# Segmented MVPN Using IP Lookup for BIER

# draft-xie-bier-mvpn-segmented-06

Jingrong Xie @Huawei
Liang Geng @China Mobile
Lei Wang @China Mobile
Mike McBride @Huawei
Gang Yan @Huawei

# Update States

- Update from -01 rev (ietf102) to -06 rev(now).
  - the term 'BIER tunnel', 'P2MP tunnel' and 'BGP-MVPN FEC'.
    - One more thing: “Pseudo VRF” on segmented point (ABR).
    - Pseudo VRF is comparable to the BGP-MVPN FEC(RootIP, RD, *,*).
    - Many BIER tunnels share the same(RootIP, RD) will be mapped to the same Pseudo VRF.
  - tunnel stitching can be between any two of mLDP/RSVP-TE/IR/BIER.
  - e2e stitched tunnel can be bound to one or many 'BGP-MVPN FEC(s)' from some IngressPE and VRF, and Ingress PE can decide to use which tunnel for which flow(s).
    - Possiblely use an BIER tunnel bound to FEC(RD,*,*), and switch to the BIER tunnel bound to an FEC(RD,S,G), but the IP Lookup followed ensures the replication optimized on the BIER segment !
  - the per-tunnel stitching of upstream tunnel and downstream tunnels, and the per-flow IP lookup for downstream BIER encapsulation(BitString), are separated/decoupled.
  - BIER-BIER, BIER-P2MP, P2MP-BIER all covered after ietf102.

# Comments addressed

- Comments-0:
  - Initial comments came from Eric when I discussed with authors of <draft-ietf-bier-mvpn> before the -00 rev of this draft.
  - The exploring of IP-lookup for BIER was lost in <draft-ietf-bier-mvpn-04>.

When segmented P-tunnels are being used, a BFER that receives a BIER-encapsulated MVPN multicast data packet may need to be forwarded on its next P-tunnel segment. The choice of the next P-tunnel segment for the packet depends upon the C-flow to which the packet belongs. Since the BFIR assigns a distinct upstream-assigned MPLS label for each C-flow, the BFER can select the proper "next P-tunnel segment" for a given packet simply by looking up the upstream-assigned label that immediately follows the BIER header. (If the BFIR had not assigned a distinct label to each C-flow, the BFER would need to maintain all the state from the Multicast Flow Overlay in order to select the next P-tunnel segment.)

When segmented P-tunnels are being used, a BFER that receives a BIER-encapsulated MVPN multicast data packet may need to be forwarded on its next P-tunnel segment. The choice of the next P-tunnel segment for the packet depends upon the C-flow to which the packet belongs. As long as the BFIR has assigned the MPLS label according to the constraints specified in Section 2.1, the BFIR will have assigned distinct upstream-assigned MPLS labels to distinct C-flows. The BFER can thus select the proper "next P-tunnel segment" for a given packet simply by looking up the upstream-assigned label that immediately follows the BIER header.

  - The join-latency was added as ‘more efficient’ in <draft-ietf-mvpn-expl-tracking>.

This document clarifies the procedures for originating and receiving S-PMSI A-D routes and Leaf A-D routes. This document also adds new procedures to allow more efficient explicit tracking. The procedures being clarified and/or extended are discussed in multiple places in the documents being updated.

  - Authors agree with the text very much, and think it worth to leverage the ‘more efficient’ tracking for BIER in segmented deployment.

# Comments addressed (cont)

- Comment-1: Is this problem a BIER-specific or a more generic one ?
  - GTM using IR has a very similar opinion to use ‘positive join’ Leaf-AD from downstream.
  - GTM using IR has a very similar opinion to use IP lookup when aggregating flows using one Label.
  - Very similar to the proposal of this draft to BIER: ‘positive join’ initiated by LIRpF, aggregating flows using one label, and per-flow replication without flooding using IP lookup.
  - And the RFC7988/7524 has text on this very well.
  - Authors think this draft is BIER-specific, and the BIER WG is the right place to discuss. It is all about: BIER-specific Segmented MVPN, leveraging the LIR-pF, and then use IP lookup for per-flow forwarding on BIER segment (sub-domain).

