MVPN/EVPN C-Multicast/SMET Route Enhancements

Zhaohui Zhang, Robert Kebler
Wen Lin, Eric Rosen

Juniper Networks

96th IETF, Berlin

MVPN C-Multicast Route

- Used to disseminate customer multicast state across provider core
  - Contains (C-S/RP, C-G) information
  - Targeted at the ingress PE – except in MVPN-RPL method for C-Bidir support
    - RD is that of the VRF on the ingress PE (wrt C-S/RP)
    - RT makes sure that only the VRF on the ingress PE import the route
- Specifications about C-multicast route in RFC 6513/6514 have some issues
  - Inter-AS propagation
  - MVPN-RPL for C-Bidir
- Procedures for MVPN-RPL with selective tunnels could be optimized
  - With enhancements to C-Multicast route procedures

C-multicast Route Inter-AS Propagation

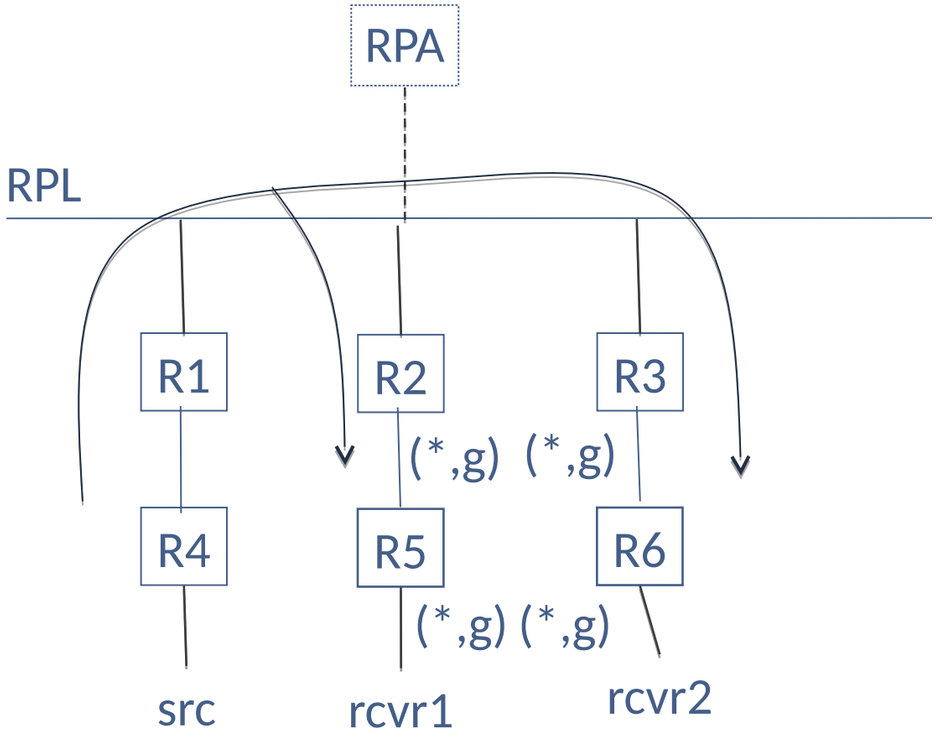
- Currently RFC6514 requires inter-as propagation through ASBRs
  - Along the reverse path of I-PMSI A-D route from the Ingress PE/AS
    - The required routes may not be available/desired in some deployments
  - Follows Option-B model for propagation
    - Regardless if forwarding is Option B (segmented) or Option C (non-segmented)
  - Built-in special procedures on ASBRs and egress PE
    - Could have used general BGP route propagation with RT Constraint but does not
  - Segmented tunnel case: requires Inter-AS I-PMSI A-D routes
    - RD of the Inter-AS I-PMSI A-D route for the source AS is used for C-Multicast route
    - Source AS number encoded in the NLRI
    - The RD and Source AS are used by ASBRs to locate Inter-AS I-PMSI A-D routes
  - Non-segmented tunnel case: requires Intra-AS I-PMSI A-D routes
    - Ingress PE’s address encoded in the Source AS field of C-Multicast route’s NLRI
      - Used by ASBRs to locate corresponding Intra-AS I-PMS A-D route
      - Does not work with IPv6 Infrastructure

