NSWI184 – Řízení počítačových sítí

Přednáška jedenáctá

Ondřej Zajíček, Kateřina Kubecová

2025-12-17

BGP EVPN Overview

- ▶ Ethernet VPN
- ▶ BGP-based control plane for L2 VPN services
- ▶ Supports both MPLS and VXLAN data planes
- ▶ RFC 7432 (MPLS) and RFC 8365 (VXLAN)

L2 vs L3 VPNs

- ▶ L3 VPN: Virtual/distributed router
  (learns routes by routing protocols)
- ▶ L2 VPN: Virtual/distributed switch
  (learns link addresses in data plane)

L2 VPN Challenges

- BUM frames (broadcast, unknown, multicast)
- MAC mobility
- Multi-homing redundancy

L2 VPN Variants

- Pseudowires (VPWS) vs LANs (VPLS)
- 'Physical' LANs vs VLANs
- RFC 4664 (framework)

BGP L2VPN History

- RFC 4761: VPLS with BGP signalling
- L2VPN AFI 25, VPLS SAFI 65
- NLRI for auto-discovery and label distribution
- MAC learning in data plane
- Later replaced by EVPN

EVPN Advantages

- ▶ Data plane MAC learning replaced by control plane
- ▶ MACs advertised via BGP routes
- ▶ Redundancy and load balancing with multi-homing
- ▶ Reduced flooding through proxy ARP/ND
- ▶ MAC mobility

EVPN-MPLS

- ▶ Original EVPN implementation over MPLS data plane
- ▶ Uses MPLS labels for forwarding
- ▶ ISP and enterprise use cases
- ▶ Integrates with existing MPLS infrastructure

EVPN-VXLAN

- EVPN control plane with VXLAN data plane
- Uses UDP encapsulation over IP networks
- Popular in data center environments
- Enables L2 extension over L3 infrastructure

EVPN and L3VPN Similarities

- P/PE/CE routers
- Route distinguishers
- Route targets
- VRFs vs MAC-VRFs
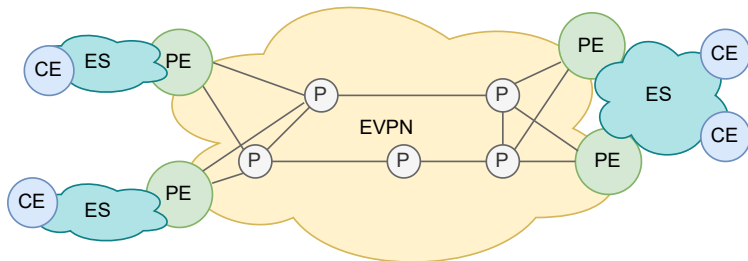
Terminology

- ▶ EVPN instance (EVI)
- ▶ Broadcast domain (~ VLAN)
- ▶ Ethernet tag (~ VLAN ID)
- ▶ MAC-VRF (~ virtual VLAN-aware bridge in PE)
- ▶ Bridge table (per broadcast domain)
- ▶ Ethernet segment (ES)
- ▶ ES identifier (ESI)

EVPN and VLANs

- **VLAN-Based Service**
  (1 EVPN instance $\sim$ 1 broadcast domain, no tag)
  Possible VLAN – EVPN Instance – VLAN mapping

- **VLAN Bundle Service**
  (1 EVPN instance $\sim$ N broadcast domains, no tag)
  Ethernet frames encapsulated with original VLAN ID header

- **VLAN-Aware Bundle Service**
  (1 EVPN instance $\sim$ N broadcast domains identified by tags)
  Possible VLAN – EVPN Instance and tag – VLAN mapping

BGP EVPN

- Carries EVPN routes between PEs
- L2VPN AFI 25, EVPN SAFI 70
- Multiple EVPN route types
- Format: TLV (1B type, 1B length, variable data)
- Route distinguisher first in data
- Route targets as extended communities
- MPLS labels / VXLAN IDs in data

EVPN Route Types

- ▶ Type 1: Ethernet Auto-Discovery (EAD)
- ▶ Type 2: MAC/IP Advertisement
- ▶ Type 3: Inclusive Multicast Ethernet Tag (IMET)
- ▶ Type 4: Ethernet Segment (ES)
- ▶ Type 5: IP Prefix (RFC 9136, for L3 services)