# Comments addressed (cont 2)

- Comment-2: How about other opinions for a better join latency ?
  - LIR-pF can initiate ‘positive join’ of per-flow Leaf A-D route, and optimize the Join latency.
  - Per-flow VpnLabel allocation will require a round-trip of SPMSI A-D and the acked Leaf AD.
    - Existing code may use a data-driven SPMSI A-D route advertising according to RFC6513. This can make the join latency even longer.
    - One opinion is to send SPMSI A-D route ASAP when being aware of any (S,G) state, for example, receive C-multicast join(S,G), or any form of Soure-Active(S,G) information.
    - One opinion is to use BIER in an I-PMSI manner ( flooding manner ) temporarily.
    - GTM using IR has a mechanism special to get better join latency: ‘positive join’ Leaf A-D route from downstream routers. This can also be solicited by using the LIR-pF explicit-tracking.

# Comments addressed (cont 3)

- Comment-3: pros and cons of this proposal
  - **Overview:**
    - IP Lookup: VpnLabel for Pseudo-VRF, (Pseudo-VRF, C-SA, C-DA) for BitString.
    - w/o IP Lookup: VpnLabel for BitString directly, allocating of Vpnlabel for every C-flow before.
  - **Forwarding cycles:**
    - IP Lookup will need more forwarding cycles.
  - **Forwarding table rows and width:**
    - IP Lookup will need T + N states.
    - T represent the stitching tunnels, and N represent the number of flows.
    - w/o IP Lookup will need N states.
    - T can be 1 in some case, and the N states of MFIB/MFIB6 is wider than VpnLabel.
    - T can be N in some case, which cost even more.
  - **IP lookup is the help and cost to leverage the more 'efficient' LIR-pF explicit-tracking.**
    - Positively initiate the C-multicast(S,G) join from receiver sites.