C-multicast Route Inter-AS Propagation Enhancements

- Allow general BGP route propagation procedures
  - No need to go through ASBRs
    - No need for relevant complicated procedures
    - No need to set the Source AS field of C-multicast route's NLRI
  - RT Constraint achieves optimal propagation
- For existing Option-B based propagation in non-segmentation case
  - Allow any I/S-PMSI A-D routes from the ingress PE to be used
  - Uses RD alone of the C-multicast route to locate the I/S-PMSI A-D route
    - No need to encode Ingress PE's router ID into the Source AS field
      - Works fine with IPv6 infrastructure

PIM-Bidir and MVPN-RPL

- For PIM-Bidir, a Rendezvous Point Address (RPA) belongs to RP Link (RPL) but may not be tied to any router
- To receive traffic for a Bidir group routers sends join towards RPA, establishing a tree rooted at the RPA with branches rooted at the routers on the RPL
- Traffic is sent along the tree bi-directionally. When upstream traffic (towards the RPA) reaches a router on the RPL, it is dumped on the RPL, picked up by others, and sent downstream on other branches rooted at those other routers.
- DF election required on transit LANs but not on RPL
- MVPN-RPL: VPN Backbone as C-Bidir RPL
  - PEs are routers on the RPL
  - Avoids DF election over the provider core
    - VPN backbone is essentially a virtual LAN



MVPN-RPL: VPN Backbone as C-Bidir RPL

- Traffic received from PE-CE interface needs to be sent across the backbone (RPL)
  - If another PE has corresponding (C-*,C-G-Bidir) state
    - As indicated by the existence of C-multicast routes that are distributed to all PEs
  - By default, inclusive tunnel is used to send to all PEs
- Selective tunnel can be used
  - Current procedures require S-PMSI AD routes for all tunnel types, plus Leaf AD routes for RSVP/IR/BIER tunnel types
    - PMSI: Provider Multicast Service Interface, a conceptual interface for a PE to send customer traffic to all or some PEs
  - Any ingress PE (receiving traffic from CE and sending to the core) need to advertise S-PMSI AD
  - Any egress PE with corresponding (C-*,C-G-Bidir) state needs to send Leaf AD in response to S-PMSI AD incase of RSVP/IR/BIER
    - Leaf AD serves Explicit Tracking purpose
    - N S-PMSI and N^2 Leaf AD routes in the worst case

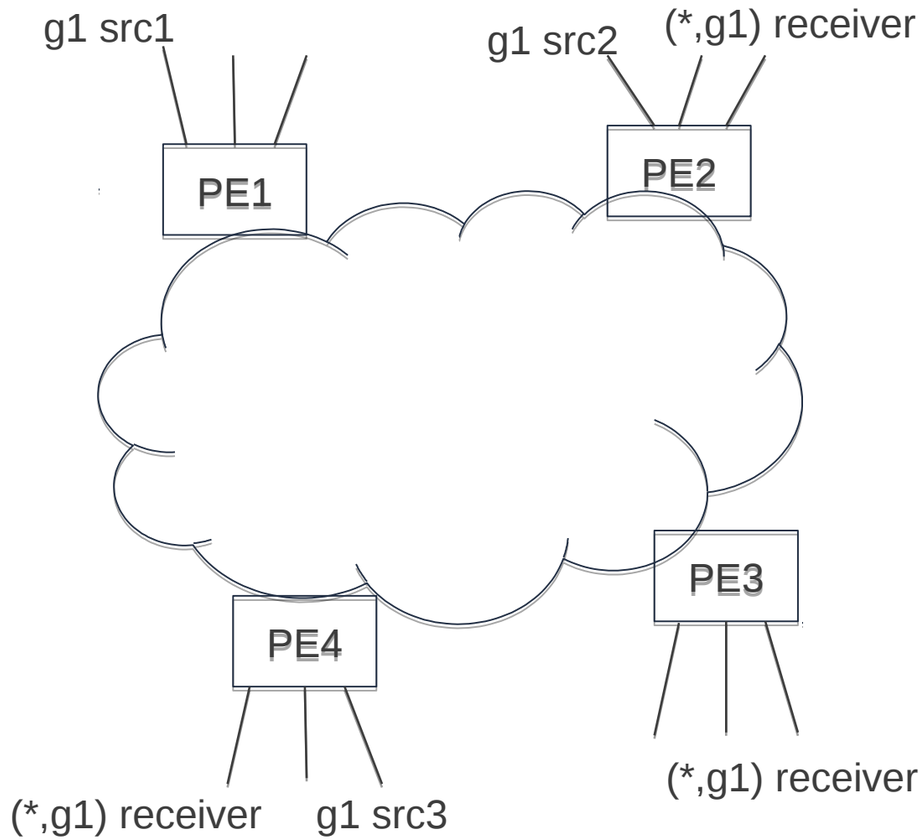
Optimizations for MVPN-RPL with Selective Tunnels

