

EVPN BUM Procedures Update

Jeffrey Zhang, Wen Lin

Jorge Rabadan, Keyur Patel

IETF 93, Prague

EVPN BUM Procedures

- RFC 7432 (EVPN) refers to RFC 7117 (VPLS Multicast) for quite some EVPN BUM procedures
 - RFC 7432 mentions selective tree for Ingress Replication and refer to RFC 7117
 - The same could apply to P2MP tunnels
 - RFC 7432 refers to RFC 7117 for P2MP inclusive tree
- RFC 7117 includes inter-as segmentation procedures
 - With complicated and VPLS-specific rules for inclusive tree
 - RFC 6514 (MVPN) has slightly different procedures for inter-as inclusive trees
 - Good to adopt MVPN instead of VPLS procedures for inter-as inclusive trees
- RFC 7524 (Inter-area segmentation) covers MVPN and VPLS
- Goals of this draft: clarifications/updates/extensions for EVPN BUM procedures
 - New EVPN route types for selective tree
 - Updated inter-as segmentation procedure for EVPN
 - Extending inter-area segmentation to cover EVPN and support inter/intra-region

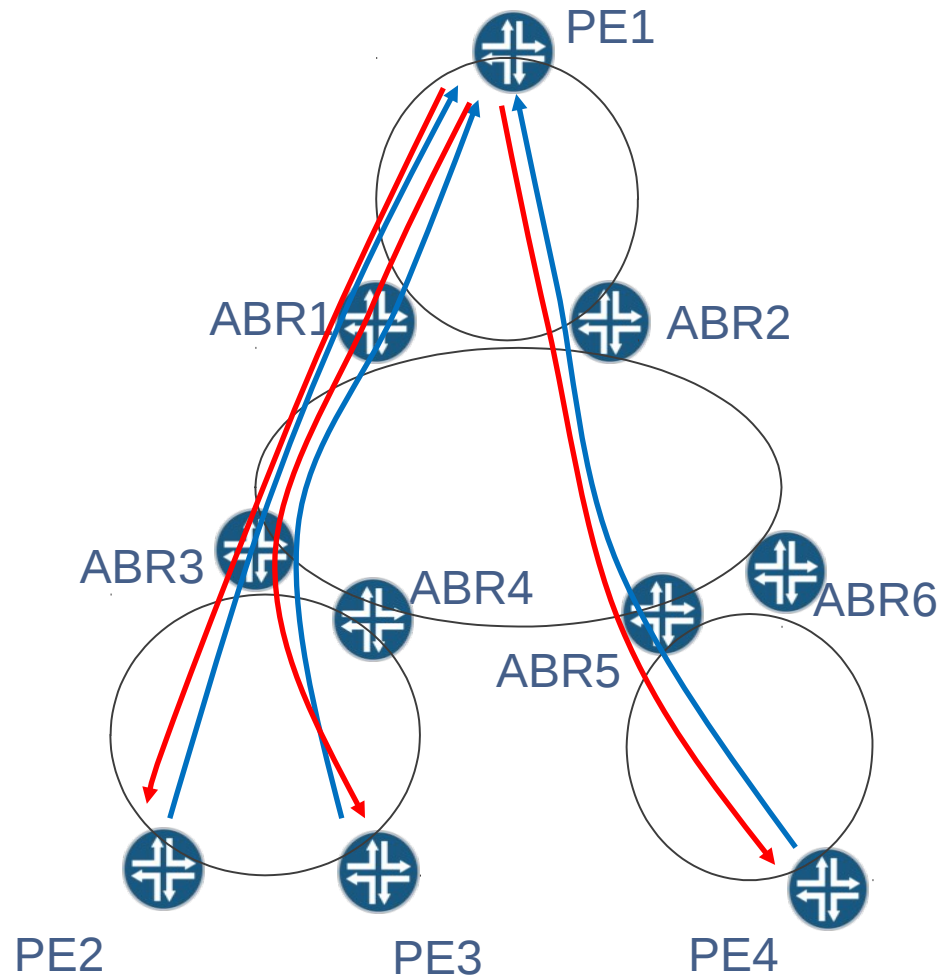
Selective Tree

- Only need to define S-PMSI and Leaf A-D routes for EVPN
 - All existing procedures in RFC 7117 apply
- An S-PMSI route announces that an ingress PE intends to send certain multicast flow on a specific tunnel (vs. the inclusive tunnel)
- An egress PE that needs to receive the traffic joins the specific tunnel
