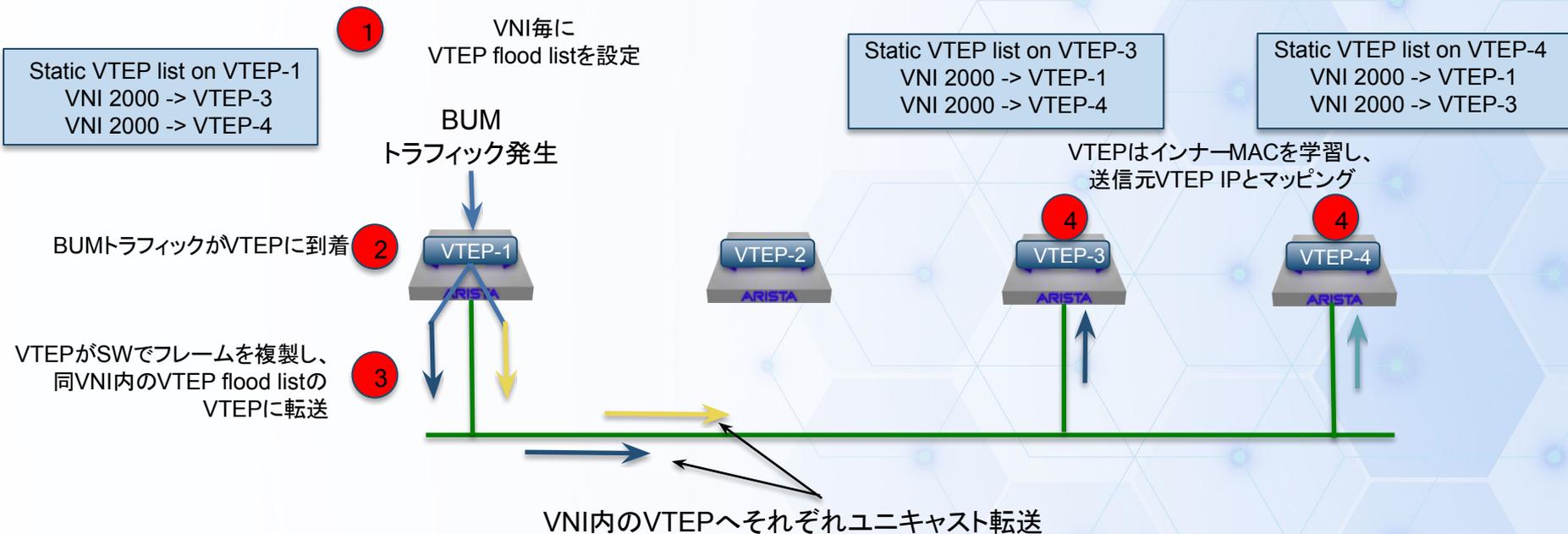
- データセンターデザインで完全なソリューション
- 標準で安定したBGPプロトコル/広く展開されたVXLANフォーマットを使用
- 特別な要求が中間ノードに存在しない

BUMトラフィックに関する課題

- VXLANはVTEPグループ間にて論理的なL2セグメントを実現
- VNIがL2ドメインとして機能するため、BUMトラフィックの処理が必要
 - B: ブロードキャストトラフィック転送
 - U: 未学習ユニキャストのMACアドレス学習
 - M: マルチキャストトラフィック転送
- 初期の実装では、BUMトラフィック処理にマルチキャストを活用しており、抵抗が大きい
- RFC7348では、マルチキャストによるデータプレーン学習だけでなく、中央権威サーバーまたは各VTEPにおけるディレクショナルリーベースの検索方法もオプションとして記載がある

Ingress Replication による BUM 処理

- 送信元VTEPが設定された「flood list」に基づき、パケットを個別に複製してユニキャスト転送することで、マルチキャスト設定を不要にする
- 未学習のMAC宛て通信はこの仕組み (Ingress Replication) を用いて全VTEPへ配送され、VXLANパケットを受信したVTEPはカプセル化解除時に送信元 MACとVTEP IPの紐づけを学習



VXLAN の設定例

- VLAN-VNIマッピングと flood-list を設定

```
vlan 10
!
interface ethernet 1
  switchport trunk allowed vlan 10
  trunk link
  switchport mode trunk
!
interface Loopback0
  description router-id
  ip address 1.1.1.15/32
!
interface Loopback1
  description VXLAN-interface
  ip address 2.2.1.15/32
!
interface Vxlan1
  vxlan source-interface Loopback1
  vxlan udp-port 4789
  vxlan vlan 10 vni 1010
  vxlan flood vtep 2.2.1.16

router bgp 65101
  router-id 1.1.1.15
  network 2.2.1.16/32
```

←ソースになるインターフェースを設定

←VLAN-ID と VNI のマッピングを設定

←BUMトラフィックの送信先を設定

EVPN 概要

VXLANコントロールプレーンの選択

- VXLANコントロールプレーンはMAC学習とフラッディングに使用される
 - リモートVTEPの背後にいるホストを発見するメカニズム
 - どのようにVTEPとVNIのメンバーを発見するのか？
 - レイヤー2セグメント(VNI)内で BUM を転送するためのメカニズム

IPマルチキャストコントロールプレーン

- VTEPはVNIの所属するマルチキャストグループにjoin
- VNI内のUnknownユニキャストはVTEPにマルチキャストで転送
 - サードパーティVTEPをサポート
- 転送と学習はIPマルチキャストを要求-限られた展開例

Ingress Replication (IR)

- BUMトラフィックはVNIの中のリモートVTEPに複製される
 - ingress VTEPで複製される
- サードパーティのVTEPをサポート
- MAC学習は転送され学習されるしかし、マルチキャストは必要が無い

HER with CloudVision eXchange (CVX)

- ローカルに学習された MACとVNIバインドはCVXに公開される
- CVXはダイナミックにリモート VTEPにステートを分配する
 - サードパーティ VTEPをサポート
- MAC学習とfloodlistの設定は自動にされる
- 信頼性の為にHAクラスタサポート

EVPN Model

- VTEP間のローカルに学習されたMACとIPの紐付けの為にBGPを使用
- ブロードキャストトラフィックはIPマルチキャストやHERによりハンドルのされる
- 設定されたBGPによりMACアドレスとVNIはダイナミックに分配される
- サードパーティVTEPをサポート

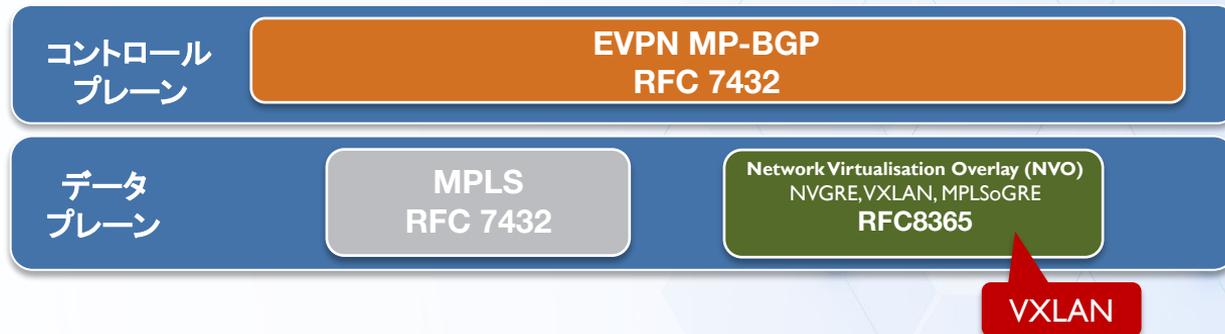
Ethernet VPN (EVPN) とは？

- **EVPN は RFC 7432で定義される**

- 同RFC中では、MPLSがデータプレーンとして定義
- BGPにて新たなアドレスファミリーを設け、MAC/IPとIPプレフィックスを広報可能に
- 多くの業界プレーヤーによって策定されている- ALU, Cisco, Juniper, AT&T, Bloomberg, Verizon など

- **データプレーンの選択肢は MPLS, NVO (RFC8365), etc**

- コントロールプレーンは共有し、ネットワーク仮想化プロトコル等をデータプレーンとして採用するべく2つのRFCが追加で定義される

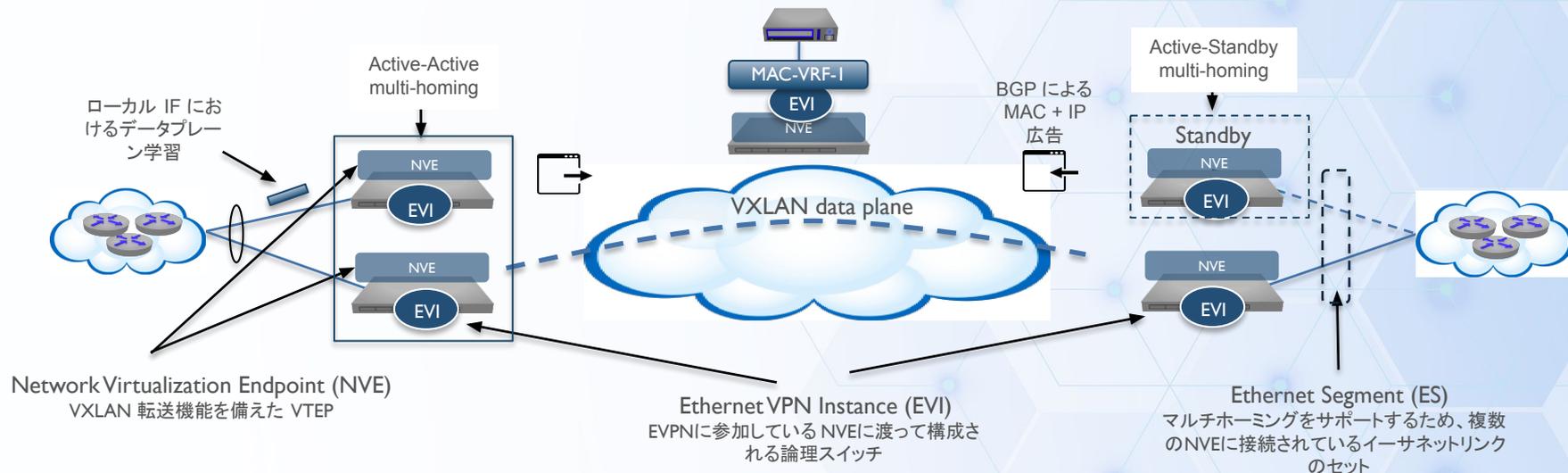


EVPN 動作概要

- **EVPNは Multi Protocol BGP (MP-BGP) を基礎とする**
 - 新たな EVPN address familyを導入
 - ≫ Address Family Identifier [AFI] 25 (Layer 2 VPN) subsequent AFI [SAFI] 70 (EVPN)
 - ≫ ホストのMAC/IPバインディングとIPプレフィックスを広報
 - ≫ Layer 2/3 情報の配布により、VXLANのブリッジングとルーティングを統合的に行える
 - **Layer 3 VPNのコンセプトである Route-Distinguisher と Route Target (RT) を用いる**
 - ≫ これによりマルチテナントのVXLANオーバーレイが可能に
 - ≫ テナント間のIPアドレスの重複にも対応可
 - **複数のテナントの NLRI が共用のBGPセッションで搬送可能**
 - ≫ テナント毎にBGPのセッションを張る必要無し

EVPN に関する用語一覧

- Network Virtualization End-point (NVE): VXLANのVTEPと同義
- EVPN Instance (EVI): EVPNに参加しているNVEに渡って構成される論理スイッチであり、テナントのホスト/サブネットを相互接続
- 冗長モード: マルチホーミングによる active-active あるいは active-standby
- MAC-VRF: Route Target (RT) および Route Distinguisher (RD) 値を持つVRFインスタンスであり、Service Interface 設定に基づき、MAC-VRFは単一あるいは複数の固有なMAC転送テーブルを保持
- Ethernet Segment Identifier (ESI): マルチホーミングの際に共有セグメントを識別するID

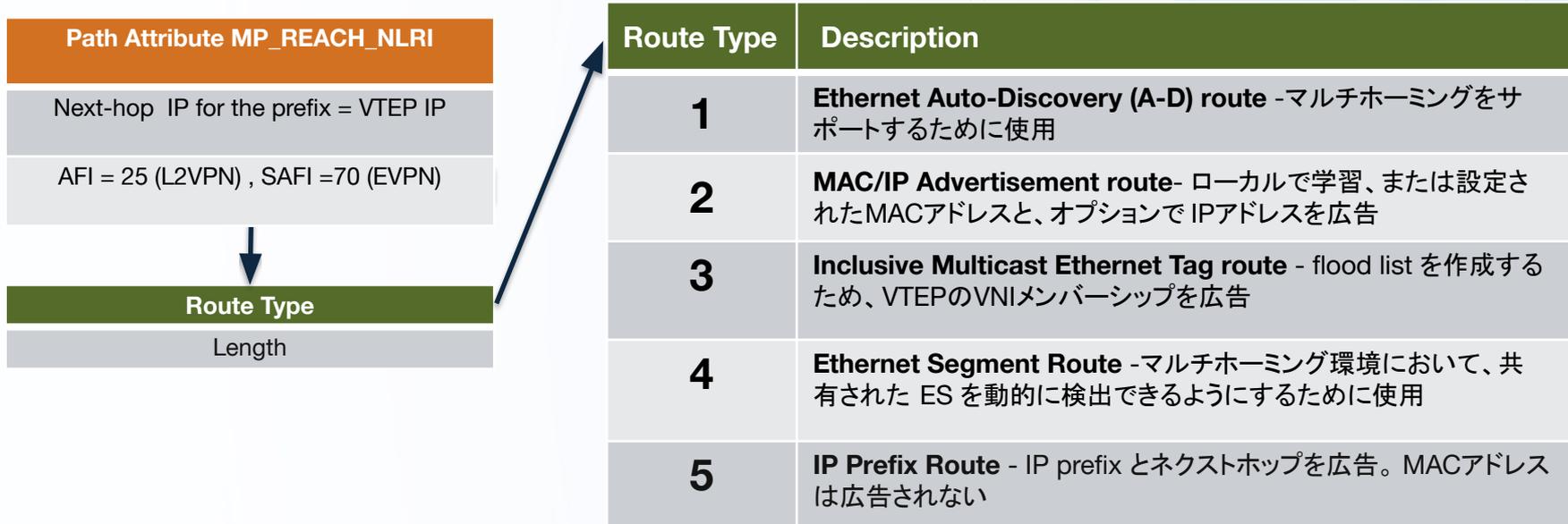


EVPNの肝となる 5つの Route Type

- 新たな NLRI で定義された 5つの Route type

- 全てのRoute typeを必ずしもサポートする必要は無い。*ベンダーの実装に任される

- Next hop (VTEP IPアドレス) は MP_REACH_NLRI パス属性で表される



EVPN でできること

- 単一の BGP による L2VPN
- マルチホーミングによる柔軟な冗長化
- 高速なコンバージェンス
- エッジ装置でのルーティング・ブリッジングを実現する IRB (Integrated Routing and Bridging)
- 単一の BGP による L3VPN
- IP-VPN (RFC4364)メカニズムを活用した RT (Route Target) によるルート制御
- ARP proxy による ARP 抑制

などなど

EVPN Bridging

※発表では、service interface は vlan-based、BUM のトランスポートは Ingress replication について解説
vlan-aware bundle や Multicast Underlay については Appendix を参照

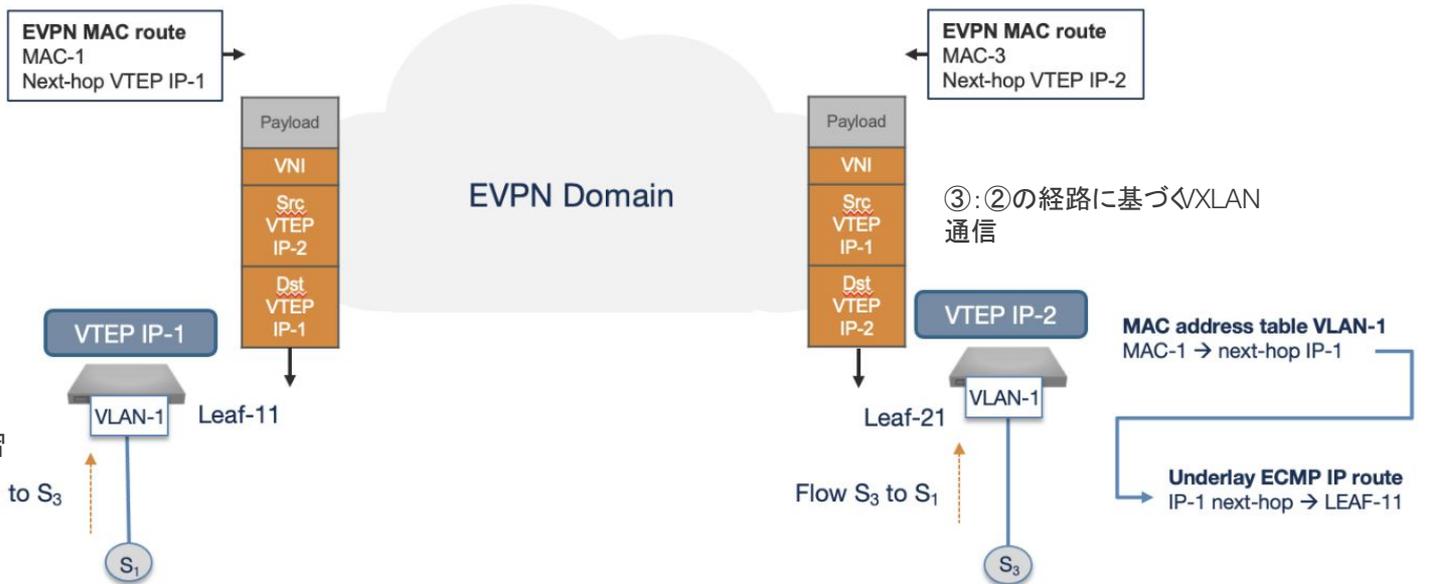
EVPN Bridging 概要

- 異なるVTEPノード上に存在するホスト間でレイヤー2接続を提供するため、EVPNは以下のプロセスを定義
 - Type-2 MAC-IP ルートによる、MACおよびMAC-IP の動的な学習と広告
 - Type-3 IMET ルートによる、BUMトラフィック処理のためのブリッジドメイン内のVTEPメンバーシップ検出