- RSVP/IR/BIER: no need for Leaf A-D routes
  - C-multicast routes can already do explicit tracking
    - Each carries the RD of the originating VRF so RRs will reflect all
    - Untargeted, explicit-tracking C-multicast routes
      - » Can also be viewed as unsolicited, untargeted Leaf A-D routes
- IR/BIER
  - No need for S-PMSI either – no need to announce the tunnel
- PIM/mLDP: no need for explicit tracking
  - A common RD (per VPN) could be used for all PEs
    - Reduces the number of routes that each PE keeps
      - A RR does not reflect every path of the same (C-*,C-G-Bidir) C-multicast route
    - BGP ADD-PATH needed
      - Up to two paths needs to be reflected by a RR

EVPN SMET Routes

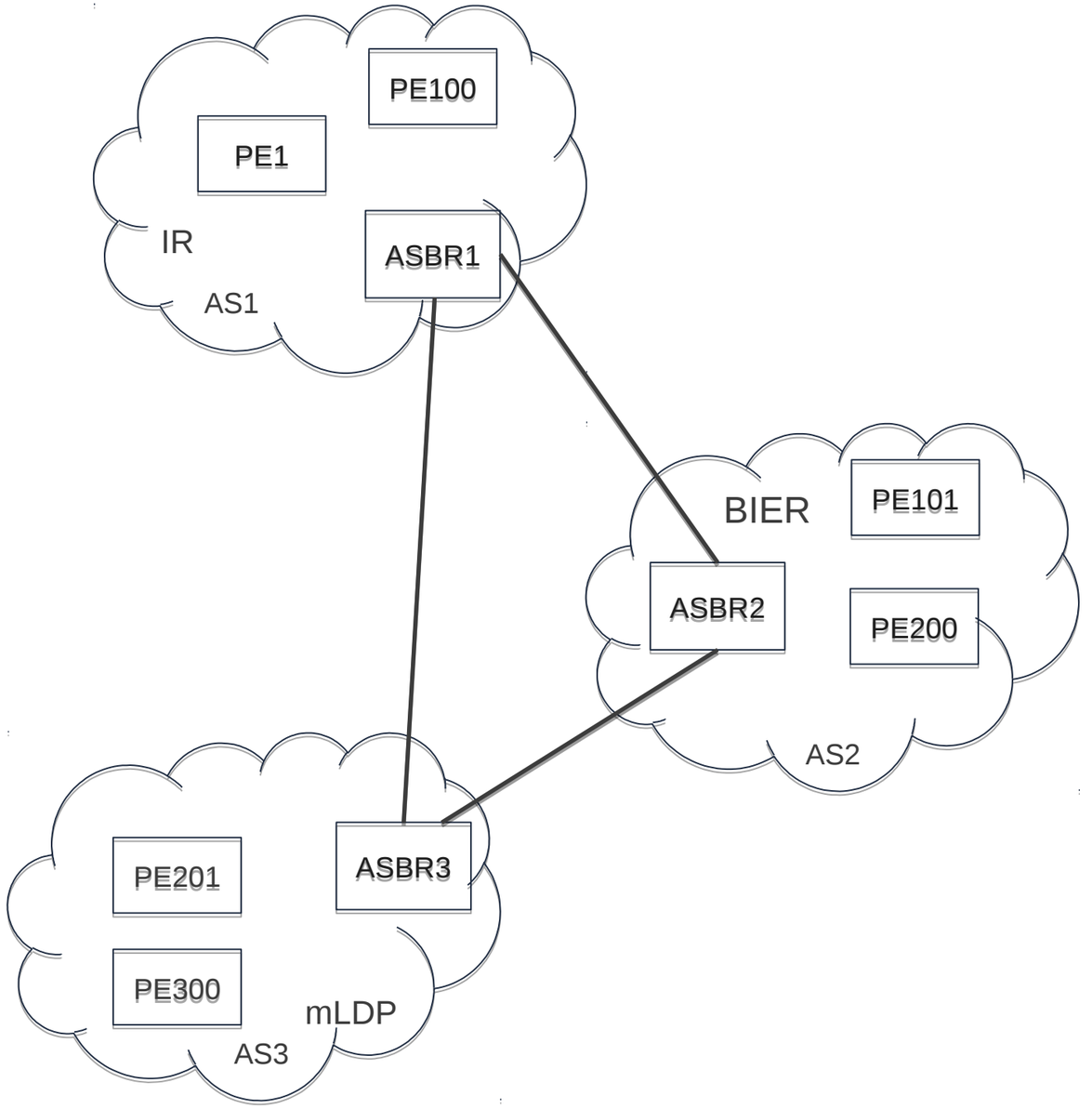
Selective Multicast: draft-sajassi-bess-evpn-igmp-mld-proxy