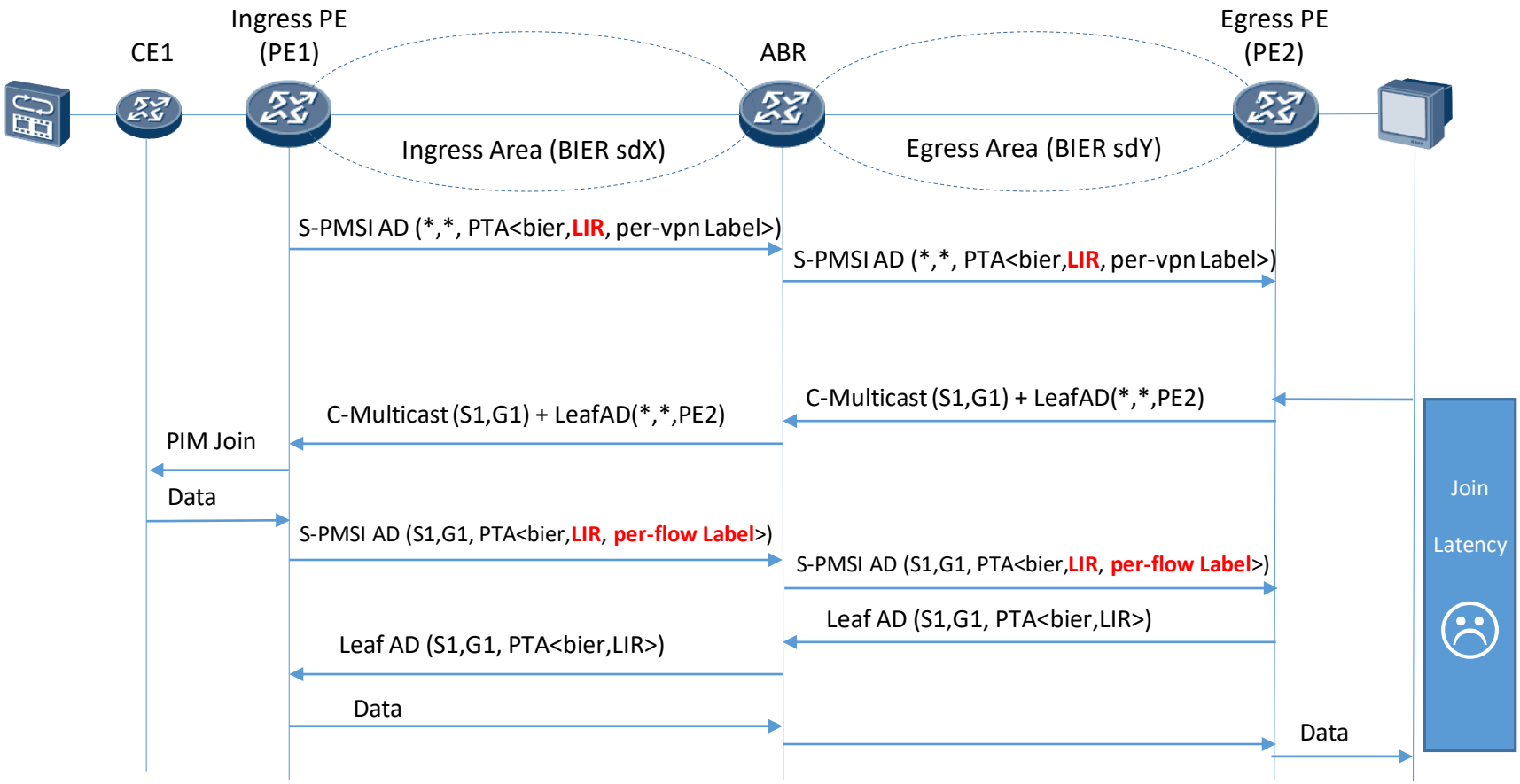
# Comments and opinions on Adoption

- Any comments ?
- Do you think it worth for adoption ?
- Backup slides of ietf102 followed below, with updated text from the today's view.

Thanks !