 - By sending mLDP label mapping or PIM join for the tunnel, or,
 - By letting ingress know explicitly about itself via Leaf A-D routes
 - This is called leaf-tracking
 - Needed for Ingress Replication, RSVP-TE P2MP tunnel, or BIER transportation
 - A Leaf A-D route is specifically in response to a particular PMSI route
- The same concept is also used for tunnel segmentation
 - Even for inclusive tree

Motivations for tunnel segmentation

- Segmentation may be necessary for Inter-AS/provider deployments
 - One AS/provider may use mLDP while the other uses RSVP-TE P2MP
 - Even if RSVP-TE P2MP can be used in both AS/providers, tunnel signaling may have to be restricted to individual AS/providers and not across
 - Technical or administrative reasons
- Segmentation may be desired in certain situations
 - Even if it's not absolutely necessary
 - Examples:
 - Different areas using different tunnel types
 - Tunnel aggregation in smaller areas for better congruency
 - Smaller per-area BIER sub-domains (hence shorter BitString)
 - Same tunnel type in multiple AS/areas but limit the Leaf A-D routes to single AS/area
 - IR in multiple AS/areas but limit replication to single AS/area and let ASBRs/ABRs relay the replication

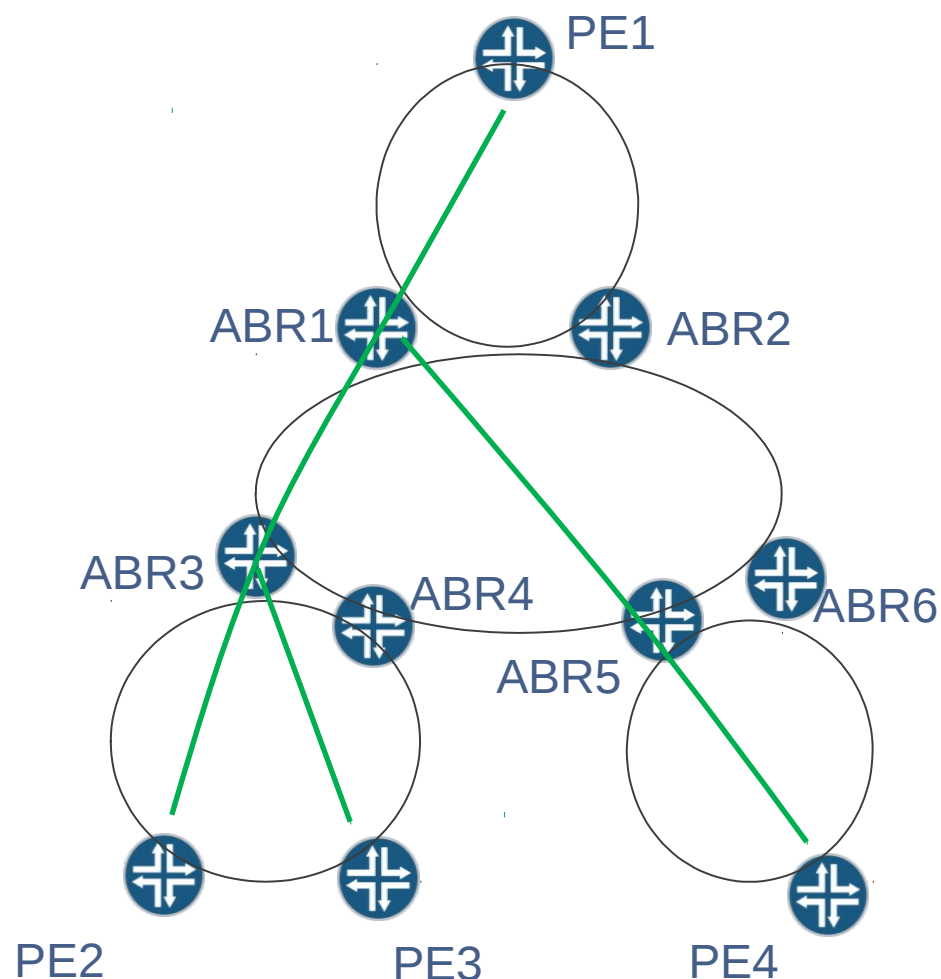
Leaf Tracking & Fan-out for Ingress Replication