- An EVPN Bridge Domain simulates a LAN
  - Hosts on the LAN may send multicast traffic for certain groups
  - Some hosts may be interested in receiving traffic for some groups
    - IGMP/MLD used to signal the interest
- A PE snoops IGMP/MLD joins on PE-CE interfaces and generate (C-S/*,C-G) Selective Multicast Ethernet Tag (SMET) BGP routes
  - Sent to all other PEs (senders could be every where)
  - Other PEs won't send traffic to this PE unless corresponding SMET route has been received from this PE
- EVPN SMET route is very similar to the untargeted explicit-tracking MVPN C-multicast route
  - Current draft assumes IR/BIER in the core
    - No need for S-PMSI/Leaf A-D procedure
  - Same optimization for PIM/mLDP/RSVP selective tunnel as in MVPN-RPL case



Provider Tunnel Segmentation

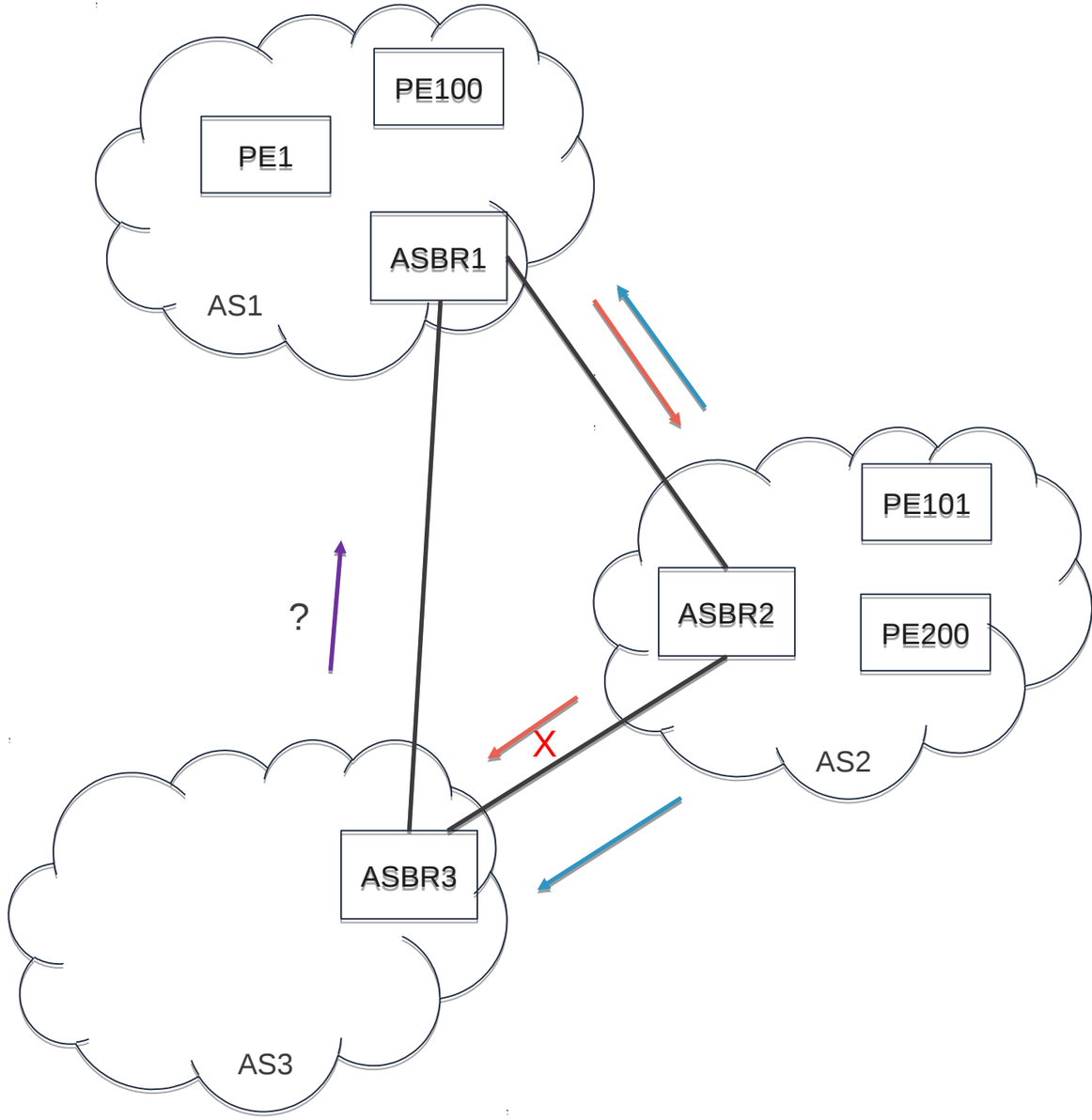
- Provider tunnel segmentation is often used to:
  - Allow different tunnels (of same or different types) in different AS
  - Aggregate many individual PE-PE tunnels to tunnels at AS level
  - Restrict per-PE PMSI/Leaf routes to the same AS
    - Only per-AS tunnels and corresponding routes across inter-as links
- Achieved by PMSI/Leaf route procedures
- Untargeted explicit-tracking C-multicast routes introduce challenges to segmentation
  - This applies to both MVPN and EVPN