MAC address 学習

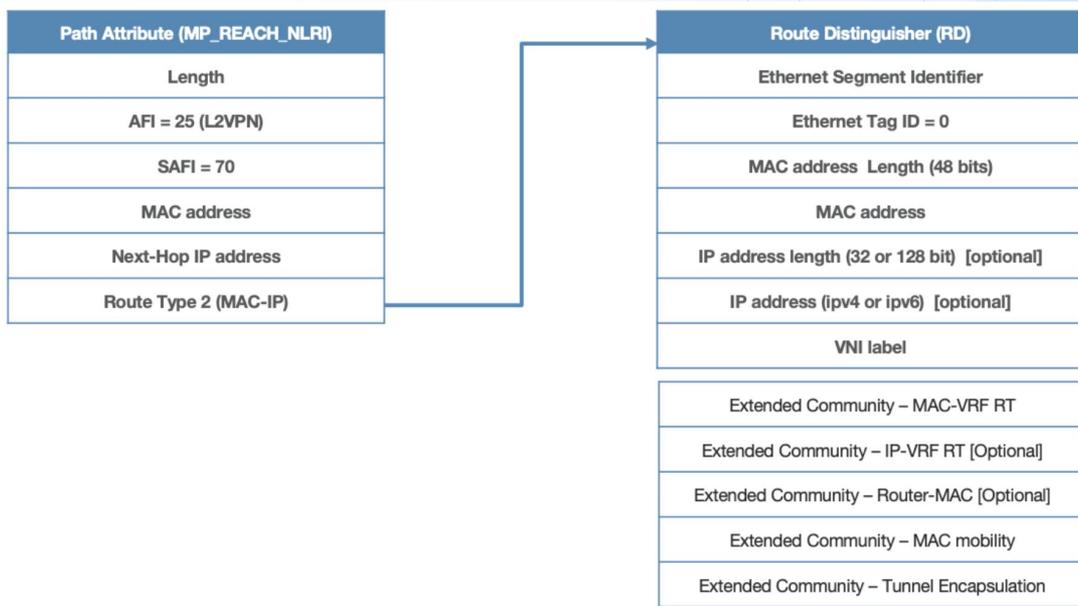
- VTEPのローカルインターフェース上でのMACアドレス学習はフローベース
 - インターフェースで受信したトラフィックに基づいてローカルMACアドレスを学習
 - インターフェースの「VLANとMAC-VRFのマッピング」に基づいて、MACアドレスがEVPN MAC-VRFに関連付け
- MACアドレスがMAC-VRFに関連付けられ、EVPNドメイン内のBGPピアに対して、Type-2 ルートのアップデートとして広告

②: ①の学習結果に基づくType-2 ルートの広告



EVPN Type-2 MAC-IP ルート フォーマット

- Type-2ルートは、MAC-VRFの RT (Route Target) と共に広告
 - 同じMAC-VRFのメンバーであるリモートVTEPは、RT import ポリシーに基づいた制御が可能
 - 経路が import されると、MACアドレスはVTEPの関連するMAC-VRF用ブリッジテーブルに追加され、ネクストホップはType-2ルートを広告したVTEPとなる
- MACアドレスに関連付けられたネクストホップVTEPへ転送される際、Type-2 ルートで広告された VNI ラベルを用いてVXLANカプセル化を行う



Type-2 ルート広告の設定例 (VLAN-based)

- VLAN-based service interface では、各VLANに対してVLAN-VNIマッピングとMAC-VRFを設定
- 以下の設定例では、VLAN 10と11に対して、VNIとMAC-VRFを割当

```
vlan 10
!
vlan 11
!
interface ethernet 1
  switchport trunk allowed vlan 10-11
  trunk link
  switchport mode trunk
!
interface Loopback0
  description router-id
  ip address 1.1.1.15/32
!
interface Loopback1
  description VXLAN-interface
  ip address 2.2.1.15/32
!
interface Vxlan1
  vxlan source-interface Loopback1
  vxlan udp-port 4789
  vxlan vlan 10 vni 1010 ←VLAN-VNI マッピング
  vxlan vlan 11 vni 1011
```

```
router bgp 65101
  router-id 1.1.1.15
  --- truncated output-----

!
vlan 10                                ←MAC-VRF の設定に対応
  rd 1.1.1.15:1010
  route-target both 1010:1010
  redistribute learned
  learnt MACs in the VLAN
!
vlan 11
  rd 1.1.1.11:1011
  route-target both 1011:1011
  redistribute learned
  learnt MACs in the VLAN
```

Type-2 ルート広告の表示例 (VLAN-based)

- ホスト1(10.10.10.101/0010.0000.0101)とホスト2(10.10.11.101/0011.0000.0101)がVTEP(DC1-Leaf11)にローカル接続されていると仮定
 - Host-1はVLAN 10のメンバー、Host-2はVLAN 11のメンバー
- 両方のMACアドレスはVTEPのローカルインターフェース(Eth1)上で学習

```
DC1-Leaf11(config-if-Po100)#show mac address-table vlan 10
Mac Address Table
```

Vlan	Mac Address	Type	Ports	Moves	Last Move
10	0010.0000.0101	DYNAMIC	Eth1	2	54 days, 21:36:13 ago

```
DC1-Leaf11(config-if-Po100)#show mac address-table vlan 11
Mac Address Table
```

Vlan	Mac Address	Type	Ports	Moves	Last Move
11	0011.0000.0101	DYNAMIC	Eth1	2	54 days, 21:36:49 ago

Type-2 ルート広告の表示例 (VLAN-based)

- VTEPのローカルインターフェースで学習された MACアドレスは、関連する MAC-VRFの RT、RD、VNI と合わせて、EVPNDメイン内に Type-2 ルートとして広告される
- 異なるVLANのメンバーである MACアドレスごとに、MAC-VRFが固有の RTとRDと共にアドバタイズされる

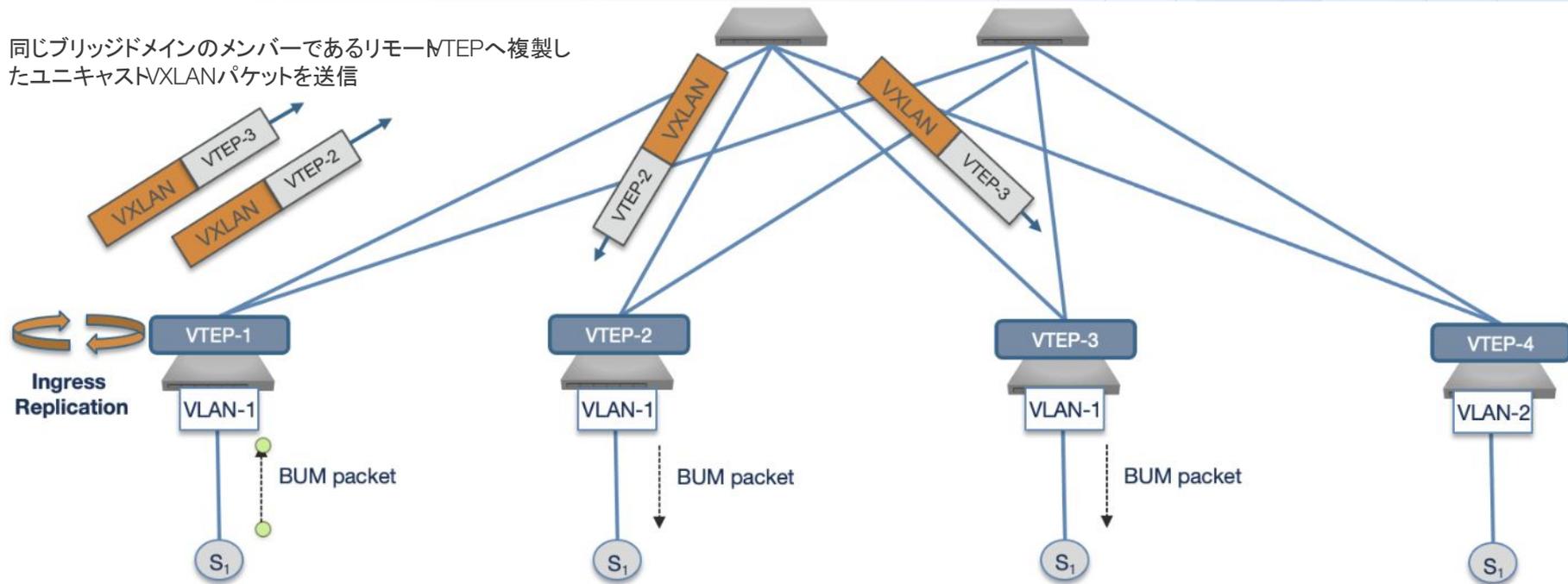
```
DC1-Spine1#show bgp evpn route-type mac-ip 0010.0000.0101 detail
BGP routing table information for VRF default
Router identifier 1.1.1.11, local AS number 65100
BGP routing table entry for mac-ip 0010.0000.0101, Route Distinguisher: 1.1.1.15:1010
Paths: 1 available
 65101
  2.2.1.15 from 1.1.1.15 (1.1.1.15)
    Origin IGP, metric -, localpref 100, weight 0, valid, external, best
    Extended Community: Route-Target-AS:1010:1010 TunnelEncap:tunnelTypeVxlan
    VNI: 1010 ESI: 0000:0000:0000:0000:0000
```

```
DC1-Spine1#show bgp evpn route-type mac-ip 0011.0000.0101 detail
BGP routing table information for VRF default
Router identifier 1.1.1.11, local AS number 65100
BGP routing table entry for mac-ip 0011.0000.0101, Route Distinguisher: 1.1.1.15:1011
Paths: 1 available
 65101
  2.2.1.15 from 1.1.1.15 (1.1.1.15)
    Origin IGP, metric -, localpref 100, weight 0, valid, external, best
    Extended Community: Route-Target-AS:1011:1011 TunnelEncap:tunnelTypeVxlan
    VNI: 1011 ESI: 0000:0000:0000:0000:0000
```

BUM traffic 転送処理

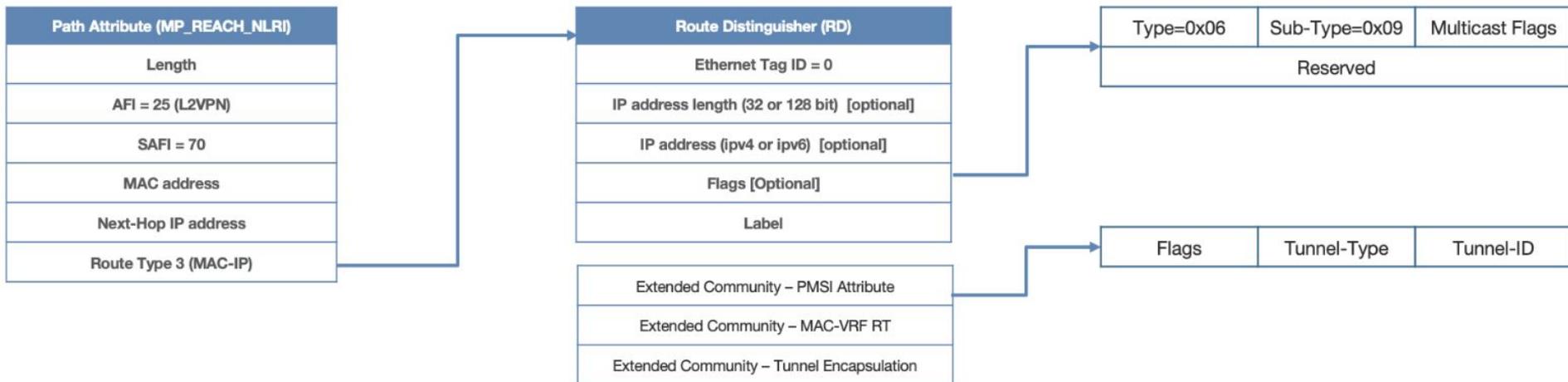
- Ingress Replication (IR) はオーバーレイBUMトラフィックを転送する方法の1つ
- IR では、BUMパケットをVXLANカプセル化し、共通のブリッジメインに参加しているVTEPに対してユニキャストVXLANパケットを送信

同じブリッジメインのメンバーであるリモートVTEPへ複製したユニキャストVXLANパケットを送信



EVPN Type-3 IMET ルート フォーマット

- BUMトラフィックの転送は、Type-3 ルートの広告によって制御
- 特定のブリッジメインのメンバーである各VTEPは、そのブリッジメインに対するType-3ルートを広告
 - Type-3 ルートには、VTEPのIPアドレス、MAC-VRF に関連付けられたRT が含まれる
 - RT に基づいてType-3ルートをインポートすることで、同じブリッジメイン内のピアVTEPを検出



Ingress replication による BUM traffic 制御の設定例

- Arista EOS の BUM トラフィック転送におけるデフォルト動作モードは IR であるため、IR 用の追加設定は不要

```
interface loopback 1
  ip address 2.2.1.15
!
interface Vxlan1
  vxlan source-interface Loopback1
  vxlan udp-port 4789
  vxlan vlan 10 vni 1010
!
router bgp 65002
  router-id 1.1.1.15
!
  vlan 10
    rd 1.1.1.15:1010
    route-target both 1010:1010
    redistribute learned
```

Ingress replication による BUM traffic 制御の表示例

- VTEP は、関連するブリッジドメインに対して、Type-3ルートを広告
 - PSMIトンネルタイプ: Ingress Replication
 - ラベル: ブリッジドメインの VNI
 - Tunnel-ID: MAC-VRF 内の受信 BUM に対する VTEP のユニキャスト IP

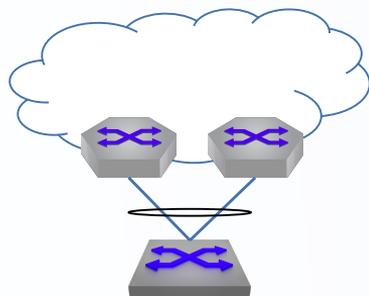
【Spine-1】

```
DC1-Spine1#show bgp evpn route-type imet detail
BGP routing table information for VRF default
BGP routing table entry for imet 2.2.1.15, Route Distinguisher: 1.1.1.15:1010
Paths: 1 available
 65101
  2.2.1.15 from 1.1.1.15 (1.1.1.15)
    Origin IGP, metric -, localpref 100, weight 0, valid, external, best
    Extended Community: Route-Target-AS:1010:1010 TunnelEncap:tunnelTypeVxlan
    VNI: 1010
    PSMI Tunnel: Ingress Replication, MPLS Label: 1010, Leaf Information Required: false, Tunnel ID: 2.2.1.15
```

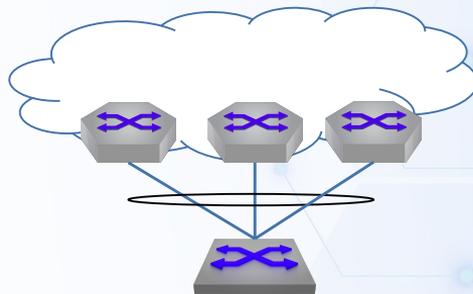
EVPN Multihoming

EVPN Multihoming 概要

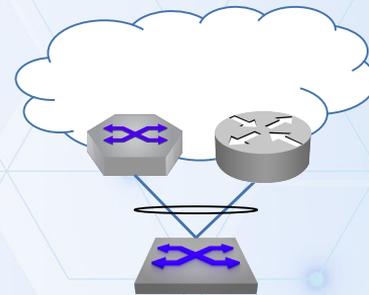
- EVPN Multihoming では、ノード間のピアリンク無しで、2台以上のノードで All-Active, Single-Active の冗長化が可能
- Ethernet Segment (ES) という概念と、2つのEVPNルートタイプである Type-1 Ethernet Auto-Discovery ルートと Type-4 Ethernet Segment ルートにより、ピアの検出、トラフィックのロードバランシング、および高速フェイルオーバーを実現可能
 - Ethernet Segment (ES): 同じダウンストリーム・ポートチャネル(LAG)に接続しているノード群は、同一の ES を共有
 - Type-1 Ethernet Auto-Discovery (EAD) ルート: ESに対するVTEPのアクティブな接続の広告、障害発生時の高速フェイルオーバーのための Mass withdraw、ESのノード間でのロードバランシングのための Aliasing に使用
 - Type-4 Ethernet Segment (ES) Route: 同じESに接続されているピア VTEPを検出し、ES上のBUMトラフィックを処理するための Designated Forwarder (DF) を選出するために使用