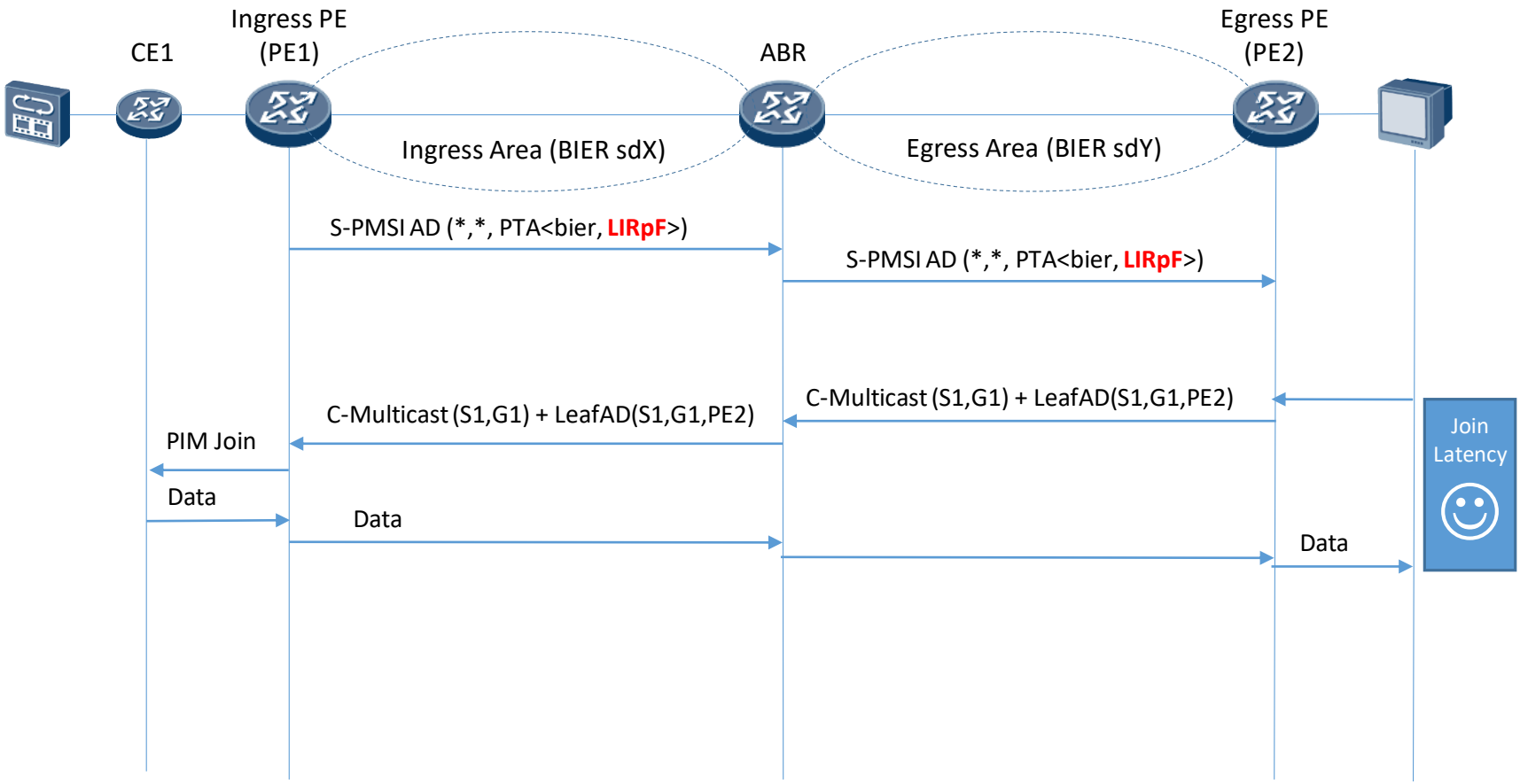
# Backup slides of ietf102

# LIR explicit-tracking for Segmented BIER



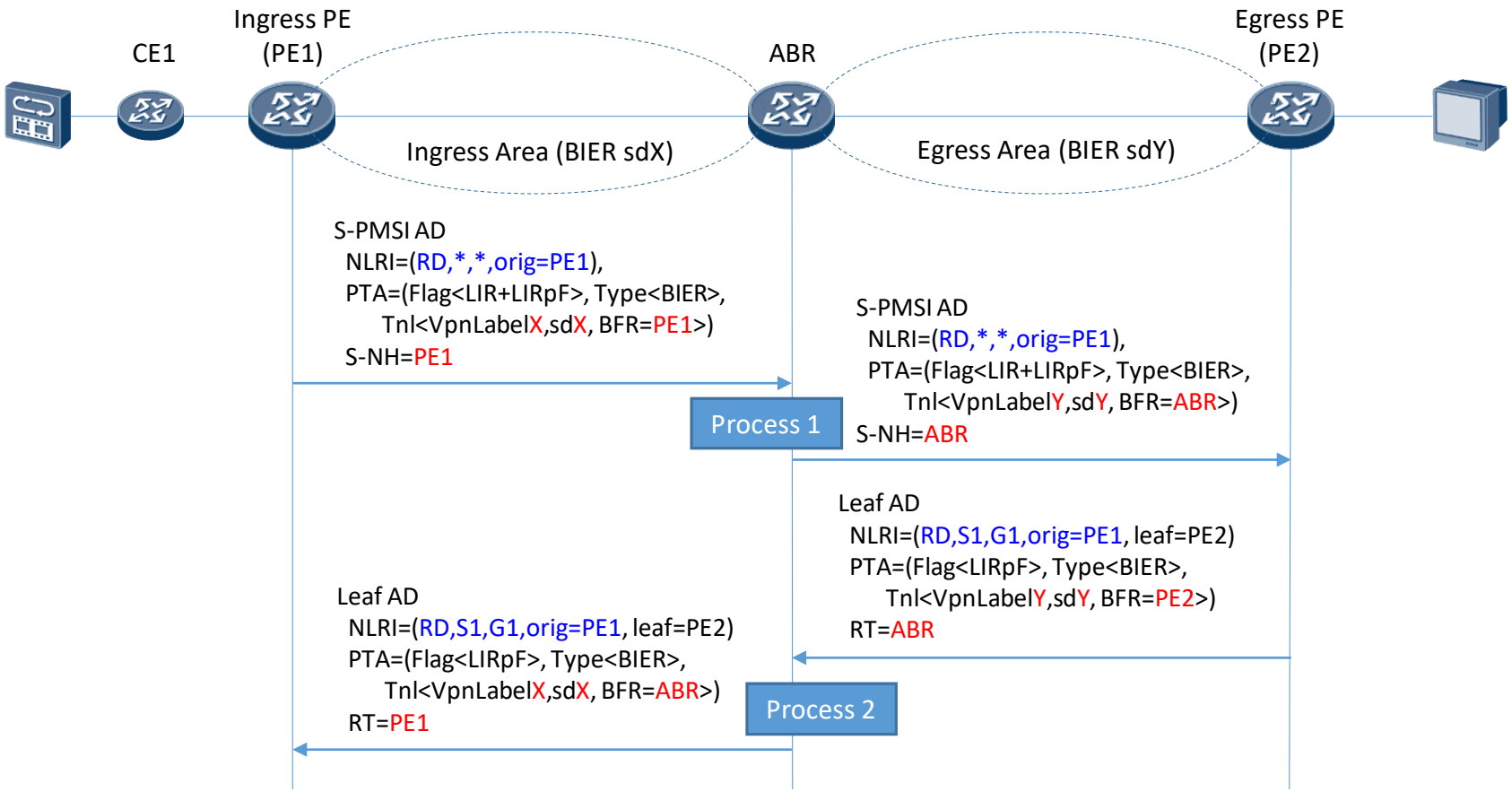

- Besides, the SPMSI(S,G) routes are 'flooded' to routers that even don't want.
- Ietf103: per-flow (S,G) need the IngressPE to initiate, possibly data-driven.