Challenge 1: Route Aggregation & Propagation

Animation In Use

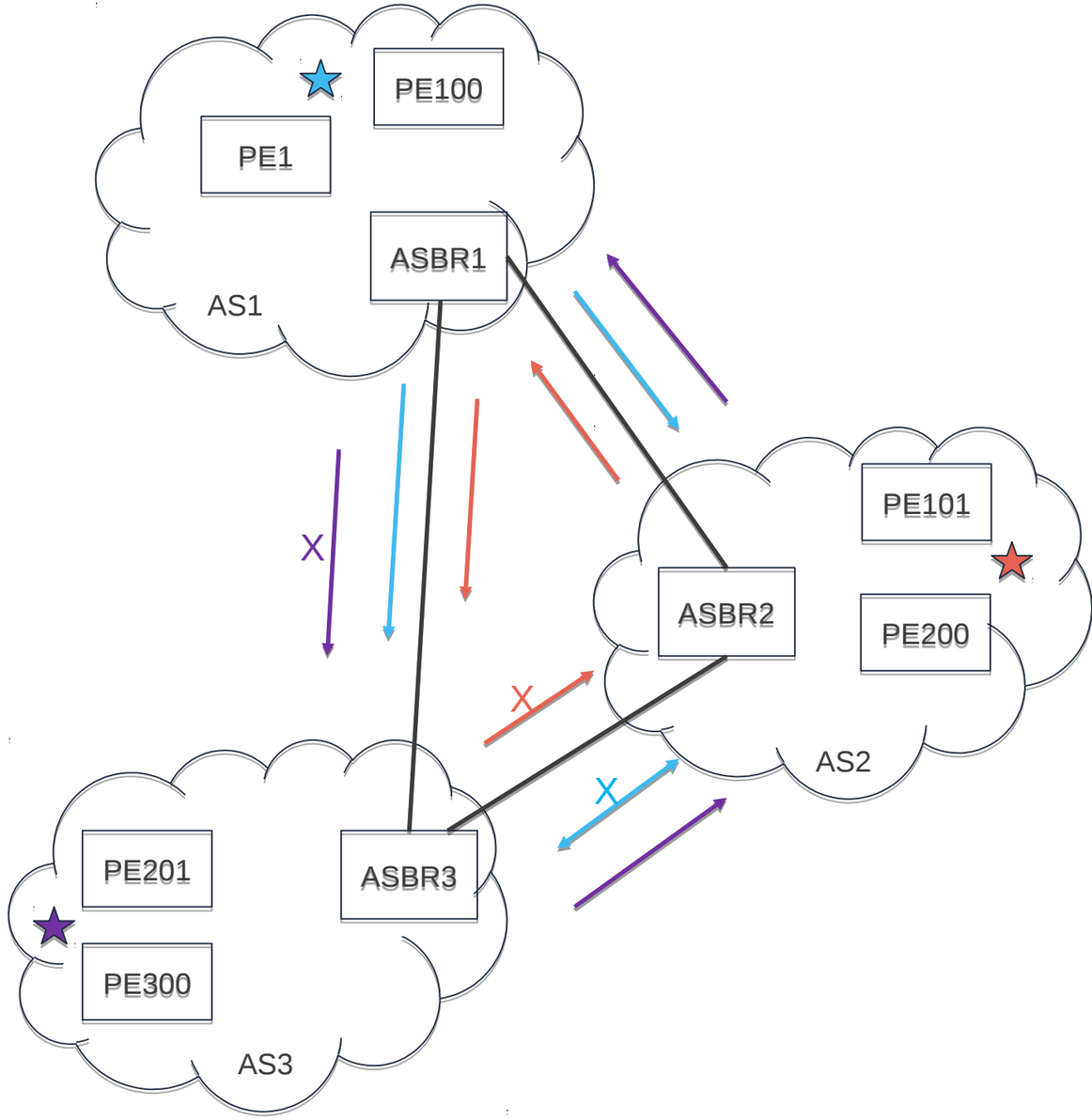
- PE1 ~ PE100 in AS1 originates 100 (*,g1) C-multicast routes; ASBR1 should aggregate those into a single one and send to AS2.
- PE101 ~ PE200 in AS2 originates 100 (*,g1) C-multicast routes; ASBR2 should aggregate those into a single one and send to AS1 & AS3.
- The aggregate one from ASBR1 should not be propagated into AS3
  - Absorbed into the one from ASBR2
  - For traffic from AS3, ASBR3 should only send one copy to ASBR2, who will forward to PE101~200 and ASBR1
- If there is an ASBR1-ASBR3 connection, should ASBR3 send a (*,g1) route to ASBR1?
  - If it's sent, ASBR3 will get AS1 traffic and send to ASBR2, who will forward duplicates to PE101~200
  - If it's not sent, and the ASBR1-ASBR2 connection is gone, then ASBR2 will not get any traffic from AS1