Active-Active
地理的制限無し



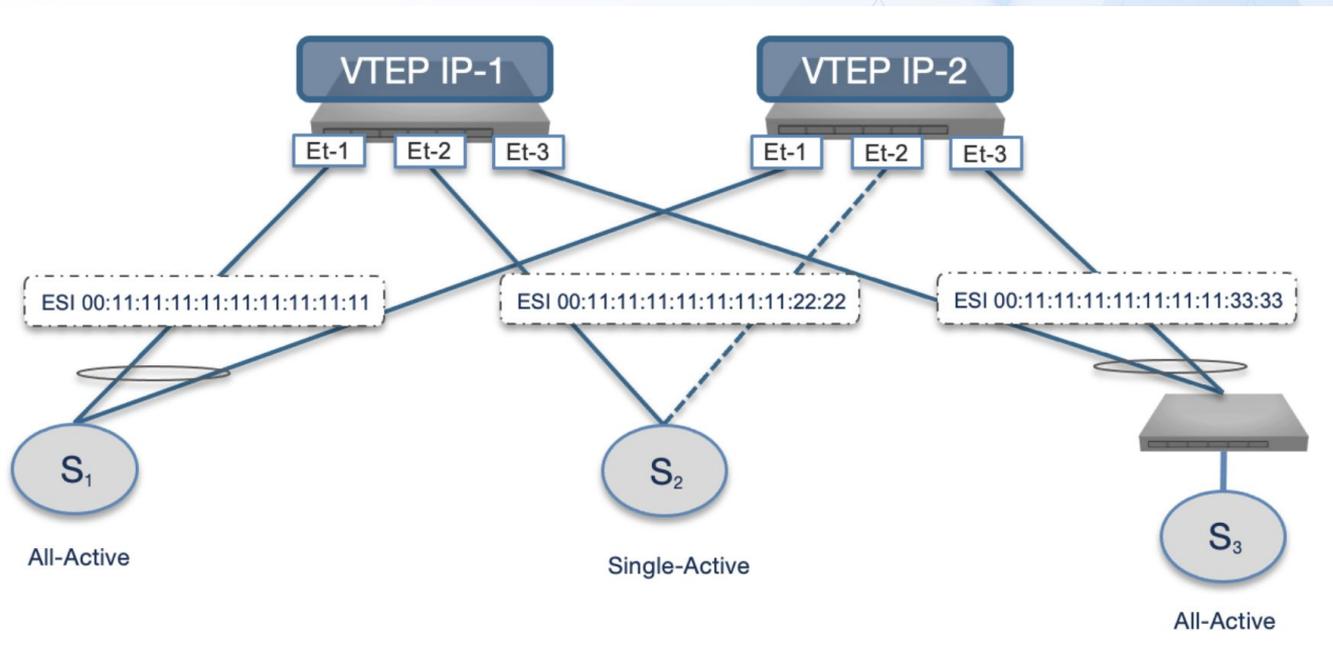
ノード3重化



マルチベンダー
サポート

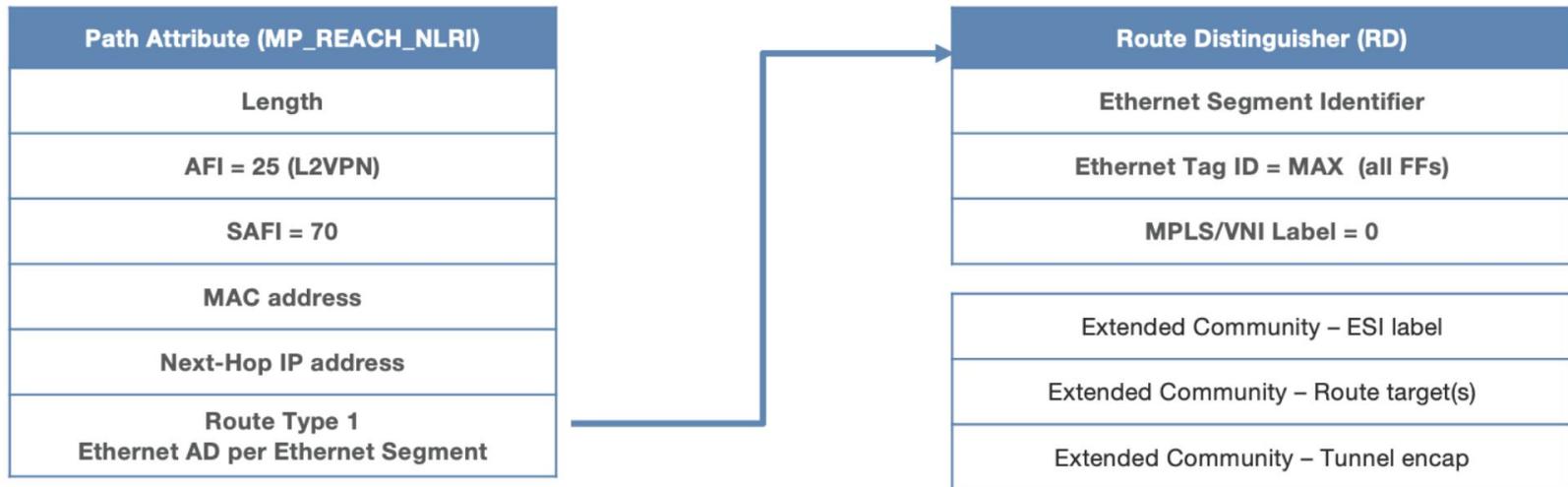
Ethernet Segment Identifier (ESI)

同じダウンストリームのサーバー/スイッチに接続する際にノード間で共有される ES は、ESI という10バイトの値を使用して識別



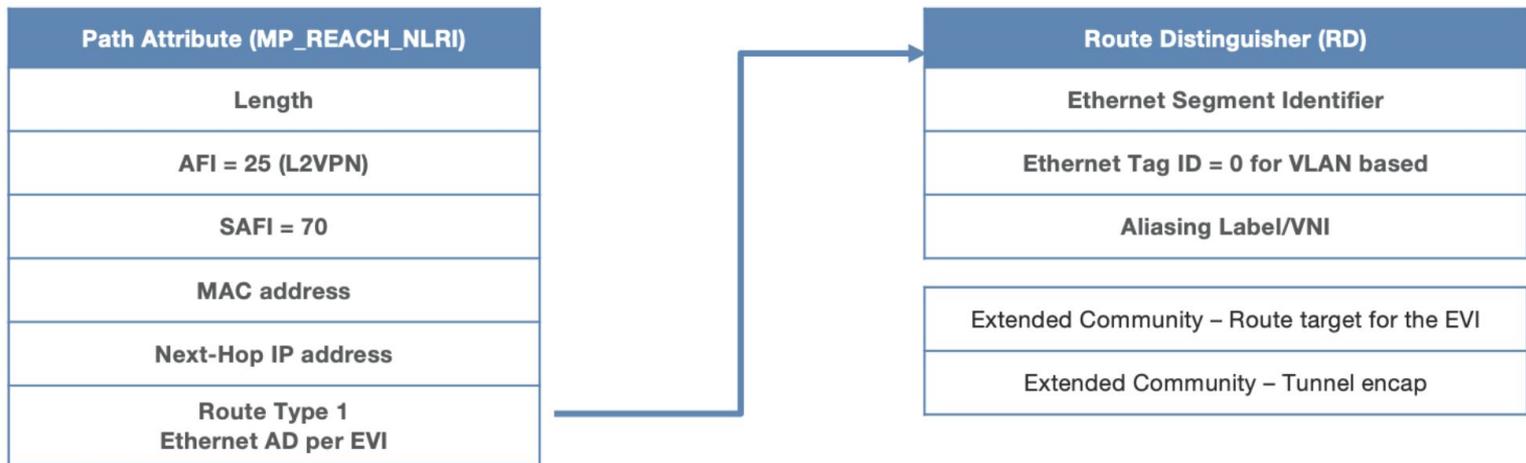
EVPN Type-1 Ethernet A-D ルート フォーマット (EAD per ES)

- VTEP は、Type-1 Ethernet A-D route を使用して、ES への接続性を広告
 - Type-1ルートには、EAD per ESと、EAD per EVI の2つのサブタイプがある
- EAD per ESルートは、**特定の ES への到達可能性を通知** するために広告
- 高速コンバージェンス (Mass Withdrawal) や、MPLS転送プレーンで展開される際の BUMトラフィック処理におけるスプリットホライズン(ループ防止)に使用



EVPN Type-1 Ethernet A-D ルート フォーマット (EAD per EVI)

- EAD per EVIルートは、**ES上の特定の EVI (MAC-VRF)への接続性を通知** するために広告
- このルートタイプはトラフィックのロードバランシング (Aliasing) に使用



EVPN All-Active ESI port-channel コンフィグ

- 下記設定例は、ESI(0000:0001:0001:0000:1112)のメンバーである2つのVTEP、Leaf-11とLeaf-12に関するもの
- このESはアクティブなポートチャネルとして設定されているため、両TEPに設定される共有LACPシステムID は、ESI値の下位6バイトから生成

```
【Leaf-11】
vlan 10
!
interface Port-Channel 100
  switchport trunk allowed vlan 10
  switchport mode trunk
!
  evpn ethernet-segment
    identifier 0000:0001:0001:0000:1112
    route-target import 00:01:00:00:11:12
    lacp system-id 0001.0000.1112
!
interface loopback 0
  description local_vtep_IP
  ip address 1.1.1.15/32
!
interface loopback 1
  description ndoe_vtep_IP
  ip address 2.2.1.15/32
!
interface vxlan1
  vxlan source-interface Loopback1
  vxlan udp-port 4789
  vxlan vlan 10 vni 1010
!
router bgp 65101
!
  vlan 10
    rd 1.1.1.15:1010
    route-target both 1010:1010
    redistribute learned
```

```
【Leaf-12】
vlan 10
!
interface Port-Channel 100
  switchport trunk allowed vlan 10
  switchport mode trunk
!
  evpn ethernet-segment
    identifier 0000:0001:0001:0000:1112
    route-target import 00:01:00:00:11:12
    lacp system-id 0001.0000.1112
!
interface loopback 0
  description local_vtep_IP
  ip address 1.1.1.16/32
!
interface loopback 1
  description ndoe_vtep_IP
  ip address 2.2.1.16/32
!
interface vxlan1
  vxlan source-interface Loopback1
  vxlan udp-port 4789
  vxlan vlan 10 vni 1010
!
router bgp 65102
!
  vlan 10
    rd 1.1.1.16:1010
    route-target both 1010:1010
    redistribute learned
```

Multihoming における Type-1, -2 ルート広告の表示例

- 下記は、2台のリーフノードによって広告されるType-1 EAD per ES および EAD per EVI ルート、ならびにポートチャンネルに直接接続されているHost-1 (00bb.cccc.aaaa) に対して Leaf-11 が広告する Type-2 ルートの表示例
- MACアドレスは Leaf-11 によってローカルに学習されるため、Leaf-11 だけが非ゼロのESI を伴う Type-2 MAC-IP ルートを広告

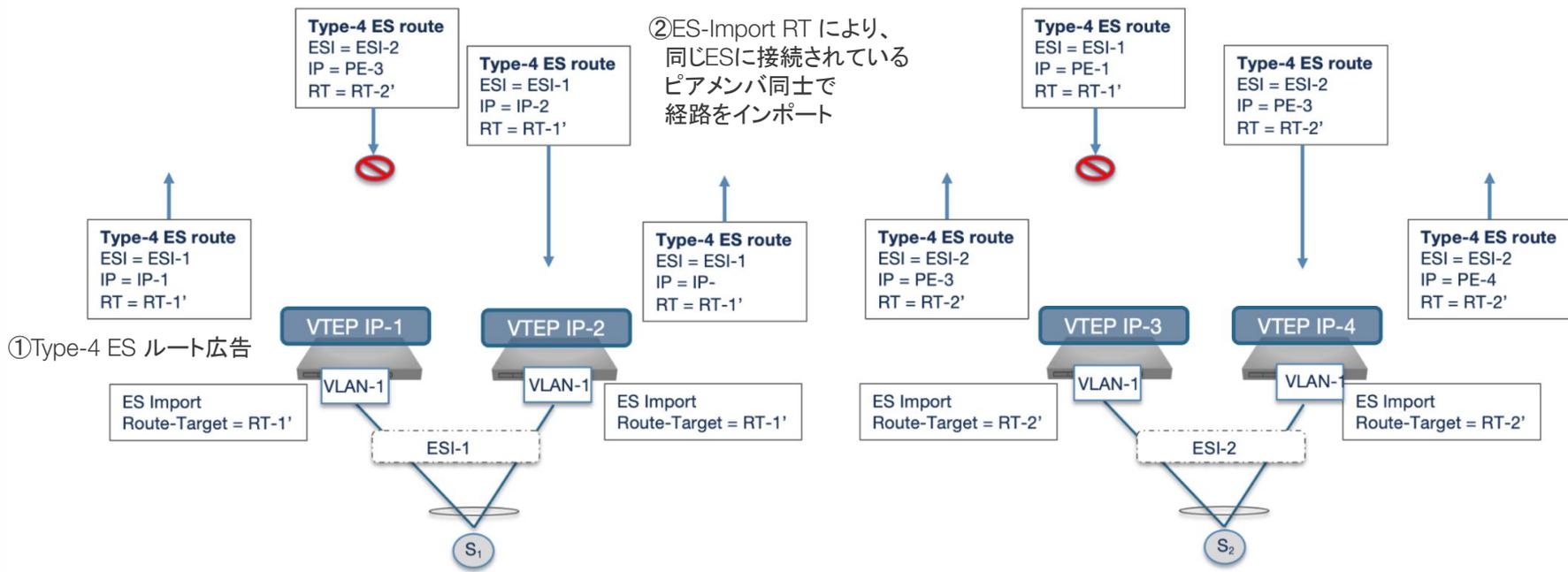
```
DC1-Leaf11(config-if-Po100)#show bgp evpn route-type auto-discovery rd 1.1.1.15:1 detail ←EAD per ES route
BGP routing table information for VRF default
Router identifier 1.1.1.15, local AS number 65101
BGP routing table entry for auto-discovery 0000:0001:0001:0000:1112, Route Distinguisher: 1.1.1.15:1
Paths: 1 available
Local
- from - (0.0.0.0)
Origin IGP, metric -, localpref -, weight 0, tag 0, valid, local, best
Extended Community: Route-Target-AS:1010:1010 Route-Target-AS:1011:1011 TunnelEncap:tunnelTypeVxlan EvpnEsiLabel:0

DC1-Leaf11(config-if-Po100)#show bgp evpn route-type auto-discovery rd 1.1.1.15:1010 detail ←EAD per EVI route
BGP routing table information for VRF default
Router identifier 1.1.1.15, local AS number 65101
BGP routing table entry for auto-discovery 0 0000:0001:0001:0000:1112, Route Distinguisher: 1.1.1.15:1010
Paths: 1 available
Local
- from - (0.0.0.0)
Origin IGP, metric -, localpref -, weight 0, tag 0, valid, local, best
Extended Community: Route-Target-AS:1010:1010 TunnelEncap:tunnelTypeVxlan
VNI: 1010

DC1-Leaf11(config-if-Po100)#show bgp evpn route-type mac-ip 00bb.cccc.aaaa detail
BGP routing table information for VRF default
Router identifier 1.1.1.15, local AS number 65101
BGP routing table entry for mac-ip 00bb.cccc.aaaa, Route Distinguisher: 1.1.1.15:1010
Paths: 1 available
Local
2.2.1.15 from - (0.0.0.0)
Origin IGP, metric -, localpref -, weight 0, tag 0, valid, local, best
Extended Community: Route-Target-AS:1010:1010 TunnelEncap:tunnelTypeVxlan
VNI: 1010 ESI: 0000:0001:0001:0000:1112
```

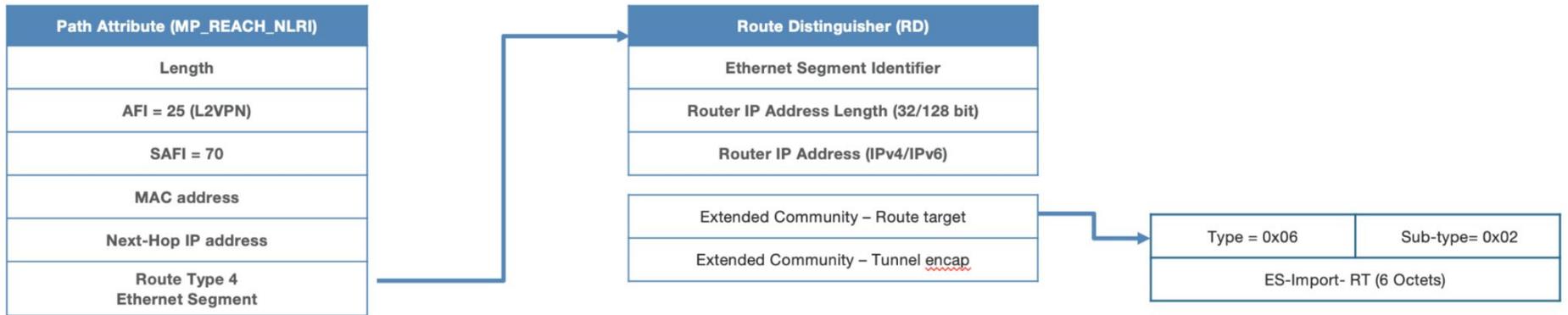
EVPN Type-4 Ethernet Segment ルート

- EVPN Multihoming に参加するVTEPは、Type-4 ES ルートを介して、ES への自身の接続性を広告し、ピアメンバーVTEPを検出
- Type-4 ES ルートは、そのESの識別子 (ESI) から導出される固有の ES-Import Route Target と共に広告され、同じESに接続されているVTEPだけがこのルートを実ポート



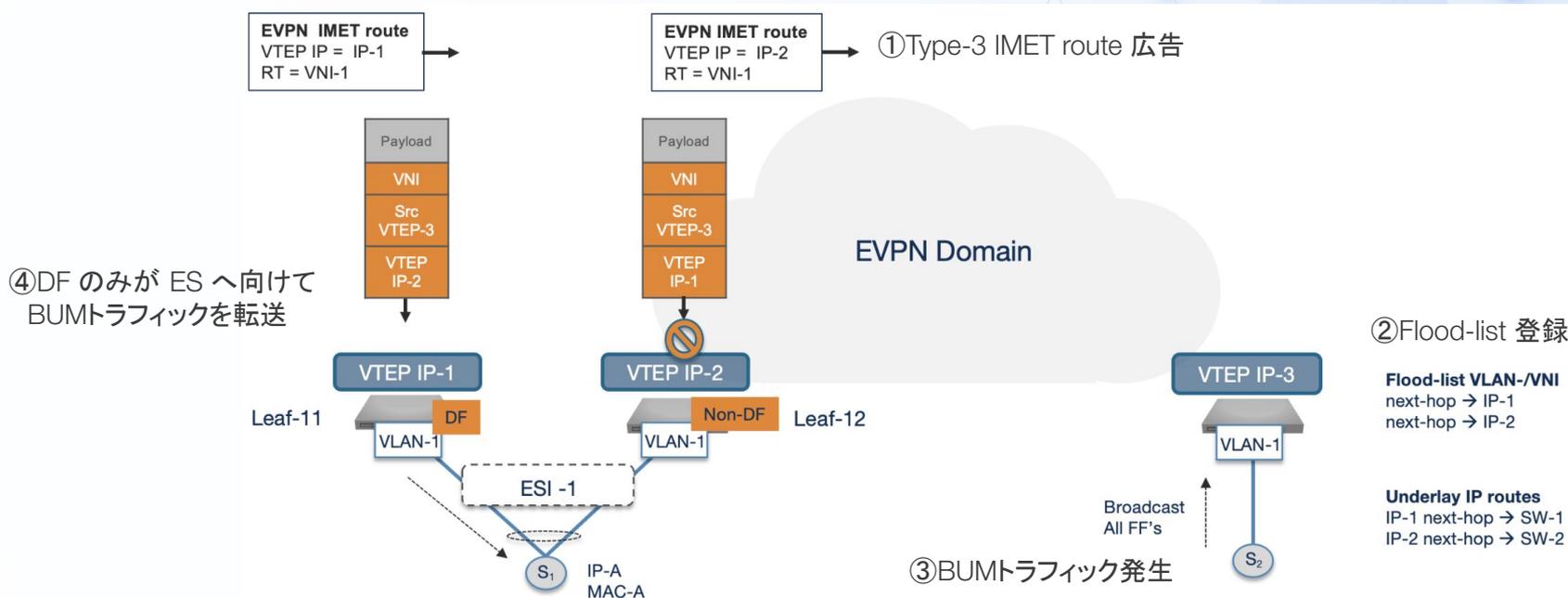
EVPN Type-4 Ethernet Segment ルート フォーマット

- ES-Import Route Target がの導出方法は、ESIタイプによって異なる
 - タイプ1 (LACP)、タイプ2 (STP MAC)、およびタイプ3 (System MAC) の場合、RTは ESI値の9オクテット部分のうち、上位6オクテットから自動的に導出
 - その他のすべてのESIタイプについては、RTを手動で設定する必要があり、設定値は ESの全ピアノード間で一致している必要がある