# LIR-pF explicit-tracking for Segmented BIER




- Accordingly, the unwanted SPMSI(S,G) routes are eliminated.
- The same benefit as Un-Segmented BIER MVPN.
- Ietf103: per-flow (S,G) 'positive join' from EgressPE using Leaf AD routes, once LIR-pF kicked-off.

# Control Plane Process on ABR




Process 1: per-vpn info: FEC=(RD,PE1), upstream (X, PE1), downstream(Y, ABR)

Process 2: per-flow info: FEC=(RD,PE1,S1,G1), upstream(X,PE1), downstream(Y,ABR,PE2)

# Per-flow state building

Process 2 : Build the control-plane state for per-flow on ABR

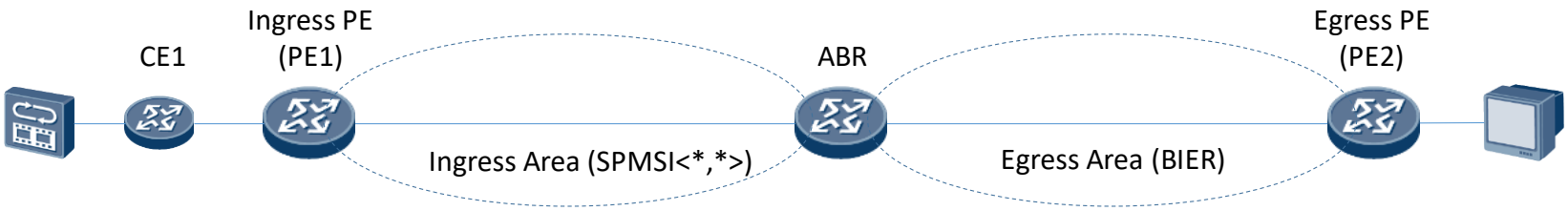
- ABR receive Leaf-AD, and form the **downstream state**: <RD, S1, G1, PE1> <Leaf=PE2> (Attrs=BFR-id)
- ABR send Leaf-AD to PE1, and form the **upstream state**: <RD, S1, G1, PE1> (umh=PE1)
- It is very similar to the PIM Join, or mLDP Mapping, which build a state driven by downstream join.
- Control-plane always need keeping a Per-flow state <RD, S1, G1, PE1>, including **upstream** and **downstream(s)** parts.
- The <RD, S1, G1, PE1> is an Per-flow FEC (I'd like to call it a BIER-FEC like RFC7524).
- The <RD, PE1> is an implicit VRF identifier for an ABR (call Per-vpn FEC).
- Ietf103: the BIER-FEC is changed to BGP-MVPN FEC.

# Per-flow state for forwarding

Process 2: Build the forwarding state for per-flow on ABR

- **upstream state**: <RD, S1, G1, PE1> (umh=PE1)
- **downstream state**: <RD, S1, G1, PE1> <Leaf=PE2> (Attrs=BFR-id)
- Do a mapping of <RD,PE1> to <virtual VRF identifier> **locally on ABR**, then
  - Disposition Process : (BIER Label<of sdX>, BFIR-id<PE1>, VpnLabelX, virtual-VRF-identifier)
  - Re-Imposition Process : (virtual-VRF-identifier, S1, G1, sdY, VpnLabelY, BitString=PE2)
  - The Re-imposition process need an IP lookup ---- actually an MFIB lookup.
- IETF103: Defined a new term 'Pseudo VRF' for clear.

# Think a little more




- Case 1 (the above diagram):
  - Ingress Area using P2MP ---- 'less specific replication', for example, SPMSI(*,*) for rep.
  - Egress Area using BIER ---- 'most specific replication', or say, 'per-flow specific replication'.
    - The LIR-pF explicit-tracking is still available ---- so do the LIR explicit-tracking.
- Case 2 (the opposite to the above diagram):
  - Ingress Area using BIER ---- 'per-flow specific replication' generally.
  - Egress Area using P2MP ----'less specific replication', for example, SPMSI(*,*) for rep.
    - Trade-off difficulty ----not only for LIR-pF, but also for LIR.
    - If Ingress Area(BIER) uses 'less specific replication' BIER ----not optimized replication.
    - If Egress Area(P2MP) uses 'most specific replication' Per-flow SPMSI ----possibly overloaded.
- IETF103: the P2MP-BIER, BIER-P2MP are covered and updated.