Challenge 2: Traffic forwarding

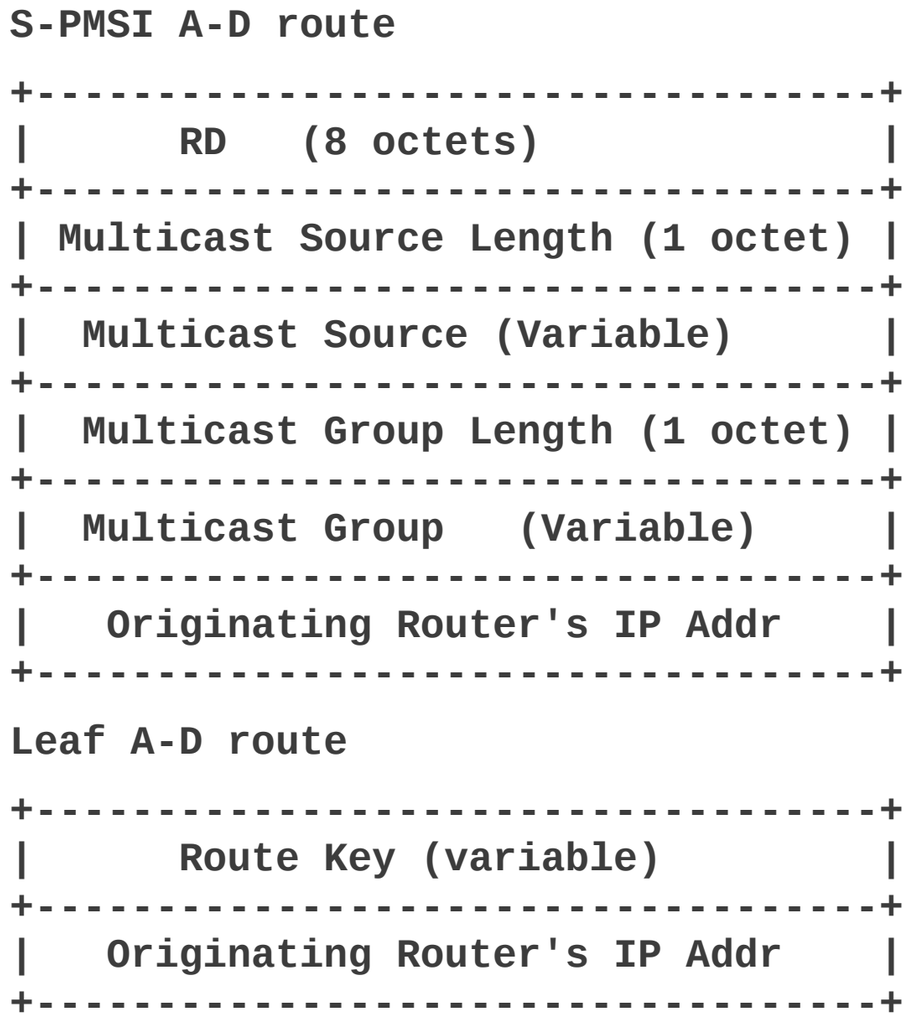
Animation In Use

- Multicast traffic forwarding must follow rooted trees
- W/o segmentation, a tree is rooted at an ingress PE with leaves being all other PEs that need to receive traffic
- W/ segmentation, an inter-as tree is rooted at the source AS, with branches beginning with ASBRs in the source AS and extending to other ASBRs along the way
- All 300 PEs need to receive (*,g1) traffic
  - Traffic from AS1 sent to ASBR2 and ASBR3, and they should not forward to each other
  - Traffic from AS2 may be sent to ASBR1 and then forwarded to ASBR3, who should not forward to ASBR2
  - Traffic from AS3 may be sent to ASBR2 and then forwarded to ASBR1, who should not forward to ASBR3



S-PMSI/Leaf A-D Route

- PMSI/Leaf A-D procedures handle the challenges very well
  - Provider Multicast Service Interface
    - A conceptual interface for a PE to send customer multicast traffic to all or some PEs
- Per RFC 6514 (MVPN)
  - A Leaf route is always generated in response to an I/S-PMSI route
  - A Leaf route's NLRI includes:
    - Route Key – corresponding PMSI route's NLRI
      - Including Originating Router's IP Addr (ingress-id)
    - Originating Router's IP Addr (egress-id)
  - A Leaf route carries a RT corresponding to either the ingress-id or the upstream ASBR
- Draft-zzhang-bess-evpn-bum-procedure-updates extends this to EVPN



Segmentation w/ Untargeted Explicit-tracking C-multicast Route: Inter-as Example

ASBRs turns them into targeted Leaf A-D routes

- PEs advertise untargeted explicit-tracking C-multicast/SMET routes if they have local receivers
- ASBRs in the local AS do not re-advertise those to other ASes
- They pretend they have received a corresponding S-PMSI route from an ASBR in each remote AS
  - Corresponding Leaf AD routes are generated and propagated upstream per existing procedures, only that the S-PMSI route is imaginary/fabricated
    - RD, Tag & Originator ID are from the active per-AS I-PMSI route for the remote AS
    - Source/Group are from received C-Multicast/SMET route
  - This builds an inter-as tree rooted at each AS
    - Different tunnel types can be used for different segments