w/o segmentation

PE1 tracks three egress PEs

PE1 sends three copies



w/ segmentation

PE1 tracks one downstream ABR1
ABR1 tracks two downstream ABR3 & 5
ABR3 tracks two egress PEs
ABR5 tracks one egress PE

Reduced control plane overhead on ingress PEs

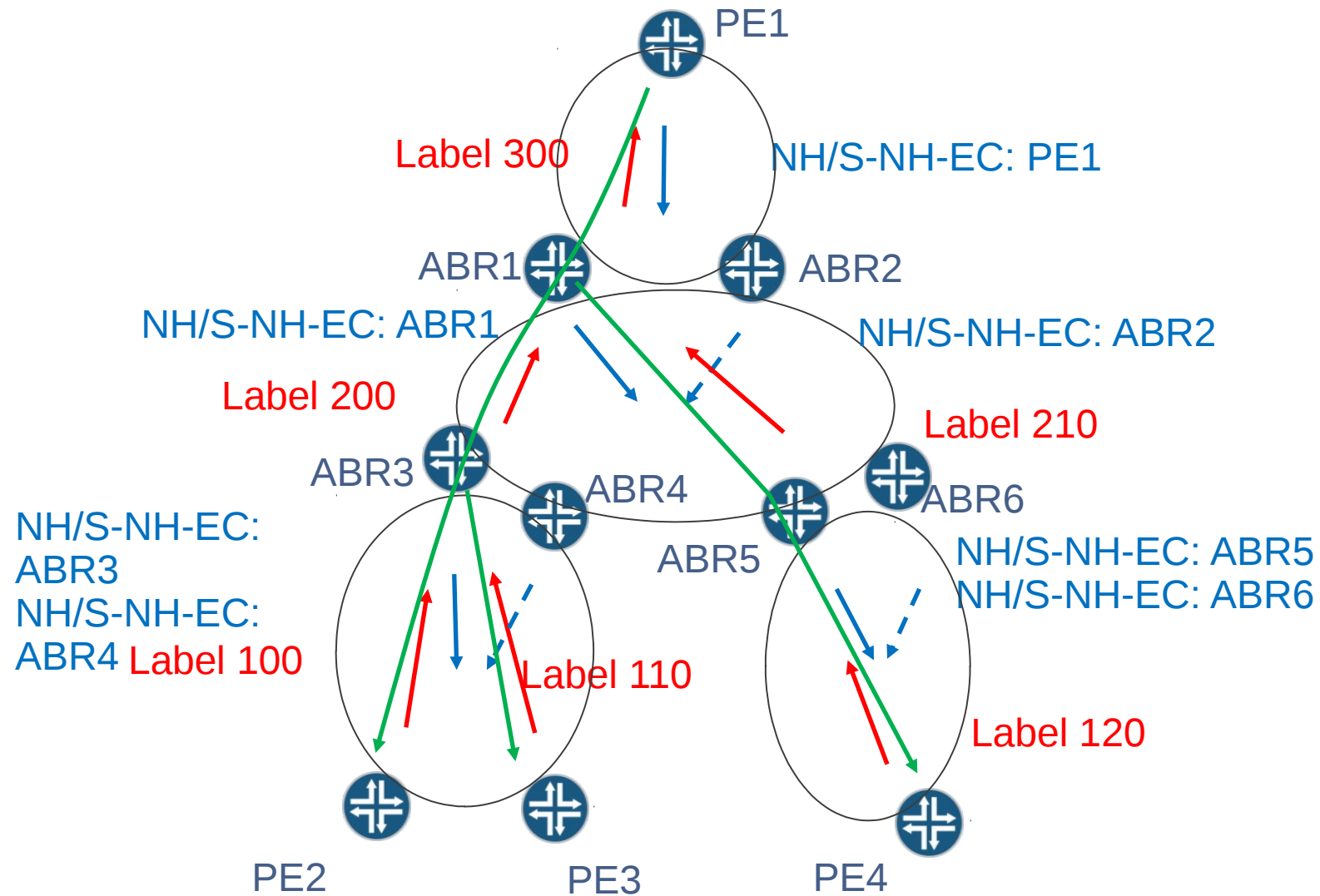
PE1 sends one copy
ABR1 sends two copies to ABR3 & 5
ABR3 sends two copies to PE2 & 3
ABR5 sends one copy to PE4