Ethernet Auto-Discovery

- Represents ethernet segment reachable through PE
- Used for fast convergence and load balancing
- Not required in single-homing setups

Ethernet Auto-Discovery

```
+----------------------------------------+
|  Route Distinguisher (RD) (8 octets)   |
+----------------------------------------+
|Ethernet Segment Identifier (10 octets)|
+----------------------------------------+
|  Ethernet Tag ID (4 octets)           |
+----------------------------------------+
|  MPLS Label (3 octets)                | X
+----------------------------------------+
```

MAC/IP Advertisement

- ▶ Represents unicast MAC reachable through PE
- ▶ Optional IP information for ARP/ND proxy
- ▶ ES identifier for multi-homing
- ▶ Ethernet tag ID for VLAN info
- ▶ Label for encapsulation

MAC/IP Advertisement

```
+---------------------------------------+
|  RD (8 octets)                        |
+---------------------------------------+
|Ethernet Segment Identifier (10 octets)| X
+---------------------------------------+
|  Ethernet Tag ID (4 octets)           |
+---------------------------------------+
|  MAC Address Length (1 octet)         |
+---------------------------------------+
|  MAC Address (6 octets)               |
+---------------------------------------+
|  IP Address Length (1 octet)          |
+---------------------------------------+
|  IP Address (0, 4, or 16 octets)      |
+---------------------------------------+
|  MPLS Label1 (3 octets)               | X
+---------------------------------------+
|  MPLS Label2 (0 or 3 octets)          | X
+---------------------------------------+
```

MAC/IP Advertisement

- Local data plane MAC learning (bridge table)
- Local IP-MAC learning (neighbor table)
- Other sources (DHCP database)
- MAC announced as Type 2 EVPN route
- Remote PEs install it as MAC entry to appropriate bridge table
- BGP next hop resolved to get immediate next hop and outer label
- Or used directly in case of VXLAN encapsulation
- When ESI is used, it is first resolved to EAD route,
  which is used for BGP next hop and labels

ARP/ND Suppression

- PE routers act as ARP/ND proxies
- Based on IP-MAC bindings from MAC/IP advertisements
- Reduces broadcast/multicast traffic in network

MAC Mobility

- Move of MACs between ethernet segments (and PEs)
- Usually handled by withdrawal and following advertisement
- Problem: new advertisement before old one was withdrawn
- MAC mobility extended community with sequence number
- New advertisement raises sequence number from the old one

Fast convergence and load balancing

- ▶ Multiple PEs connected to a segment
- ▶ Each announces EAD route and there are many MAC/IP advertisements
- ▶ Remote PEs can resolve any EAD route regardless of who advertised MAC
- ▶ When link on one PE fails just one EAD is withdrawn

Inclusive Multicast Ethernet Tag

- Represents BUM destination reachable through PE
- Announced for all VLANs in EVPN instance regardless of known MACs
- BUM frames are sent to all destinations known from IMET routes
- PMSI Tunnel attribute (RFC 6514) describes method of delivery
- Ingress replication or multiple variants of multicast encapsulation

Inclusive Multicast Ethernet Tag

```
+---------------------------------------+
|  RD (8 octets)                        |
+---------------------------------------+
|  Ethernet Tag ID (4 octets)           |
+---------------------------------------+
|  IP Address Length (1 octet)          |
+---------------------------------------+
|  Originating Router's IP Address      |
|           (4 or 16 octets)            |
+---------------------------------------+
```

Ethernet Segment Route

- ▶ Represents ethernet segment reachable through PE
- ▶ Used for autodetection of multi-homing setups
- ▶ Used for designated forwarder election
- ▶ Not used in single-homing setups

Ethernet Segment Route

```
+---------------------------------------+
|  RD (8 octets)                        |
+---------------------------------------+
|Ethernet Segment Identifier (10 octets)|
+---------------------------------------+
|  IP Address Length (1 octet)          |
+---------------------------------------+
|  Originating Router's IP Address      |
|          (4 or 16 octets)             |
+---------------------------------------+
```

Designated Forwarder Election

▶ Only one PE per ES should propagate BUM packets
▶ Otherwise they are duplicated
▶ Election based on a set of ES routes for given ES
▶ Non-elected PEs block egress BUM delivery