Designated Forwarder (DF)

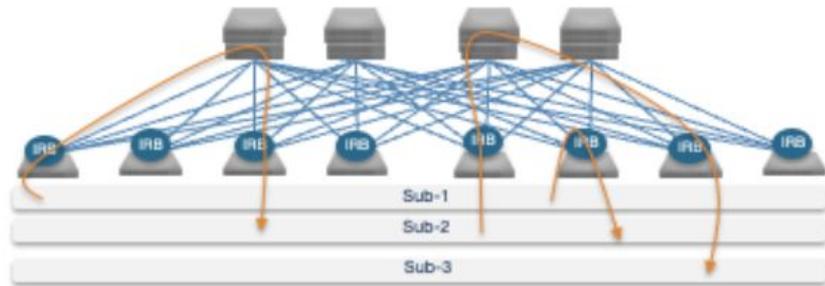
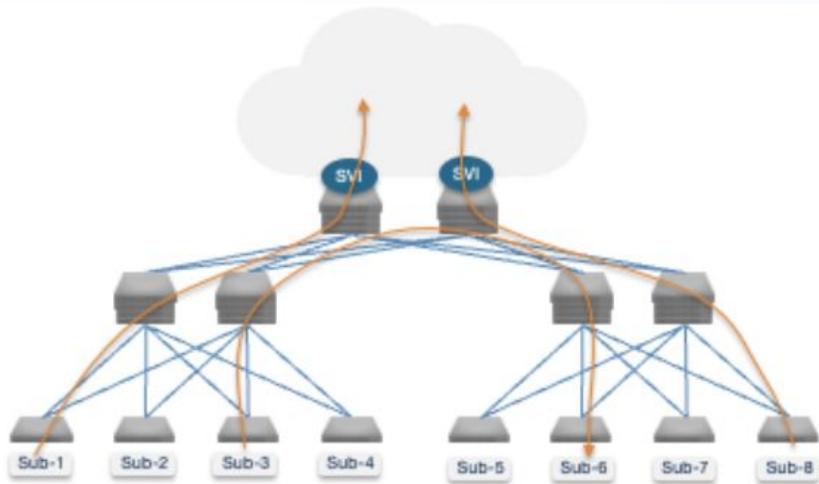
- Type-4 ESルートは、ES 上での DF の選出に使用
 - EVIに対して選出されたDFだけが、ESへ向けてBUMトラフィックを転送可能
- DFの役割が必要になる理由は、ESに接続された各 VTEPがType-3 IMETルートを広告するため
 - リモートVTEPはES上のVTEPごとにIMETルートを受信し、それに基づいてフラッドリストを作成
 - ES上のVLAN宛てのBUMトラフィックは、そのESのメンバーであるすべてのVTEPに対してフラディング
 - 選出されたDFだけがBUMトラフィックをESへ転送することで、パケットの重複中継を防ぐ



EVPN IRB

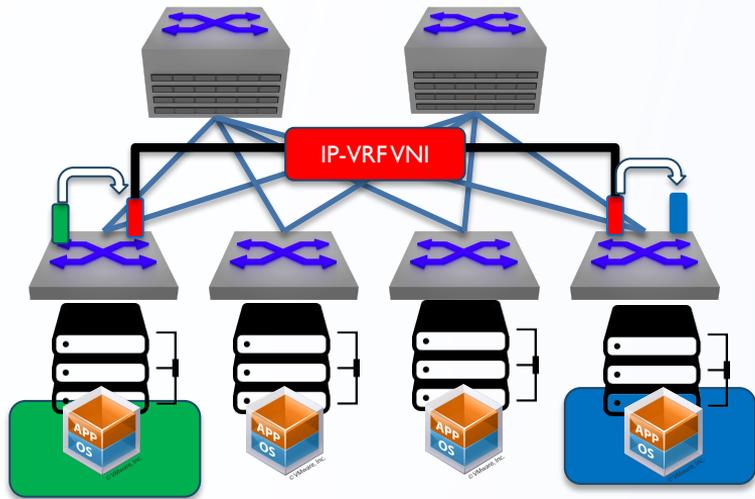
Integrated Routing and Bridging (IRB) 概要

- North-to-Southトラフィックに最適化された層トポロジーでは、サブネット間の転送は中央集権型のルーターによって提供
- East-to-Westトラフィックが大部分の仮想化サーバーインフラ向けに設計されたマルチテナンEVPNデータセンター環境では、サブネットが複数の Leaf ノードにまたがって広がっている
 - east-west でサブネットが複数リーフにまたがる環境では、中央集権型の設計は、非効率的な帯域幅使用と最適ではない転送につながる
- 最適な転送モデルを実現し、トラフィックのヘアピン現象を回避するためRFC 9135 (IRB in EVPN) では、分散ルーティングモデルを導入
 - IRB 機能がトポロジー内の各VTEPによって提供され、ルーティング操作をエンドホストの近くで行うことが可能
 - さらに、RFC 9136では Type-5 IP prefix ルートが導入され、サブネット間のルーティングのためにoverlay prefix を広報することが可能。EVPNコントロールプレーンを使用したL3 IP-VPNの提供のために、IRBと組み合わせて、あるいは単独で使用することが可能



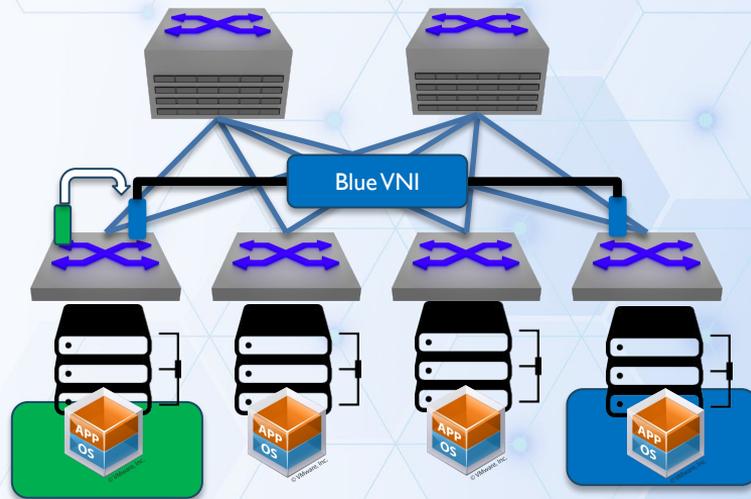
Integrated Routing and Bridging (IRB) 概要

- RFC9135では、IRB機能について Asymmetric IRB、Symmetrical IRB の2つの転送モデルを提案



Symmetric IRB

- Ingress VTEPでEgress VTEPとの共通のメンバーであるIP-VRFへルーティングされ、その後はIP-VRF介して通信される
- Egress VTEPではIP-VRFからルーティングを行う

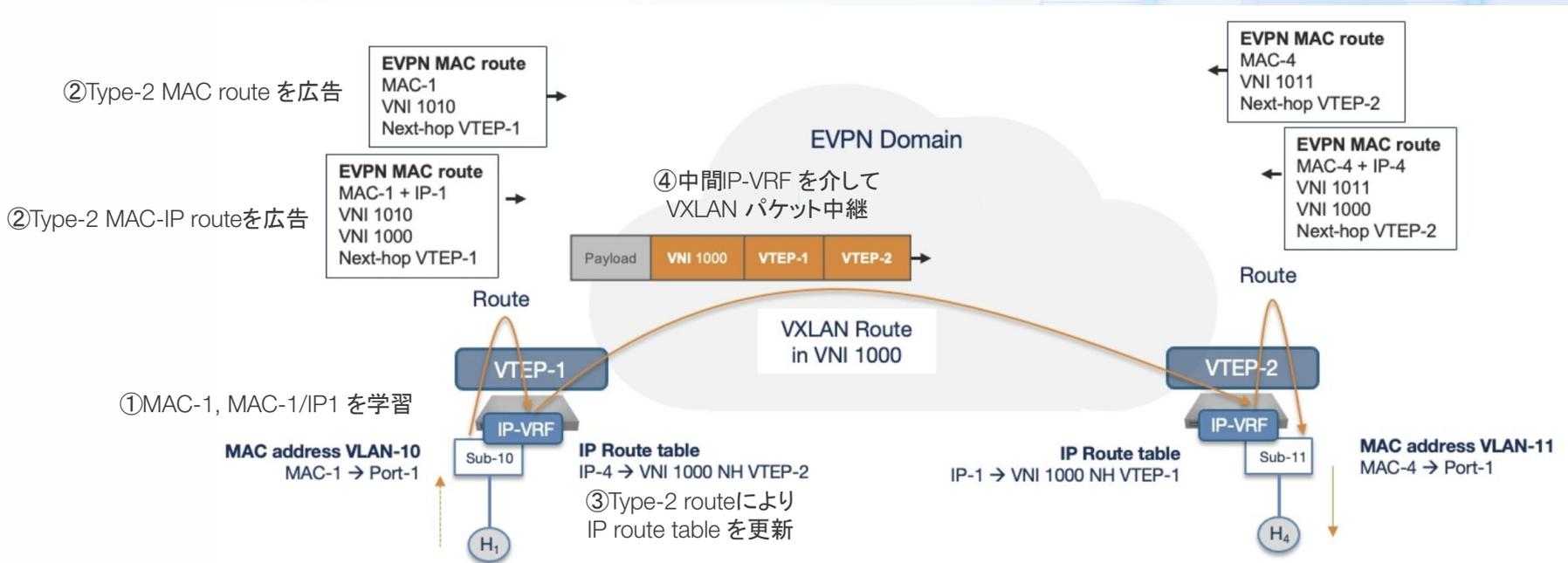


Asymmetric IRB

- Ingress VTEPでルーティング処理が行われる。その後宛先VTEPにVXLANブリッジされる
- Egress VTEPでは通常のVXLAN処理を行う
- 帰りでは逆の処理を行う
- VTEPは全テナントのAnycast GWの情報を持っている必要がある

Symmetric IRB

- ・ リモートVTEP上にある非ローカルサブネットへの接続は、中間にある共有VRFを介して実現
- ・ サブネット間ルーティングにおいて、インGRESS VTEPが中間IP-VRFを介して宛先サブネットにトラフィックをルーティングし、その後、イーグレスVTEPがそのIP-VRFから、ローカルで設定されている宛先サブネットへパケットをルーティングする
- ・ 直接接続されたホストにファーストホップ・ルーティングを提供するため各サブネットでエニキャストIPアドレスを利用。
- ・ サブネットに直接接続されたホストが存在するVTEP 上にのみ Anycast IP アドレスを設定
- ・ 直接接続されたホストについては、VTEPはデータプレーン学習とARPスヌーピングの組み合わせを介して、ホストのMACおよびMAC-IPバインディングを学習し、それぞれに対してType-2のMACおよびMAC-IPルートを広報



Symmetric IRB の設定例

Leaf ノードが、IP-VRF「tenant-a」のメンバーであるMAC-VRF (VLAN 10およびVLAN 11) の Anycast GW として動作

```
ip virtual-router mac-address 00:aa:aa:aa:aa:aa
!
vrf instance tenant-a
ip routing vrf tenant-a
!
interface Loopback0
    ip address 1.1.1.15/32
!
interface Loopback1
    ip address 2.2.1.15/32
!
interface Vxlan1
    vxlan source-interface Loopback1
    vxlan udp-port 4789
    vxlan vlan 10 vni 1010
    vxlan vlan 11 vni 1011
    vxlan vrf tenant-a vni 1000
!
vlan 10
!
vlan 11
!
interface Vlan10
    vrf vrf-a
    ip address virtual 10.10.10.254/24
!
interface Vlan11
    vrf vrf-a
    ip address virtual 10.10.11.254/24
```

```
router bgp 65101
    router-id 1.1.1.15
    maximum-paths 4
    neighbor SPINE peer group
    neighbor SPINE remote-as 65100
    neighbor SPINE send-community
    neighbor SPINE_EVPN peer group
    neighbor SPINE_EVPN remote-as 65100
    neighbor SPINE_EVPN update-source Loopback0
    neighbor SPINE_EVPN ebgp-multihop 5
    neighbor SPINE_EVPN send-community extended
    neighbor 1.1.1.11 peer group SPINE_EVPN
    neighbor 1.1.1.12 peer group SPINE_EVPN
    neighbor 172.11.15.11 peer group SPINE
    neighbor 172.12.15.12 peer group SPINE
    redistribute connected
!
vlan 10
    rd 1.1.1.15:1010
    route-target both 1010:1010
    redistribute learned
!
vlan 11
    rd 1.1.1.15:1011
    route-target both 1011:1011
    redistribute learned
!
address-family evpn
    neighbor SPINE_EVPN activate
!
address-family ipv4
    no neighbor SPINE_EVPN activate
!
vrf tenant-a
    router-id 1.1.1.15
    rd 1.1.1.15:1000
    route-target import evpn 1000:1000
    route-target export evpn 1000:1000
    redistribute connected
```

まとめ

- VXLAN により、レイヤ3 (L3)上で論理的なL2接続を実現可能
 - オーバーレイ接続によって物理的なIPファブリックに対して透過的
 - L3 IPファブリック上でL2の論理的な仮想接続を構成
- EVPN/VXLAN により、下記を実現可能
 - 単一の BGP による L2VPN
 - マルチホーミングによる柔軟な冗長化
 - 高速なコンバージェンス (Mass withdraw)
 - エッジ装置でのルーティング・ブリッジングを実現する IRB (Integrated Routing and Bridging)
 - 単一の BGP による L3VPN
 - IP-VPN (RFC4364)メカニズムを活用した RT (Route Target) によるルート制御
 - ARP proxy によるフラッディング抑制

リファレンス

- RFC 7348 Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks
 - <https://datatracker.ietf.org/doc/html/rfc7348>
- RFC 7432 BGP MPLS-Based Ethernet VPN
 - <https://datatracker.ietf.org/doc/html/rfc7432>
- RFC 8365 A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)
 - <https://datatracker.ietf.org/doc/html/rfc8365>
- RFC 9135 Integrated Routing and Bridging in Ethernet VPN (EVPN)
 - <https://datatracker.ietf.org/doc/html/rfc9135>
- RFC 9136 IP Prefix Advertisement in Ethernet VPN (EVPN)
 - <https://datatracker.ietf.org/doc/html/rfc9136>
- Arista Tech Library
 - https://tech-library.arista.com/data_center/evpnvxlan/feature_references/evpn_intro/

ARISTA

Thank You

www.arista.com



Appendix:

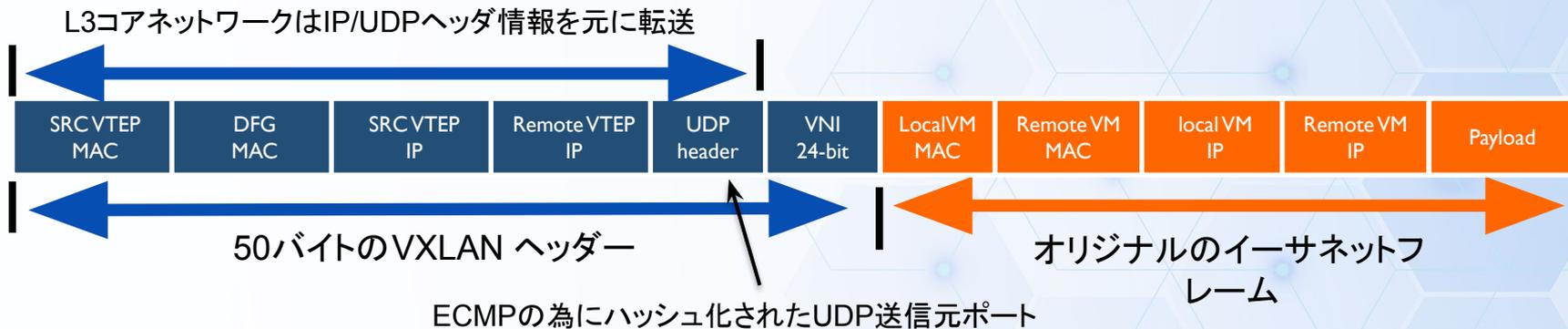
- VXLAN 概要 追加情報
- EVPN 概要 追加情報
- EVPN Bridging 追加情報
- EVPN Multihoming 追加情報
- EVPN IRB 追加情報
- EVPN GW

VXLAN 概要 追加情報

- VXLAN フレームフォーマット

VXLAN フレームフォーマット

- イーサネットヘッダーはローカルVTEPのMACアドレスとデフォルトゲートウェイのMACアドレス (14バイト+4バイト(802.1q)のオプション)
- IPヘッダーは送信元及び宛先VTEPのアドレス (20バイト)
- UDPヘッダーの送信元ポートはオリジナルのイーサネットヘッダーに基づいたハッシュ値、宛先ポートはIANA定義値 (8バイト) *VXLAN非対応のコアネットワークにおいてもECMPによるロードバランスを可能とする設計
- VXLANヘッダーは24ビットのVNIIによって1600万超のL2ドメインを実現 (8バイト)



VXLAN フレームフォーマットの各パラメータ

• 宛先ポート:

- IANAはVXLANのポートとして 4789をアサインした。このポートをデフォルトの宛先ポートとして使う

• 送信元ポート:

- UDPソースポート番号は、ロードバランスの際のハッシュの計算に使用される。動的にプライベートポート範囲49152-65535である事が推奨される

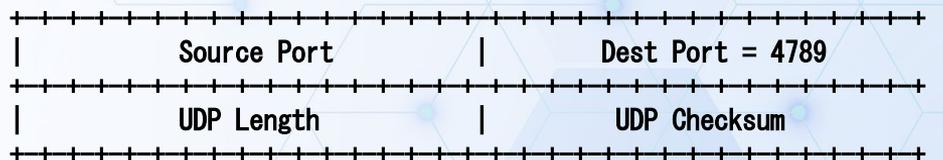
• フラグ(8 bits):