Reduced bandwidth usage and replication burden

Tunnel Segmentation Procedures

- Segmentation points are ASBRs/ABRs
 - RFC 6513/6514: MVPN Inter-AS segmentation
 - RFC 7117: VPLS Inter-AS segmentation
 - RFC 7524 (Seamless MPLS Multicast for MVPN/VPLS): Inter-Area Segmentation
- When an ASBR/ABR re-advertises a PMSI/IMR A-D route:
 - It changes the PTA to specify a new tunnel for the downstream segment
 - ASBR and EVPN ABR: changes BGP next hop to its own address
 - MVPN/VPLS ABR: changes S-NH-EC to specify its own address
 - Segmented Next Hop Extended Community
- A downstream ASBR/ABR/PE joins that tunnel segment
 - PIM join or mLDP label mapping, or
 - Leaf A-D routes towards the upstream ASBR/ABR
 - Either the BGP next hop or specified in S-NH-EC of corresponding PMSI/IMR A-D route
 - The upstream ASBR/ABR further joins its upstream tunnel segment
 - And stitches the upstream segment to the downstream segment in forwarding path

Inter-Area IR Segmentation Example



- ABR3: L200 -> (L100, base LSP to PE2)
(L110, base LSP to PE3)
- ABR6: L210 -> (L120, base LSP to PE4)
- ABR1: L300 -> (L200, base LSP to ABR3)
(L210, base LSP to ABR6)
- PE1: (S,G) -> (L300, base LSP to ABR1)

- PMSI A-D to all RR clients
- Dashed lines for those not preferred by receivers
- Leaf A-D
- To upstream ABR/PE specified in NH or S-NH-EC of PMSI A-D

Inter-Area Segmentation Extensions

- Inter-area → Inter-region
- While RFC 7524 is based on areas, the area/ABR can be replaced by region/RBR
 - A region can be a subset of an area, defined by a BGP peer group
 - A region can even be an entire AS plus its external link
- Inter-region → intra-region
 - With inter-region, a RBR changes next hop and tunnel type/id when it re-advertise a PMSI/IMR route into other regions, and stitches the segments in different regions together
 - With intra-region, a RBR could do the same even when it re-advertise into the same region, and stitches the segments together
 - Use cases include assisted Ingress Replication and Virtual Hub and Spoke

Inter-AS Inclusive Tunnel Segmentation

- VPLS requires that only one ASBR re-advertises a VPLS A-D route into an AS
 - That requires complicated and VPLS-specific election procedures that do not apply to EVPN
- This document proposes for EVPN to follow generic MVPN inter-as segmentation procedures
 - Any ASBR can re-advertise IMR routes into its AS, and the egress PEs or downstream ASBRs will accept traffic from their own choice of upstream ASBR – as in the selective tree case and inter-area case
 - ASBRs could also consume and aggregate individual per-PE IMR routes into per-AS IMR routes that they originate
 - This achieves better scaling and backwards compatibility

Summary and Next Steps

- Clarified/Updated/Extended EVPN BUM procedures
 - New route types for selective tree
 - Updated inter-as inclusive tunnel segmentation
 - Extended inter/intra-region segmentation
- Next steps
 - Seek and address comments from the community
 - Add more details
 - Current revision focuses on discussing the ideas and concepts
 - Will seek WG adoption afterwards