EVPN S-PMSI A-D route

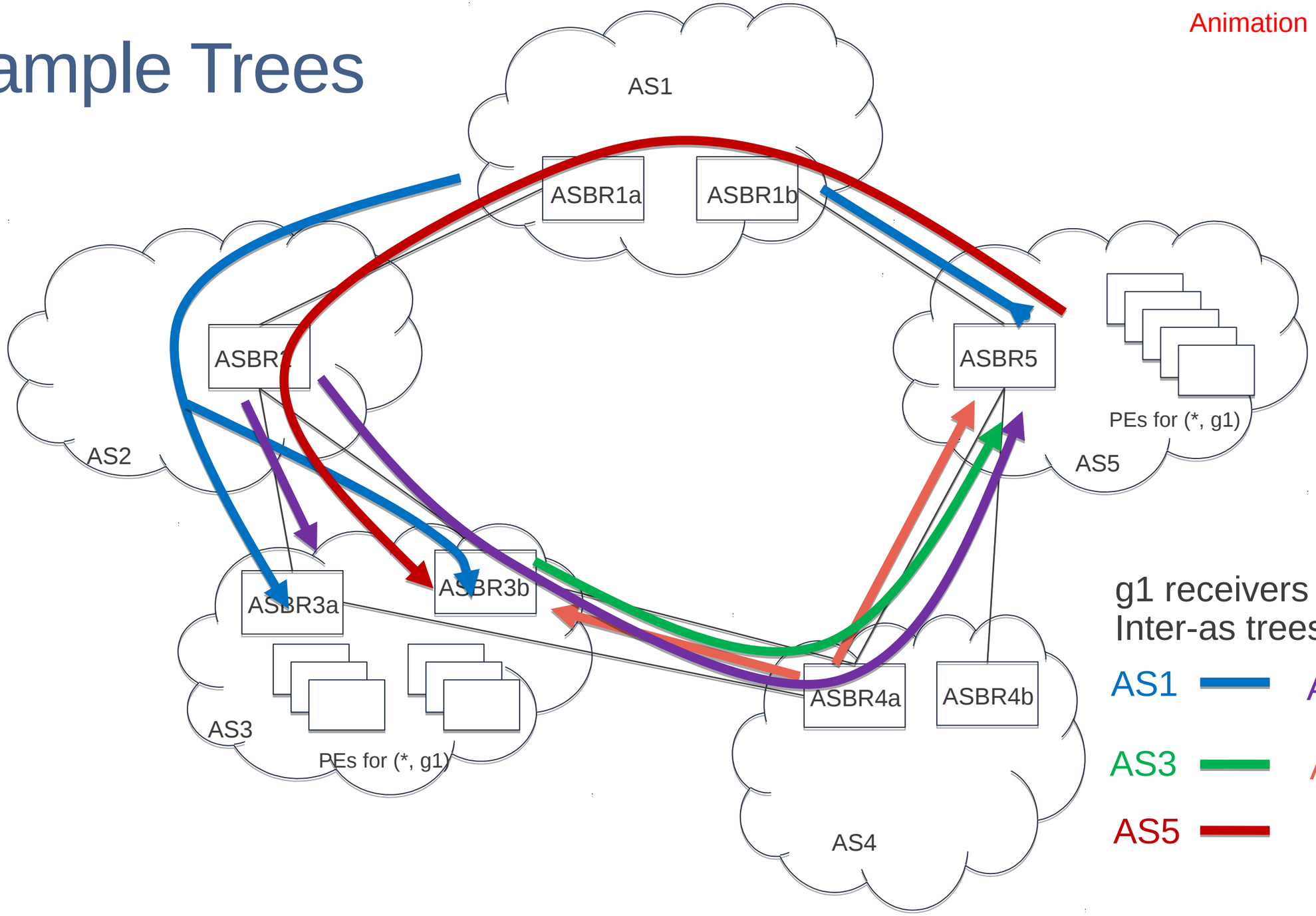
```
+-----------------------------------+
|      RD   (8 octets)              |
+-----------------------------------+
|  Ethernet Tag ID (4 octets)       |
+-----------------------------------+
| Multicast Source Length (1 octet) |
+-----------------------------------+
|  Multicast Source (Variable)      |
+-----------------------------------+
|  Multicast Group Length (1 octet) |
+-----------------------------------+
|  Multicast Group   (Variable)     |
+-----------------------------------+
|   Originating Router's IP Addr    |
+-----------------------------------+
```

Leaf A-D route

```
+-----------------------------------+
|      Route Key (variable)         |
+-----------------------------------+
|   Originating Router's IP Addr    |
+-----------------------------------+
```

Example Trees

Animation In Use



g1 receivers in AS3 & 5.
Inter-as trees rooted at:

AS1 AS2

AS3 AS4

AS5

Summary

- MVPN C-multicast routes are used to disseminate customer multicast state across provider core
  - Inter-AS propagation procedures are updated
  - MVPN-RPL selective tunnel procedures are optimized
- EVPN SMET routes are very similar to MVPN-RPL's C-multicast routes
  - Above mentioned optimizations for MVPN-RPL are either:
    - Already the specified behavior for EVPN SMET routes, e.g. Explicit Tracking
    - Or could be applied to EVPN SMET routes, e.g. when Explicit Tracking is not needed
- Common segmentation procedures are proposed for both MVPN-RPL C-multicast routes and EVPN SMET routes.