- 有効なVXLANのネットワークIDの場合、1フラグは1をセットしなければならない。他の 7ビット(R)は予約フィールドで送信時に0をセットしなければならない、受信時に無視される

• VXLANネットワーク識別子 (VNI):

- これは、通信するVMが配置されている個々のVXLANオーバーレイネットワークを指定するために使用される24ビットの値です。異なるVXLANオーバーレイネットワーク内のVMは互いに通信できません。

Outer UDP Header:



VXLAN Header:



EVPN 概要 追加情報

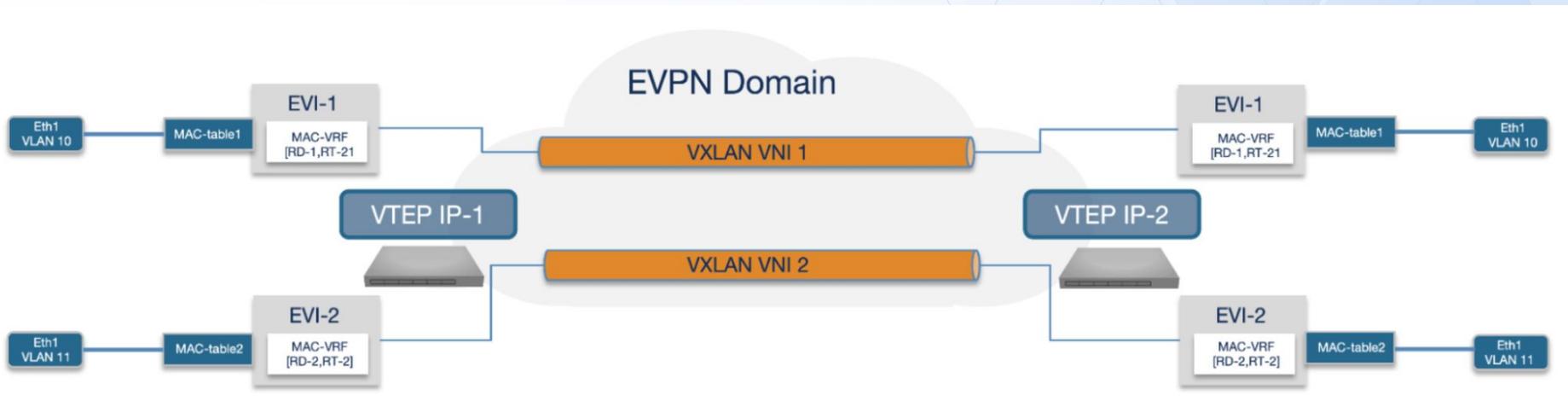
- EVPN Service モデル
 - VLAN-based service interface
 - VLAN-aware bundle service interface

EVPN Service モデル

- EVI は、EVPNドメイン内の複数のVTEPにまたがる論理スイッチを定義し、EVPNコントロールプレーンにおいて特定の MAC-VRF とその関連する RT および RD 値にマッピングされる
- 1つのEVI(すなわち1つのMAC-VRF)は、1つまたは複数のVLANを含むことができる
- VLAN(単数または複数)とEVI/MAC-VRFインスタンスとの関連付けは、EVPNサービスタイプによって定義
- 本資料では、**vlan-based** と **vlan-aware-bundle** という2つのサービスタイプを紹介

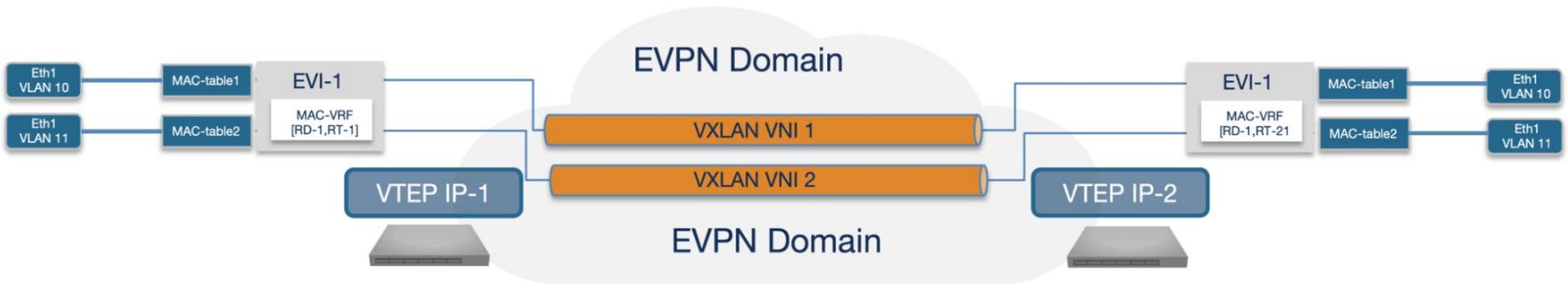
VLAN-based service interface

- VLAN-IDとEVIのMAC-VRFの間には1対1でマッピング
- EVIに関連付けられたRTにより、VLANを一意的に識別することが可能
- 各VTEPでどのType-2 (MAC)およびType-3 (IMET) ルートをインポートするかについて、VLANレベルで制御が可能
- 各VLANで、MAC-VRFインスタンスとそれに関連するRTおよびRDを設定する必要(設定量が増加)



VLAN-aware bundle service interface

- 複数のVLANが単一のEVI/MAC-VRFにマッピング
- MAC-VRFは関連付けられた各VLANに対して、固有のブリッジテーブルを保持
- コントロールプレーンにおいては、MAC-VRF内の個々のVLANとそれに関連するブリッジテーブルを識別するために、すべてのEVPNルートアップデート内でVLANタグが含まれる
- 転送プレーンにおいては、MAC-VRF内の各VLANに固有のVNIラベルが割り当てられる
- VLANとMAC-VRF間のN対1のマッピングにより、RTは個々のVLANではなく、MAC-VRFのみを識別することになる
- MAC-VRF内の特定のVLANの有無にかかわらず、同じMAC-VRFが存在していれば、そのMAC-VRFに対する全てのType-2およびType-3ルートがリモートノード上でインポートされる
- 単一のMAC-VRFインスタンスとその関連RTおよびRDの配下に複数のVLANを設定できるため、設定の簡素化が可能



Service interface 比較

- 2つの Service Interface は、**制御と設定の簡素化** という点で、異なるメリットを提供。どちらのサービスタイプが適切かは、導入環境の要件によって決まる
- データセンター環境の場合
 - 需要に応じてVLANが動的に展開・再展開されるような環境では、VLAN based サービス が適している
 - VTEP上で必要なType-2およびType-3ルートのみがインポートされることが保証されるため、ノード上のハードウェアおよびソフトウェアリソースを節約できる
- 動的変化の少ない環境の場合
 - オペレーターが一連のVTEP間で、常に同じVLANのバンドル(束)をトンネリングさせたいような環境では、VLAN aware bundle サービス が適している
 - 対象となる全VLANのType-2およびType-3ルートをインポートする必要があるため、設定の手間が削減できる

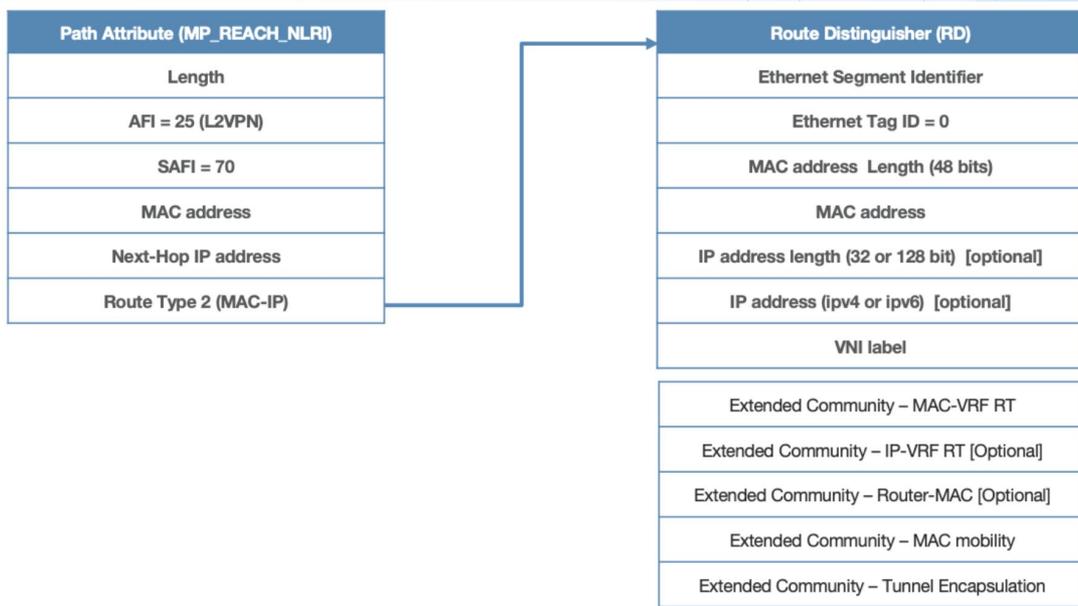
Service Type	VLANとEVIの関係	Route-Target	メリット	デメリット
VLAN based	VLANとEVIの間で 1対1 のマッピング	Type-2およびType-3ルート のRTは、EVIの VLAN を表 す	RTポリシーにより、MAC/IMETルートのイ ンポートを VLANレベル できめ細かく制 御できる	各テナントVLANごとに、MAC-VRFとRTを設定 する必要がある(設定量が多い)
VLAN aware bundle	複数のVLANとEVIの間で N対1 のマッピング	Type-2およびType-3ルート のRTは、VLANではなく EVI を表す	EVIにマッピングされるすべてのVLANに 対して、 単一のEVI を設定するだけで済 む(設定がシンプル)	RTポリシーによるType-2およびType-3ルート のインポート制御は、VLANレベルではなく EVI レベル となる(個別のVLAN単位での制御が難 しい)

EVPN Bridging 追加情報

- Type-2 route フォーマット詳細
- Vlan-aware bundle における Type-2 route 広告の参考リンク
- MAC Mobility
- Type-3 route フォーマット詳細
- Multicast underlay における BUM 処理の参考リンク

EVPN Type-2 MAC-IP route フォーマット

- Type-2 route は、MAC-VRFの RT と共に広告
 - 同じMAC-VRFのメンバーであるリモートVTEPは、RT import ポリシーに基づいた制御が可能
 - 経路が import されると、MACアドレスはVTEPの関連するMAC-VRF用ブリッジテーブルに追加され、ネクストホップは Type-2ルートを広告したVTEPとなる
- MACアドレスに関連付けられたネクストホップVTEPへ転送される際、Type-2 route で広告された VNI ラベルを用いて VXLAN encap. を行う



EVPN Type-2 MAC-IP route フォーマットの各パラメータ

- **Multiprotocol Reachable NLRI (MP_REACH_NLRI):** 広告されるルート「ネクストホップ」を運ぶために使用。VXLAN転送プレーンの文脈においては、広告を行っているVTEPの送信元アドレス(VTI)となる
- **Route-Distinguisher (RD):** MACアドレスがローカルで学習された、生成元MAC-VRFのRD(ルート識別子)
- **Ethernet Segment Identifier (ESI):** MACアドレスがEVPNマルチホーミングトポロジーのES上で学習された場合、その値を格納。ES上で学習されていない場合、このフィールドはゼロ
- **Ethernet Tag ID:** VLAN-aware bundle モードで動作している場合、MACアドレスがローカル学習されたブリッジドメインのイーサネットタグが入る。VLAN-based モードの場合はゼロに設定。VXLANカプセル化においてVLAN-aware bundleモードを使用する場合、イーサネットタグとしてVNI値 が使用
- **MAC address length:** ホストのMACアドレスの長さ
- **MAC address:** 広告されるホストのMACアドレス
- **IP address length [オプション]:** 広告されるIPアドレスの長さ(IPv4なら32ビット、IPv6なら128ビット)。このフィールドはオプションであり、ホストのMAC-IPバインディングを広告する設定になっている場合のみ格納
- **IP address [オプション]:** 広告されるホストのIPアドレス。このフィールドはオプションであり、ホストのMAC-IPバインディングを広告する設定になっている場合のみ格納
- **Label:** ホストが学習されたブリッジドメイン/MAC-VRFに対応する VNIラベル
- **Extended Community MAC-VRF route-target:** 広告されたMACアドレスが学習されたMAC-VRFのRT
- **Extended Community IP-VRF route-target:** 広告されたホストIPが学習されたIP-VRFのRT。これはオプションの拡張コミュニティであり、IRB転送に使用
- **Extended Community Router MAC:** 広告を行っているVTEPのルーターMACアドレス。これはオプションの拡張コミュニティであり、IRB転送に使用
- **Extended Community MAC Mobility:** 広告されるMACアドレスに関連付けられたシーケンス番号。これはMACアドレスの移動を示し、VTEP上でのMACテーブルの更新を促すために使用
- **Extended Community Tunnel encapsulation:** データプレーンにおいてVTEPが使用しているトンネルタイプを表す(=VXLAN)

Type-2 Route 広告の設定例 / 表示例 (VLAN-aware bundle)

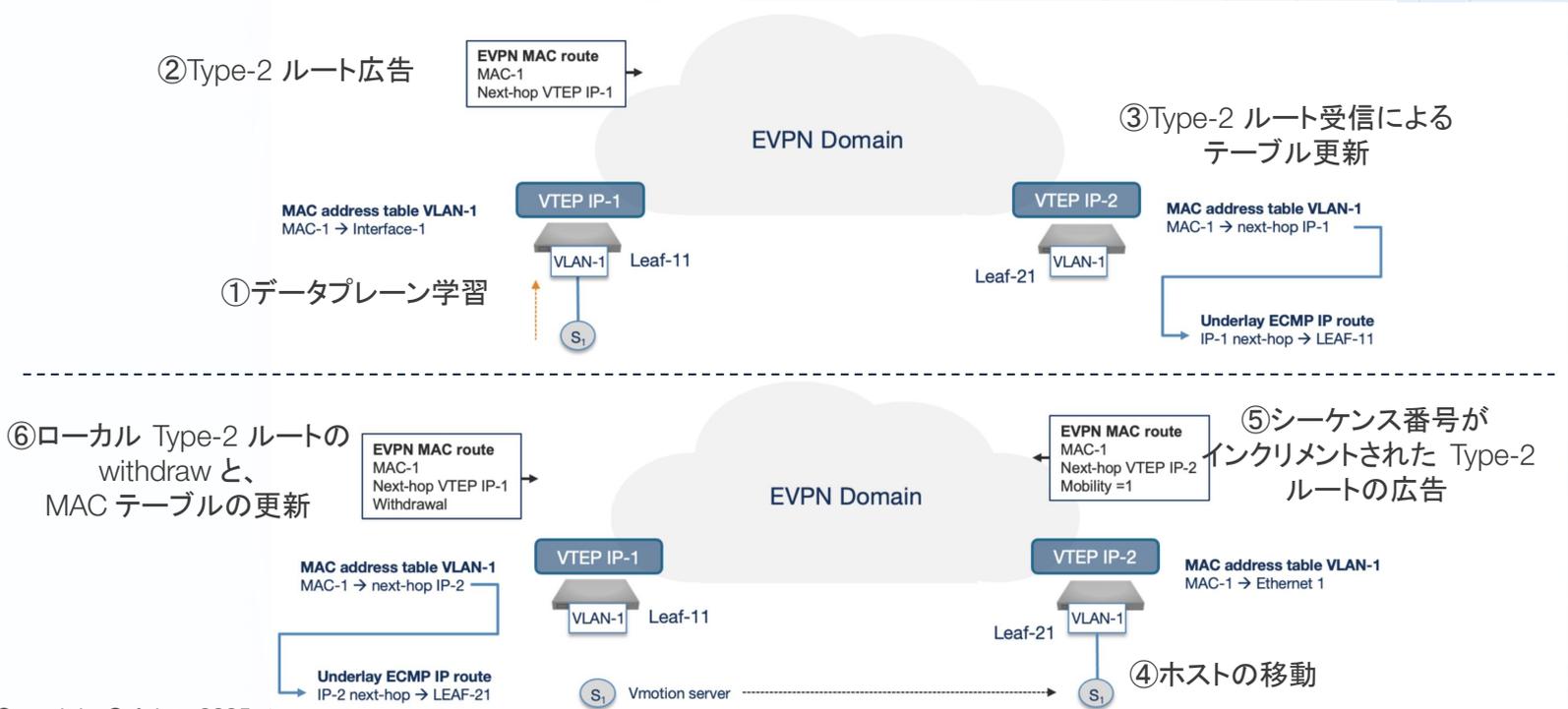
VLAN-aware bundle service interface における Type-2 route 広告の設定例/表示例は下記を参照ください

https://tech-library.arista.com/data_center/evpnvxlan/feature_references/evpn_bridge_operation/#212-type-2-route-advertisement-with-vlan-aware-bundle

(Arista Tech Library > Data Center > Reference Guides “EVPN/VXLAN” > 2. EVPN Bridging Operations > 2.1.2 Type-2 Route advertisement with VLAN aware bundle)

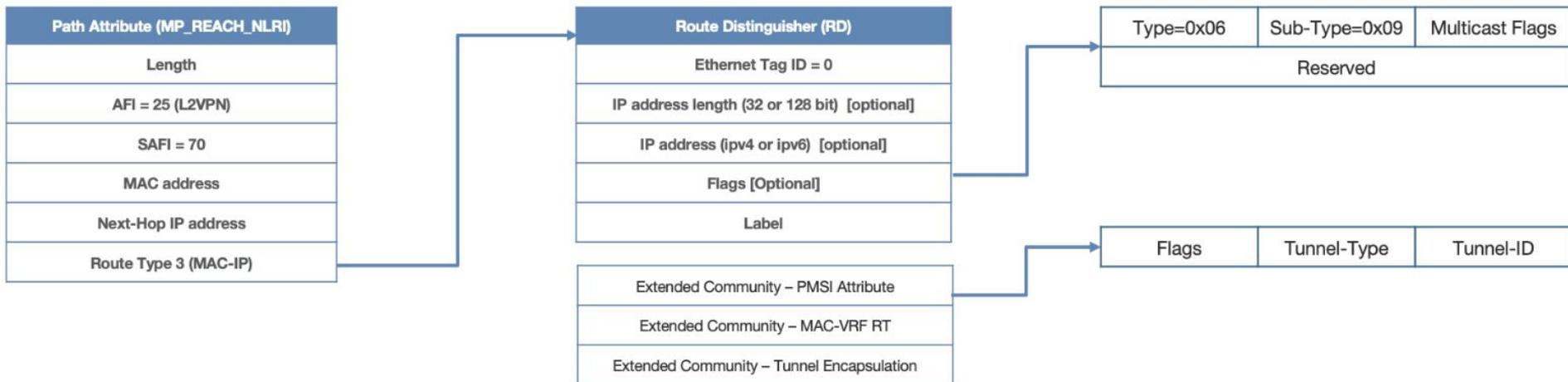
MAC Mobility

- ホスト (VMなど) が EVPN ドメイン内を移動した場合でも、全てのVTEPがMACテーブルを正しいネクストホップVTEPへ同期させるため、VTEPがMACの移動を検出すると、MAC Mobility拡張コミュニティを付与したType-2 MAC広告
- MACの移動が検出された場合、シーケンス番号がインクリメントされ、そのMACに対するType-2ルートのアップデートとして広告
- 元々そのMACを広告していたVTEPは、より高いシーケンス番号を持つType-2 MACルートを受信すると、自身のローカルMACルートをWithdraw
- EVPNドメイン内の他のすべてのVTEPは、古いMACルートを破棄し、より高いシーケンス番号を持つ新しいType-2ルートでテーブルを更新



EVPN Type-3 IMET ルート フォーマット

- BUMトラフィックの転送は、Type-3 ルートの広告によって制御
- 特定のブリッジメインのメンバーである各VTEPは、そのブリッジメインに対するType-3ルートを広告
 - Type-3 ルートには、VTEPのIPアドレス、MAC-VRF に関連付けられたRT が含まれる
 - RT に基づいてType-3ルートをインポートすることで、同じブリッジメイン内のピアVTEPを検出



EVPN type-3 IMET route フォーマットの各パラメータ

- **Multiprotocol Reachable NLRI:** 広告されたルートのネクストホップを運ぶために使用。VXLAN転送プレーンの文脈において、広告を行っているVTEPの送信元アドレス(VTI)に対応
- **Route-Distinguisher (RD):** IMETルートが広告されているMAC-VRFのRD
- **Ethernet Tag ID:** IMETルートが広告されているMAC-VRF内のブリッジドメインのイーサネットタグ。VLAN-based サービスの場合は値がゼロ。VLAN-aware bundle サービスの場合は関連するVNIが入る
- **Router IP length/address:** Type-3ルートを広告しているVTEPのIP(IPv4またはIPv6)アドレスを運ぶために使用
- **Flags fields:** IMETルート内のオプションフィールドであり、ブリッジドメイン内でのノードのEVPNマルチキャスト機能を広告するために使用
- **Extended Community MAC-VRF route-target:** IMETが関連付けられているMAC-VRFのRT
- **Extended Community PMSI Tunnel Attribute:** BUMトラフィックのトンネル転送メカニズムを広告するために使用。NVO標準内で定義されているサポートオプションは、Ingress Replication と IP Multicast

BUM traffic forwarding with multicast underlay

Multicast underlay による BUM 転送の概要/設定例/表示例は下記を参照ください

https://tech-library.arista.com/data_center/evpnvxlan/feature_references/evpn_bridge_operation/#222-bum-traffic-forwarding-with-multicast-underlay

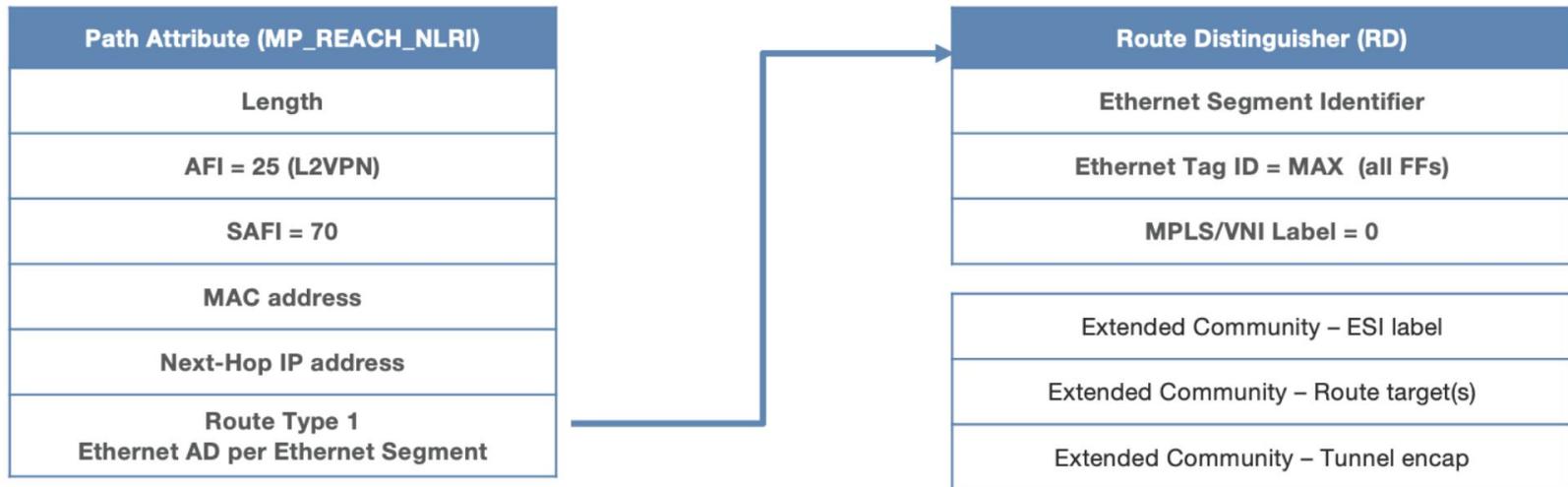
(Arista Tech Library > Data Center > Reference Guides “EVPN/VXLAN” > 2. EVPN Bridging Operations > 2.2.2 BUM traffic forwarding with multicast underlay)

EVPN Multihoming 追加情報

- Type-1 route フォーマット詳細
- 共通の ES に接続された VTEP における Type-2, Type-3 ルート広告
- Type-1 EAD per ES ルートによる Mass withdrawal
- Type-1 EAD-EVI ルートによる Aliasing
- DF Election アルゴリズムの種類
- Modulus DF Election アルゴリズムの概要/設定例/表示例
- All-Active トポロジにおける BUM Traffic 処理

Ethernet A-D per ES route フォーマット

- ノードは、Type-1 Ethernet AD route を使用して、ES への接続性を広告
 - タイプ1ルートには、Ethernet AD per Ethernet Segment と、Ethernet AD per EVI の2つのサブタイプがある
- Ethernet A-D per ESルートは、特定の ES への到達可能性を通知するために広告
- 高速コンバージェンス (Mass MAC Withdrawal) や、MPLS転送プレーンで展開される際のBUMトラフィック処理におけるスプリットホライズン(ループ防止)に使用

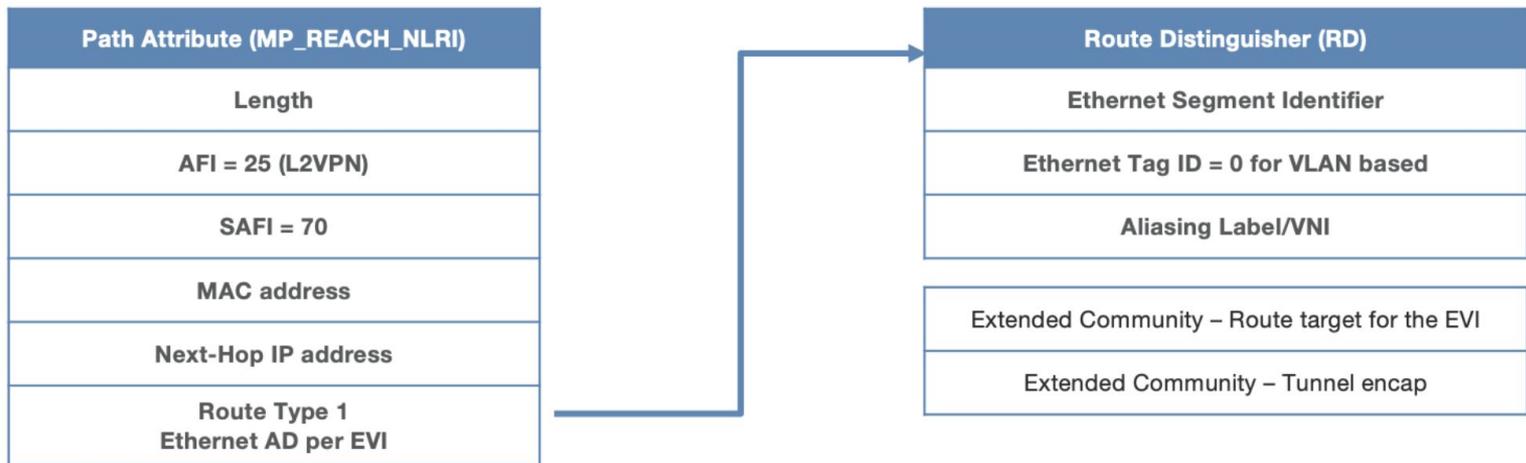


Ethernet A-D per ES route フォーマットの各パラメータ

- **Multiprotocol Reachable NLRI (MP_REACH_NLRI):** 広告されたルートのネクストホップを運ぶために使用。VXLAN転送プレーンの文脈においては、広告を行っているVTEPのIPアドレス
- **Route Distinguisher (RD):** Type 1 RDであり、通常は広告を行っているVTEPのループバックIPアドレスと、それに続く2バイトのローカル割り当て値で構成
- **Ethernet Segment Identifier (ESI):** 広告されるイーサネットセグメントの識別子
- **The Ethernet Tag ID field:** AD per ES route において、MAX_ETID(値は 0xFFFFFFFF)に設定
- **Label field:** A-D per ES route の場合、0 に設定
- **Extended Community ESI-label:** ESIラベルは、MPLS転送プレーンで展開される際のスプリットホライズンに使用。VXLAN転送プレーンの場合、ゼロに設定。
この拡張コミュニティには1オクテットのフラグも含まれており、その下位ビットは「Single-Active」ビットと呼ばれる。ES上でSingle-Active冗長モードが設定されている場合、ESI Label拡張コミュニティのフラグ内の「Single-Active」ビットは1に設定。一方、All-Activeが動作モードである場合、「Single-Active」ビットは0に設定
- **Extended Community Route-target:** 広告されたES上で設定されている各EVI(MAC-VRF)のRT が含まれる
- **Extended community Tunnel Encapsulation:** 転送プレーンで使用されるカプセル化方式 [MPLS | VXLAN]

Ethernet A-D per EVI route フォーマット

- Ethernet A-D per EVIルートは、ES上の特定のEVI(MAC-VRF)への接続性を通知するために広告
- このルートタイプは Aliasing 使用

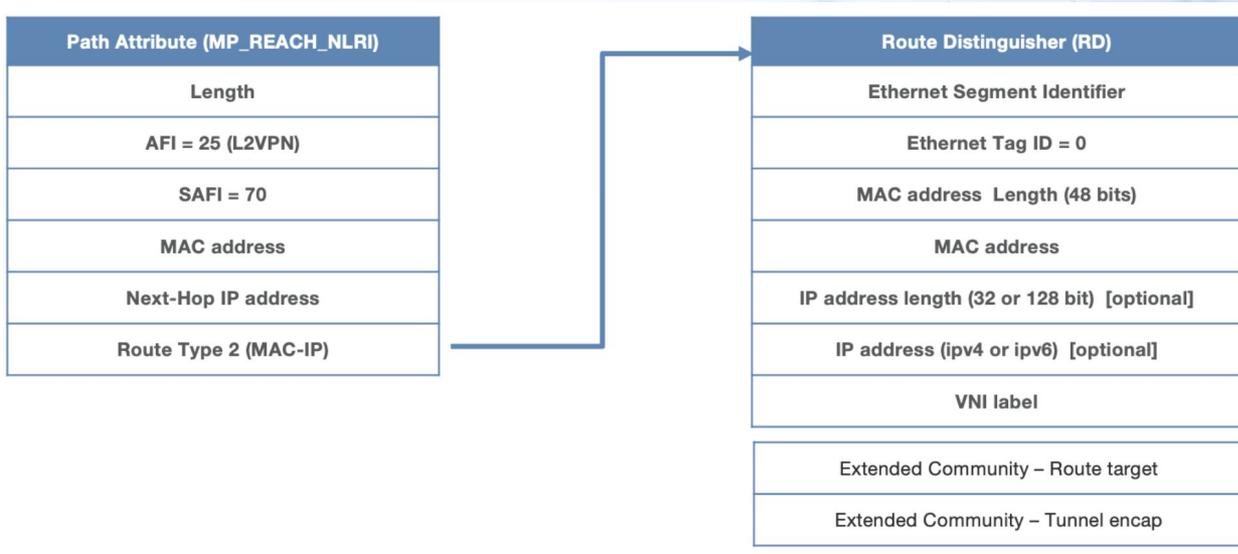


Ethernet A-D per EVI route フォーマットの各パラメータ

- **Multiprotocol Reachable NLRI (MP_REACH_NLRI):** 広告されたルートのネクストホップを運ぶために使用。VXLAN転送プレーンの文脈においては、広告を行っている VTEPのIPアドレス
- **Route Distinguisher (RD):** ルートを広告している MAC-VRF用の Type 1 RDです。通常、広告を行っているVTEPのループバックIPアドレスと、それに続くMAC-VRFに固有の番号で構成
- **Ethernet Segment Identifier (ESI):** 広告されるイーサネットセグメントの識別子
- **The Ethernet Tag ID field:** VLAN based サービスの場合は 0。VLAN-aware bundle サービスの場合は、そのVLANに関連付けられたVNIが入る
- **Aliasing-label:** VTEPによってまだ広告されていない MACアドレスに対してロードバランシングを行う際に使用されるラベル。
VXLAN転送プレーンの場合、これはその MACに関連付けられたブリッジドメインの VNI
- **Extended community Tunnel Encapsulation:** 転送プレーンで使用されるカプセル化方式 [MPLS | VXLAN]
- **Extended Community Route-target:** AD per EVIルートの場合、関連する EVI/MAC-VRFのRT

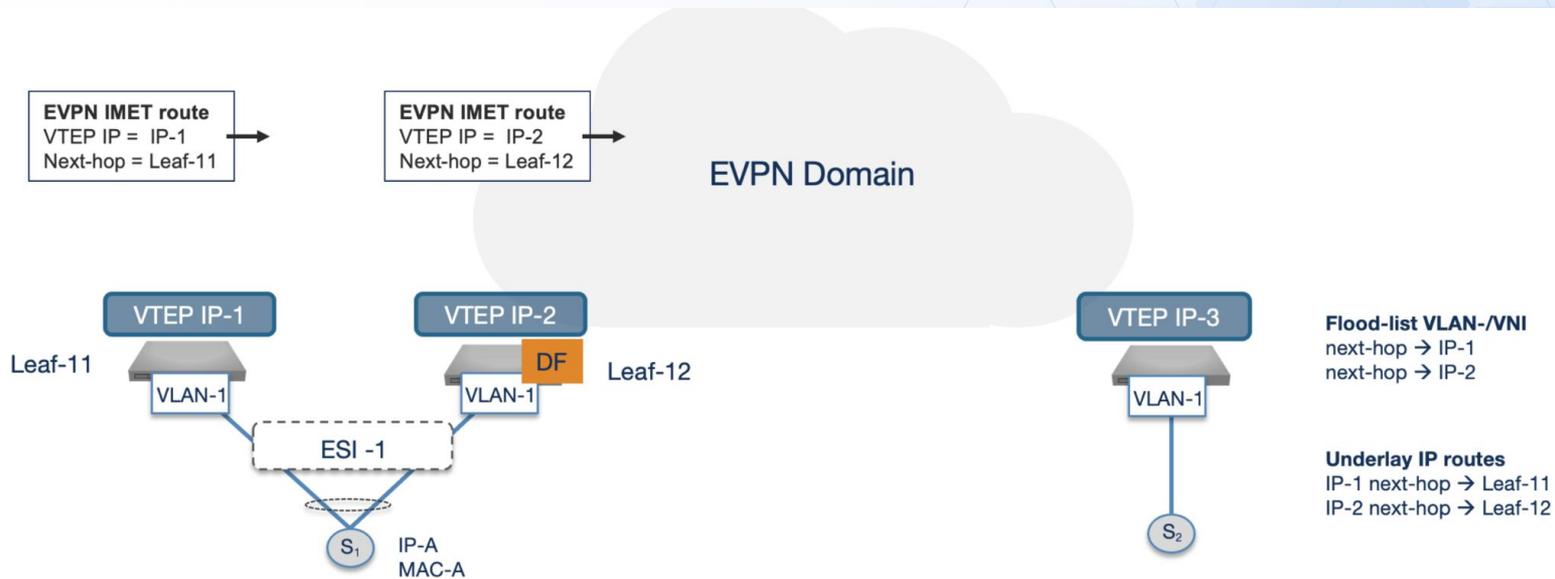
共通の ES に接続された VTEP における Type-2 ルート広告

- ホストのMACアドレスは、Type-2 ルート広告を通じて特定の ESに関連付けられる
- VTEPが、あるESのメンバーであるローカルリンク上で、データプレーン学習または静的設定によって MACアドレスを学習すると、そのホストに対する Type-2 ルートのESIフィールドには、その ESに関連付けられたESI値が格納される
 - Type-2ルート内のESI値は、そのMACアドレスを、同じESのメンバーであるピアノードと関連付けるために使用
- ESI値を持つType-2ルートをインポートしたリモート VTEPは、その特定の ESへの接続性を広告している全てのVTEPのネクストホップエントリを作成
 - 各VTEPは、Type-1ルート広告を通じてESへの接続性を広告



共通の ES に接続された VTEP における Type-3 ルート広告

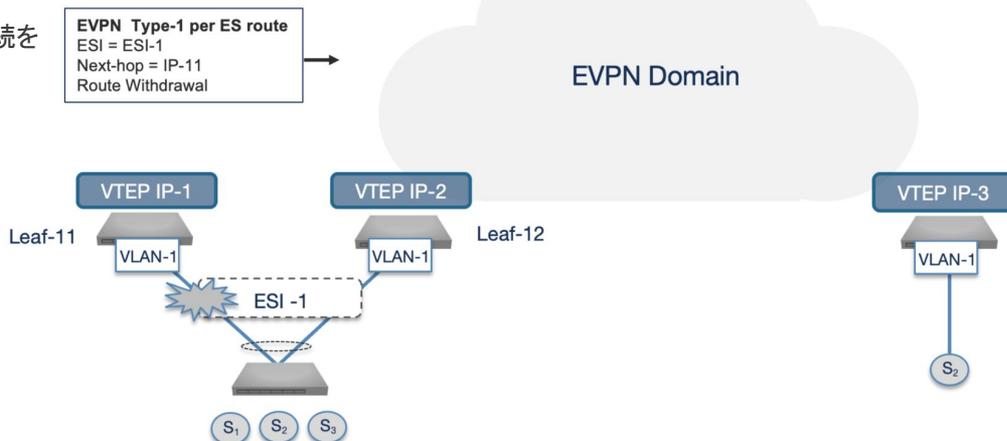
- 共通の ES に接続された各 VTEP は、トンネル ID として 固有の VTEP IP を含んだ Type-3 IMET ルート を広告
- リモート VTEP は ES 上の VTEP ごとに Type-3 ルートを受信し、Flood-list を作成
- ES 上の VLAN 宛ての BUM トラフィックは、その ES のメンバーであるすべての VTEP に対してフラディング
- 重複した BUM パケットが ES へ転送されるのを防ぐため、ES 上の VLAN に対して 1 つの VTEP が Designated Forwarder (DF) として選出され、その VTEP が ES への BUM トラフィック転送を担当
 - 残りのメンバーは非 DF として動作し、BUM トラフィックを破棄



Type-1 EAD per ES ルートによる Mass withdrawal

- Type-1 AD per ES ルートを使用して、ES への接続喪失後にMACの高速コンバージェンス(Mass Withdrawal)を提供
- 下記図では、VTEP-1とVTEP-2はESI-1のAll-Activeトポロジーにあり、サーバ1~3がデュアルホームされている
- VTEP-1がESへの接続を失った場合、VTEP-1は Type-1 AD-per-ES ルートを Withdraw
- VTEP-3はそのESIに関連して広告されていたすべてのMACアドレスのネクストホップからVTEP-1を削除
 - VTEP-3はVTEP-1がESIに接続されたホストごとのType-2ルートを個別に withdraw するのを待つ必要がなくなり、単一のType-1 AD per ESルートの withdraw だけで、関連するすべてのMACに対してVTEP-1を即座にネクストホップから外すことができるため、高速コンバージェンスが可能
- ピアノード (VTEP-2) への影響
 - ESIのピアノード (VTEP-2)もVTEP-1からのType-1撤回を受信
 - VTEP-2 が関連するESI上でMACをローカルに学習していた場合、Type-1ルートの撤回はMACエンタリに影響しない
 - VTEP-2 が関連するESI上でMACをローカルに学習しておらず、VTEP-1からのType-2広告に基づいてESIのローカルリンクに関連付けられている場合、そのMACアドレスはピアノードの転送テーブルから削除
 - VTEP-1のリンク障害により、VTEP-3 はトラフィックを残りのノード (ESのもう一方である VTEP-2) へロードバランスするようになり、その結果、そのノードがデータプレーンでMACを学習し、そのMACに対する新しいType-2ルートを広告

①VTEP-1 が ES への接続を失った際、EAD per ES ルートを withdraw



②ES に関連して広告されていた全ての MAC 転送テーブルのネクストホップからVTEP-1 を削除

MAC table 1

MAC-1, ESI-1 → IP-1
→ IP-2
MAC-2, ESI-1 → IP-1
→ IP-2
MAC-3, ESI-1 → IP-1
→ IP-2

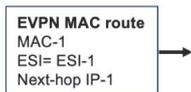
Type-1 EAD-EVI ルートによる Aliasing

- Aliasing では、あるESIに接続されたホストへの到達可能性は、「関連するSを含むType-2 MACルート」と、「Type-1 EAD per ES および Type-1 EAD per EVI ルートを通じてそのESへの接続性を広告しているVTEP」の組み合わせによって決定
- サーバーのポートチャネルでのロードバランシングにより、トラフィックはVTEP-1で受信され、サーバーのMACアドレスはVTEP-1でのみローカルに学習
- ローカル学習されたMACアドレスは、VTEP-1にて port-channel に関連付けられたESIを含むType-2ルートを広告
- VTEP-2はこのMACアドレスを学習しないためtype-2ルートを生成しない。
- サーバーのVLAN(MAC-VRF)に対するType-1 EAD per EVIルートと、MACアドレスが学習されているESIに対するType-1 EAD per ESルートを広告
- リモートノード(VTEP-3)は、VTEP-1からESI(ESI-1)付きのType-2ルートを受信すると、広告されたtype-1ルート(EAD per EVIおよびEAD per ES)から、そのESIに接続されている他のVTEPやES上のEVIを知ることが可能
- VTEP-3はこの情報を使って、VTEP-1とVTEP-2の両方をネクストホップに含むECMPセットとしてMAC転送テーブルを登録
- VTEP-2からはType-2ルートが広告されていないため、VTEP-3はそのMACに関連付けられたAD per EVIルートで広告されているAliasing Labelを使用して、VTEP-2へトラフィックを転送

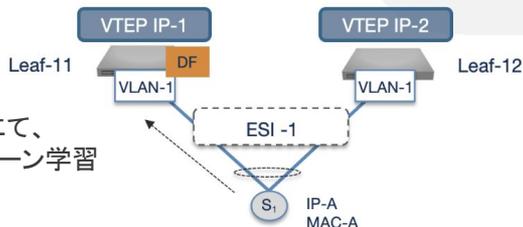
③VTEP-1,2にて、
Type-1 ルートを広告



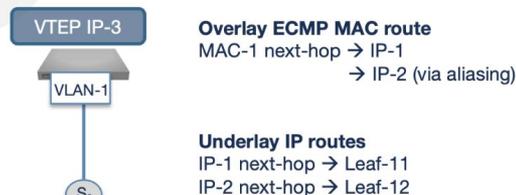
②VTEP-1にて、
ESIを含むType-2
ルートを広告



①VTEP-1にて、
データプレーン学習

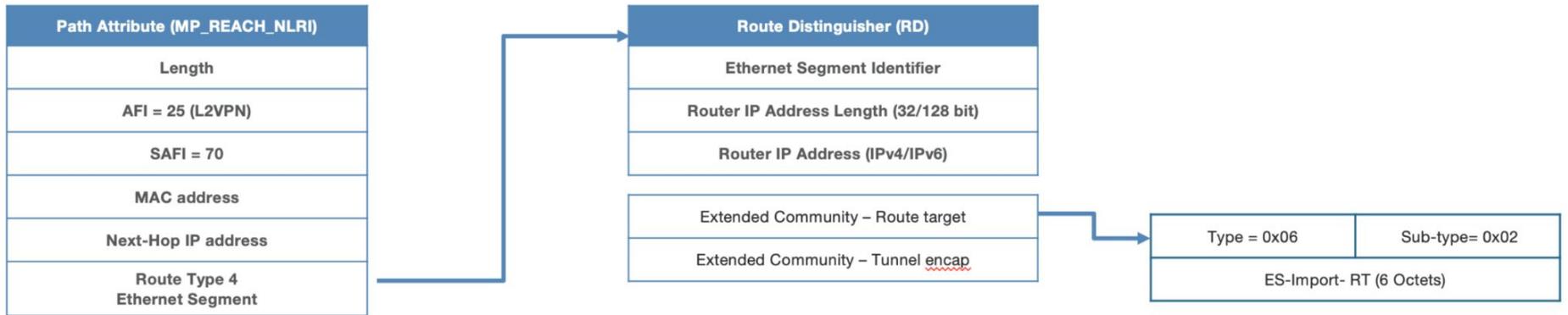


④Type-1, Type-2 ルートにより、
VTEP-1,2 をネクストホップに含む
ECMP セットとして
MAC 転送テーブルを登録



EVPN Type-4 Ethernet Segment route フォーマット

- ES-Import Route Target への導出方法は、ESIタイプによって異なる
 - タイプ1 (LACP)、タイプ2 (STP MAC)、およびタイプ3 (System MAC) の場合、ルートターゲットはESI値の9オクテット部分のうち、上位6オクテットから自動的に導出
 - その他のすべてのESIタイプについては、RTを手動で設定する必要があり、設定値はESの全ピアノード間で一致している必要がある



EVPN Type-4 ES route フォーマットの各パラメータ

- **Multiprotocol Reachable NLRI (MP_REACH_NLRI):** 広告されたルートネクストホップを運ぶために使用。VXLAN転送プレーンの文脈においては、広告を行っているVTEPのIPアドレス
- **Route Distinguisher (RD):** これはType 1 RDであり、通常は広告を行っているVTEPのルーパックIPアドレスと、それに続く2バイトのローカル割り当て値で構成
- **Ethernet Segment Identifier (ESI):** 広告されるイーサネットセグメントの識別子
- **IP address length:** OriginatorのルーターIPアドレスの長さ
- **Originator Router IP address:** Type-4ルートを生成したノードのIPアドレス
- **Extended community ES-Import RT:** 自動導出または手動設定が可能な、ESにおける固有のRT
- **Extended community Tunnel Encapsulation:** 転送プレーンで使用されるカプセル化方式 [MPLS|VXLAN]

DF Election アルゴリズムの種類

- EVPN標準(RFC 7432)は Modulus DFアルゴリズム の動作を定義
- 管理者による制御性の向上や、ESノード間でのDF機能の分散最適化などを目的として、Modulus DFアルゴリズムを改良するためのIETF標準がある

Standard (標準)	Comment (概要)
RFC 7432 (EVPN standard)	デフォルトの Modulus DF Election アルゴリズムを定義
RFC 8584 Framework for Ethernet VPN Designated Forwarder Election Extensibility	HRWアルゴリズムと、新しい DF Election Extended Community を定義
RFC 9785 Preference-Based EVPN Designated Forwarder (DF) Election	RFC 8584の手順を更新し、プリファレンス(優先度)ベースのアルゴリズムを導入

HRW DF Election アルゴリズム、Preference DF Election アルゴリズムの概要/設定例/表示例は下記を参照

https://tech-library.arista.com/data_center/evpnvxlان/feature_references/evpn_multihoming/#327-evpn-type-4-ethernet-segment-route

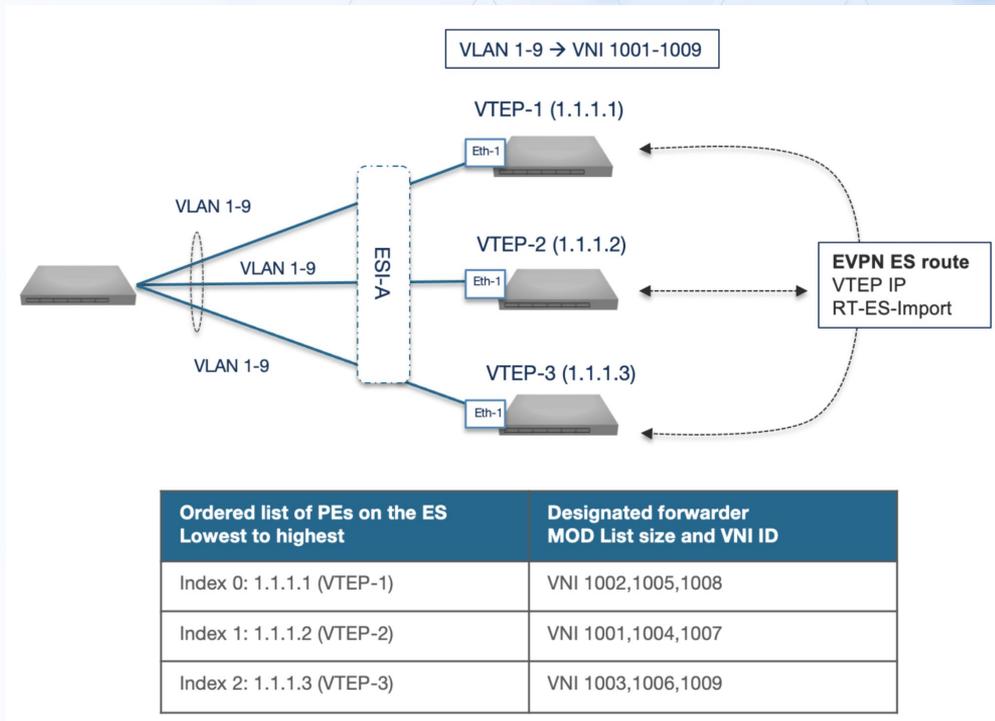
(Arista Tech Library > Data Center > Reference Guides “EVPN/VXLAN” > 3. EVPN Multihoming models > 3.2.7 EVPN Type-4 Ethernet Segment Route

Modulus DF Election アルゴリズム

- EVPN 標準 (RFC7432) の Modulus DF アルゴリズムは、ES のノード間で、DF機能を EVI 単位 で分散させることが可能

Modulus DF アルゴリズムの動作手順は下記のとおり

1. ローカル接続された ES ごとに、VTEPは関連する ES-import拡張コミュニティを含む Type-4 ES ルートを広告
2. VTEPは DF選出タイマー を開始し、同じ ES に接続された他のVTEPがType-4ルートを広告してくるのを待つ(デフォルトのDF選出タイマーは3秒)
3. タイマー満了後、ESに接続された各VTEPは、Type-4 ルートの生成元IPアドレスに基づいて、そのESへの接続性を広告しているVTEPの順序付きリストを作成
4. リスト内のVTEPには、最も低いIPアドレスを持つノードを「0」として、インデックス番号(0, 1, 2...)が付与
5. インデックスを使用して、MOD演算に基づき、ES上の特定のEVIに対してどのVTEPがDFになるかを決定



Modulus DF Election コンフィグ

- Modulus DF アルゴリズムは、Arista EOS のデフォルト動作モードであり、All-Active ES構成の設定をすることで Modulus DF アルゴリズムとして動作

```
vlan 10
!  
interface loopback 0  
  description local_vtep_IP  
  ip address 1.1.1.15/32  
!  
interface loopback 1  
  description shared_vtep_IP  
  ip address 2.2.1.15/32  
!  
interface vxlan1  
  vxlan source-interface Loopback1  
  vxlan udp-port 4789  
  vxlan vlan 10 vni 1010  
!
```

```
interface Ethernet1  
  description Harness  
  load-interval 5  
  channel-group 100 mode active  
!  
interface Port-Channel100  
  switchport trunk allowed vlan 10  
  switchport mode trunk  
!  
  evpn ethernet-segment  
    identifier 0000:0001:0001:0000:1112  
    route-target import 00:01:00:00:11:12  
  lacp system-id 0001.0000.1112  
!  
router bgp 65101  
  vlan 10  
    rd 1.1.1.15:1010  
    route-target both 1010:1010  
    redistribute learned
```

Modulus DF Election に関する表示例

- ES (0000:0001:0001:0000:1112) に対して Leaf-11, 12 が広告する Type-4 ESルート

```
DC1-Leaf11(config-evpn-es)#show bgp evpn route-type ethernet-segment detail
BGP routing table information for VRF default
Router identifier 1.1.1.15, local AS number 65101
BGP routing table entry for ethernet-segment 0000:0001:0001:0000:1112 2.2.1.15, Route Distinguisher: 2.2.1.15:1
Paths: 1 available
Local

- from - (0.0.0.0)
  Origin IGP, metric -, localpref -, weight 0, tag 0, valid, local, best
  Extended Community: TunnelEncap:tunnelTypeVxlan EvpnEsImportRt:00:01:00:00:11:12
```

```
DC1-Leaf12(config-evpn-es)#show bgp evpn route-type ethernet-segment detail
BGP routing table information for VRF default
Router identifier 1.1.1.16, local AS number 65102
BGP routing table entry for ethernet-segment 0000:0001:0001:0000:1112 2.2.1.16, Route Distinguisher: 2.2.1.16:1
Paths: 1 available
Local

- from - (0.0.0.0)
  Origin IGP, metric -, localpref -, weight 0, tag 0, valid, local, best
  Extended Community: TunnelEncap:tunnelTypeVxlan EvpnEsImportRt:00:01:00:00:11:12
```

Modulus DF Election に関する表示例

- ES上のEVI(VLAN 10)に対する DF Election の結果、Leaf-11 が DFノードに選出され、Leaf-12は非DFノードとして動作

```
DC1-Leaf11(config-if-Et49/1)#show bgp evpn instance vlan 10
EVPN instance: VLAN 10
Route distinguisher: 1.1.1.15:1010
Route target import: Route-Target-AS:1010:1010
Route target export: Route-Target-AS:1010:1010
Service interface: VLAN-based
Local VXLAN IP address: 2.2.1.15
VXLAN: enabled
MPLS: disabled
Remote ethernet segment:
ESI: 0000:0001:0001:0000:2122
Active TEPs: 1.1.1.17
Local ethernet segment:
ESI: 0000:0001:0001:0000:1112
Interface: Port-Channel100
Mode: all-active
State: up
ES-Import RT: 00:01:00:00:11:12
DF election algorithm: modulus
Designated forwarder: 2.2.1.15
Non-Designated forwarder: 2.2.1.16
```

```
DC1-Leaf12(config-if-Et49/1)#show bgp evpn instance vlan 10
EVPN instance: VLAN 10
Route distinguisher: 1.1.1.16:1010
Route target import: Route-Target-AS:1010:1010
Route target export: Route-Target-AS:1010:1010
Service interface: VLAN-based
Local VXLAN IP address: 2.2.1.16
VXLAN: enabled
MPLS: disabled
Remote ethernet segment:
ESI: 0000:0001:0001:0000:2122
Active TEPs: 1.1.1.17
Local ethernet segment:
ESI: 0000:0001:0001:0000:1112
Interface: Port-Channel100
Mode: all-active
State: up
ES-Import RT: 00:01:00:00:11:12
DF election algorithm: modulus
Designated forwarder: 2.2.1.15
Non-Designated forwarder: 2.2.1.16
```

All-Active トポロジにおける BUM Traffic 処理

A-A (All-Active) EVPN トポロジ内のノードによる BUM Traffic の処理は、その BUM Traffic を生成した送信元 VTEP が誰か (リモート VTEP なのか、同じ ES に接続されたピア VTEP なのか) に依存

- リモート VTEP から送信された (ES への Ingress) BUM Traffic の場合
 - リモートノードは、その VLAN に対する Type-3 IMET ルートで広告されている「送信元の VLAN に接続されたすべてのノード」へ BUM Traffic を複製
 - A-A トポロジでは、ES に接続された各ノードが BUM Traffic のコピーを受信
 - その VLAN に対して選出された DF が Traffic を ES へ転送する責任を負い、非 DF ノードは BUM Traffic を受信しますが破棄
- 同じ ES に接続されたピア VTEP から送信された (ES からの Egress) BUM Traffic の場合
 - Traffic ループを回避するためにスプリットホライズンを使用
 - VXLAN 転送プレーンにおいて、各ノードは Type-4 ES ルート広告を通じて、同じ ES に接続されている他のノードの VTEP IP アドレスを追跡
 - ノードは VXLAN カプセル化された BUM パケットを受信すると、VXLAN フレームの送信元 IP アドレスを検査し、送信元 VTEP と共有している ES に対して、そのパケットを転送しない
 - このスプリットホライズンフィルタリングメカニズムにより、送信元 VTEP は、アクセスポートから到着した BUM Traffic に対して、直接接続されているすべての ES への複製を常に行
 - この仕組みをローカルバイアス (Local Bias) という (RFC 8365 で定義)

EVPN IRB 追加情報

- Symmetric IRB における Type-2 MAC-IP ルートフォーマット
- Symmetric IRB の設定例/表示例
- Anycast GW

EVPN Type-2 ルートフォーマット (Symmetric IRB)

- Type-2 MAC-IP ルートは、VTEPのルーター-MACとともに、MACアドレスが学習されたMAC-VRFとそれに関連するIP-VRFに対応する2つのラベルと2つのRTで広報される
 - 受信側VTEPにIP-VRFが存在する場合、Type-2 MAC-IP ルートがVRFルーティングテーブルにホストルート(32)としてインストールされる
 - ルートが広報元VTEPのRT、VNI、およびルーター-MACとともに広報される
- 対応するMAC-VRFがVTEPに存在する場合、MAC-IPバインディングもローカルARPテーブルとしてインストールされ、ARP抑制に使用。このインポート動作は、MAC-VRF の RT および VNI と共にルートが広告されることによって制御される

Type-2 MAC only route

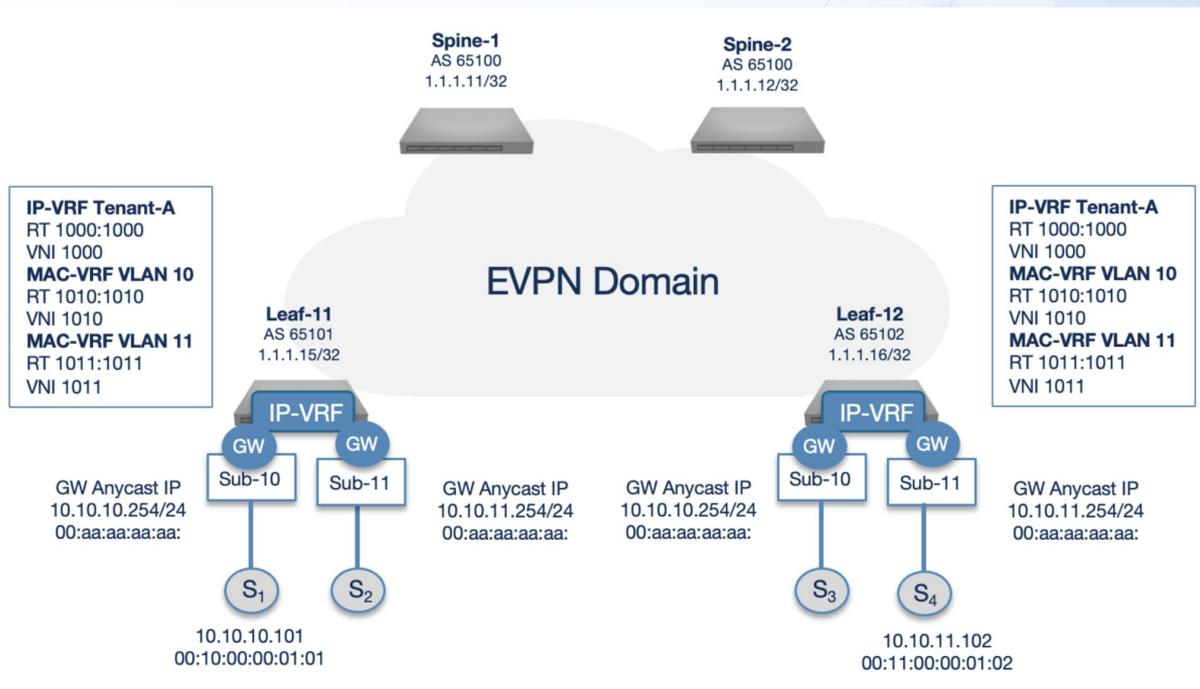
Route Distinguisher (RD)
Ethernet Segment Identifier
Ethernet Tag ID = 0
MAC address Length (48 bits)
MAC address
IP address length (32 or 128 bit) = 0
IP address (IPv4 or IPv6) = 0
VNI label
Extended Community – MAC-VRF RT
Extended Community – MAC mobility
Extended Community – Tunnel Encapsulation

Type-2 MAC-IP route

Route Distinguisher (RD)
Ethernet Segment Identifier
Ethernet Tag ID = 0
MAC address Length (48 bits)
MAC address
IP address length (32 or 128 bit)
IP address (IPv4 or IPv6)
VNI label (MAC-VRF)
VNI label (IP-VRF) [optional]
Extended Community – MAC-VRF RT
Extended Community – IP-VRF RT [optional]
Extended Community – Router-MAC [optional]
Extended Community – MAC mobility
Extended Community – Tunnel Encapsulation

Symmetric IRB の設定例

- VRFに関連付けられたMAC-VRFのデフォルトの動作は、Symmetric IRB動作用にType-2ルートを広告
 - ローカルで学習されたホストのMACアドレスに対して、タイプ2(MACのみ)ルートを広告
 - ARPスヌーピングを通じて学習されたARPバインディングに対して、デュアルVNIおよびRT(L2 VNIとL3 VNIの両方)を含む2つ目のType-2(MAC-IP)ルートを広告



Symmetric IRB の設定例

Leaf ノードが、IP-VRF「tenant-a」のメンバーであるMAC-VRF (VLAN 10およびVLAN 11) の Anycast GW として動作

```
ip virtual-router mac-address 00:aa:aa:aa:aa:aa
!
vrf instance tenant-a
ip routing vrf tenant-a
!
interface Loopback0
    ip address 1.1.1.15/32
!
interface Loopback1
    ip address 2.2.1.15/32
!
interface Vxlan1
    vxlan source-interface Loopback1
    vxlan udp-port 4789
    vxlan vlan 10 vni 1010
    vxlan vlan 11 vni 1011
    vxlan vrf tenant-a vni 1000
!
vlan 10
!
vlan 11
!
interface Vlan10
    vrf vrf-a
    ip address virtual 10.10.10.254/24
!
interface Vlan11
    vrf vrf-a
    ip address virtual 10.10.11.254/24
```

```
router bgp 65101
    router-id 1.1.1.15
    maximum-paths 4
    neighbor SPINE peer group
    neighbor SPINE remote-as 65100
    neighbor SPINE send-community
    neighbor SPINE_EVPN peer group
    neighbor SPINE_EVPN remote-as 65100
    neighbor SPINE_EVPN update-source Loopback0
    neighbor SPINE_EVPN ebgp-multihop 5
    neighbor SPINE_EVPN send-community extended
    neighbor 1.1.1.11 peer group SPINE_EVPN
    neighbor 1.1.1.12 peer group SPINE_EVPN
    neighbor 172.11.15.11 peer group SPINE
    neighbor 172.12.15.12 peer group SPINE
    redistribute connected
!
vlan 10
    rd 1.1.1.15:1010
    route-target both 1010:1010
    redistribute learned
!
vlan 11
    rd 1.1.1.15:1011
    route-target both 1011:1011
    redistribute learned
!
address-family evpn
    neighbor SPINE_EVPN activate
!
address-family ipv4
    no neighbor SPINE_EVPN activate
!
vrf tenant-a
    router-id 1.1.1.15
    rd 1.1.1.15:1000
    route-target import evpn 1000:1000
    route-target export evpn 1000:1000
    redistribute connected
```

Symmetric IRB type-2 route の表示例

- Leaf-11がServer-1 (0010.0000.0101 / 10.10.10.101)についてローカルで学習したMACとIPのバインディングは、デュアルVNI および RT と共に広告

```
DC1-Spine1(config)#show bgp evpn route-type mac-ip 0010.0000.0101 detail
BGP routing table information for VRF default
Router identifier 1.1.1.11, local AS number 65100
BGP routing table entry for mac-ip 0010.0000.0101, Route Distinguisher: 1.1.1.15:1010
Paths: 1 available
65101
  2.2.1.15 from 1.1.1.15 (1.1.1.15)
  Origin IGP, metric -, localpref 100, weight 0, valid, external, best
  Extended Community: Route-Target-AS:1010:1010 TunnelEncap:tunnelTypeVxlan
  VNI: 1010
BGP routing table entry for mac-ip 0010.0000.0101 10.10.10.101, Route Distinguisher: 1.1.1.15:1010
Paths: 1 available
65101
  2.2.1.15 from 1.1.1.15 (1.1.1.15)
  Origin IGP, metric -, localpref 100, weight 0, valid, external, best
  Extended Community: Route-Target-AS:1000:1000 Route-Target-AS:1010:1010
TunnelEncap:tunnelTypeVxlan EvpnRouterMac:74:83:ef:95:fc:5e
  VNI: 1010 L3 VNI: 1000
```

Symmetric IRB bridging and routing tables

- Leaf-11 によって広告された2 つの Type-2 ルートが Symmetric IRB フォーマットで広告された結果、Leaf-12 は MAC のみのルートを、Leaf-11 をネクストホップとするVLAN 用ローカル MAC テーブルにインポート
- MAC-IP ルートによって、ARP 抑制用のローカルARP エントリがソフトウェア上に作成され、VRF のルーティングテーブルにはホストルートが作成

```
DC1-Leaf12(config-if-Vl10)#show mac address-table dynamic address 0010.0000.0101
Mac Address Table
```

```
-----
Vlan      Mac Address          Type          Ports          Moves          Last Move
-----
10        0010.0000.0101      DYNAMIC      Vx1            1              36 days, 6:18:09 ago
Total Mac Addresses for this criterion: 1
```

```
DC1-Leaf12(config-if-Vl10)#show vxlan address-table
Vxlan Mac Address Table
```

```
-----
VLAN      Mac Address          Type          Prt  VTEP          Moves          Last Move
-----
10        0010.0000.0101      EVPN          Vx1  2.2.1.15      1              36 days, 6:20:31 ago
Total Remote Mac Addresses for this criterion: 1
```

```
DC1-Leaf12(config-if-Vl10)#show arp vrf tenant-a
```

```
Address          Age (sec)  Hardware Addr  Interface
10.10.10.101      -          0010.0000.0101 Vlan10, Vxlan1
```

Symmetric IRB bridging and routing tables

- Leaf-11 によって広告された2 つの Type-2 ルートが Symmetric IRB フォーマットで広告された結果、Leaf-12 は MAC のみのルート、Leaf-11 をネクストホップとするVLAN 用ローカル MAC テーブルにインポート
- MAC-IP ルートによって、ARP 抑制用のローカルARP エントリがソフトウェア上に作成され、VRF のルーティングテーブルにはホストルートが作成

```
DC1-Leaf11(config-if-Vl10)#show ip route vrf tenant-a

VRF: tenant-a

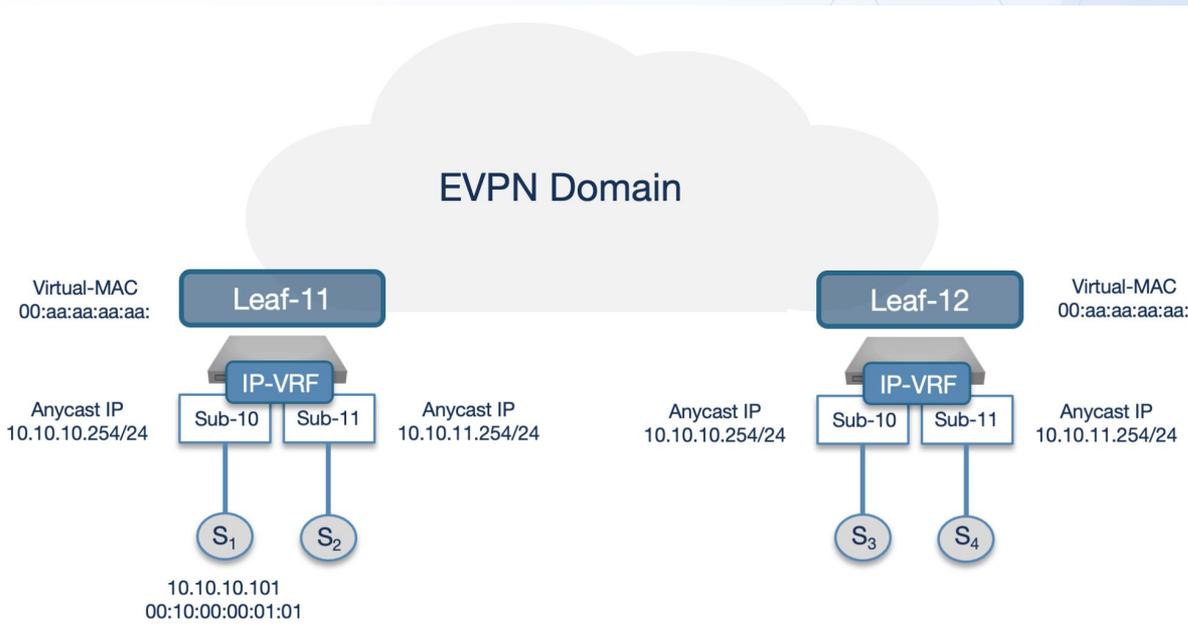
--- output truncated-----

Gateway of last resort is not set

B E      10.10.10.101/32 [200/0]
          via VTEP 2.2.1.15 VNI 1000 router-mac 74:83:ef:96:12:0c
C        10.10.11.0/24
          directly connected, Vlan11
C        10.10.31.0/24
          directly connected, Vlan31
```

Anycast Gateway

- IRB において、VTEP はドメイン内のサブネットに対する分散 Anycast GW として機能
- ip virtual router コマンドを使用する従来の VARP アプローチでは、VARP アドレス(仮想アドレス)に加えて、各サブネットの各 VTEP 上に「固有の物理 IP アドレス」を設定する必要があり、結果として IP アドレスプールの利用効率が低下する(無駄に消費する)という問題があった
- EVPN トポロジーにおいてこの問題を解決するために、ip address virtual コマンドを導入
- ip address virtual コマンドを使用すると、関連する各 VTEP 上に物理 IP アドレスを設定することなく、Anycast GW アドレスを設定することが可能



Anycast Gatewayの設定例

- 2つのサブネット (Sub-10およびSub-11) の Anycast GWとして動作している両方の VTEPには、各サブネットに対して「共通の仮想ルーター MAC」と「エニキャスト GW IP」を設定
- Anycast GW IPは、interface vlan 配下で、ip address virtual コマンドを使用して設定
- この設定により、各ノードで追加の固有 IPアドレスを消費することなく、複数のノードにまたがって Anycast GW IPを設定することが可能

```
ip virtual-router mac-address 00:aa:aa:aa:aa:aa
```

←設定された “ip address virtual に対するARP応答において、
VTEP が使用する仮想ルーターの MACアドレス

```
!  
vrf instance vrf-a  
ip routing vrf vrf-a
```

```
!  
vlan 10
```

```
!  
vlan 11
```

```
!  
interface Vlan10  
  vrf vrf-a  
  ip address virtual 10.10.10.254/24
```

```
!  
interface Vlan11  
  vrf vrf-a  
  ip address virtual 10.10.11.254/24
```

Anycast Gatewayの動作例

- ip address virtual を使用して Anycast GW IP を設定すると、仮想 IP とやり取りされる ARP は、設定された 仮想ルーター MAC アドレス を送信元にする
- エンドホストへ送出されるルーティングパケットや、VTEP 上に設定された仮想アドレス宛での ICMP トラフィックについては、関連する VTEP の システム MAC を送信元として使用
- 下記キャプチャ結果は、Leaf-11 に直接接続された Server-1 (10.10.10.101 / 00:10:00:00:01:01) と、設定されたエニキャスト GW IP (10.10.10.254 / 00:aa:aa:aa:aa:aa) の間の通信フロー
 - パケットキャプチャでは、Server が Anycast GW IP に対して ARP リクエストを送信しており、それに対する ARP レスポンスは Leaf-11 から送信され、送信元 MAC は設定された仮想 MAC アドレス 00:aa:aa:aa:aa:aa になっている
 - 続いて Anycast IP に対して ICMP リクエストを実行すると、Leaf-11 は ICMP レスポンスにおいて、自身の システム MAC を使用。これは、VTEP によってルーティングされるすべてのパケットについても同様

【Server から Anycast GW IP への ARP リクエスト】

```
00:10:00:00:01:01 > Broadcast, ethertype 802.1Q (0x8100), length 64: vlan 10, p 0, ethertype ARP, Request who-has 10.10.10.254 tell 10.10.10.101, length 46
```

```
00:aa:aa:aa:aa:aa > 00:10:00:00:01:01, ethertype 802.1Q (0x8100), length 60: vlan 10, p 0, ethertype ARP, Reply 10.10.10.254 is-at 00:aa:aa:aa:aa:aa, length 42
```

【Server から Anycast GW IP への ICMP Ping】

```
00:10:00:00:01:01 > 00:aa:aa:aa:aa:aa, ethertype 802.1Q (0x8100), length 102: vlan10, p0, ethertype IPv4, 10.10.10.101 > 10.10.10.254: ICMP echo request, id 26481, seq 10, length 64
```

```
74:83:ef:96:12:0c > 00:10:00:00:01:01, ethertype 802.1Q (0x8100), length 102: vlan10, p0, ethertype VTEP IPv4, 10.10.10.254 > 10.10.10.101: ICMP echo reply, id 26481, seq 10, length 64
```

EVPN GW

- EVPN GW 参考リンク

EVPN GW

EVPN GW の詳細/設定例/表示例は下記を参照

- EVPN VXLAN to VXLAN Gateway
 - https://tech-library.arista.com/data_center/evpnvxlan/feature_references/evpn_gw_vxlan_to_vxlan/
- EVPN VXLAN to MPLS Gateway
 - https://tech-library.arista.com/data_center/evpnvxlan/feature_references/evpn_gw_vxlan_to_mpls/