EVPN Inter-subnet Multicast Forwarding

draft-lin-bess-evpn-irb-mcast-03

Wen Lin
Jeffrey Zhang
John Drake
Jorge Rabadan
Ali Sajassi

IETF 98, March 2017

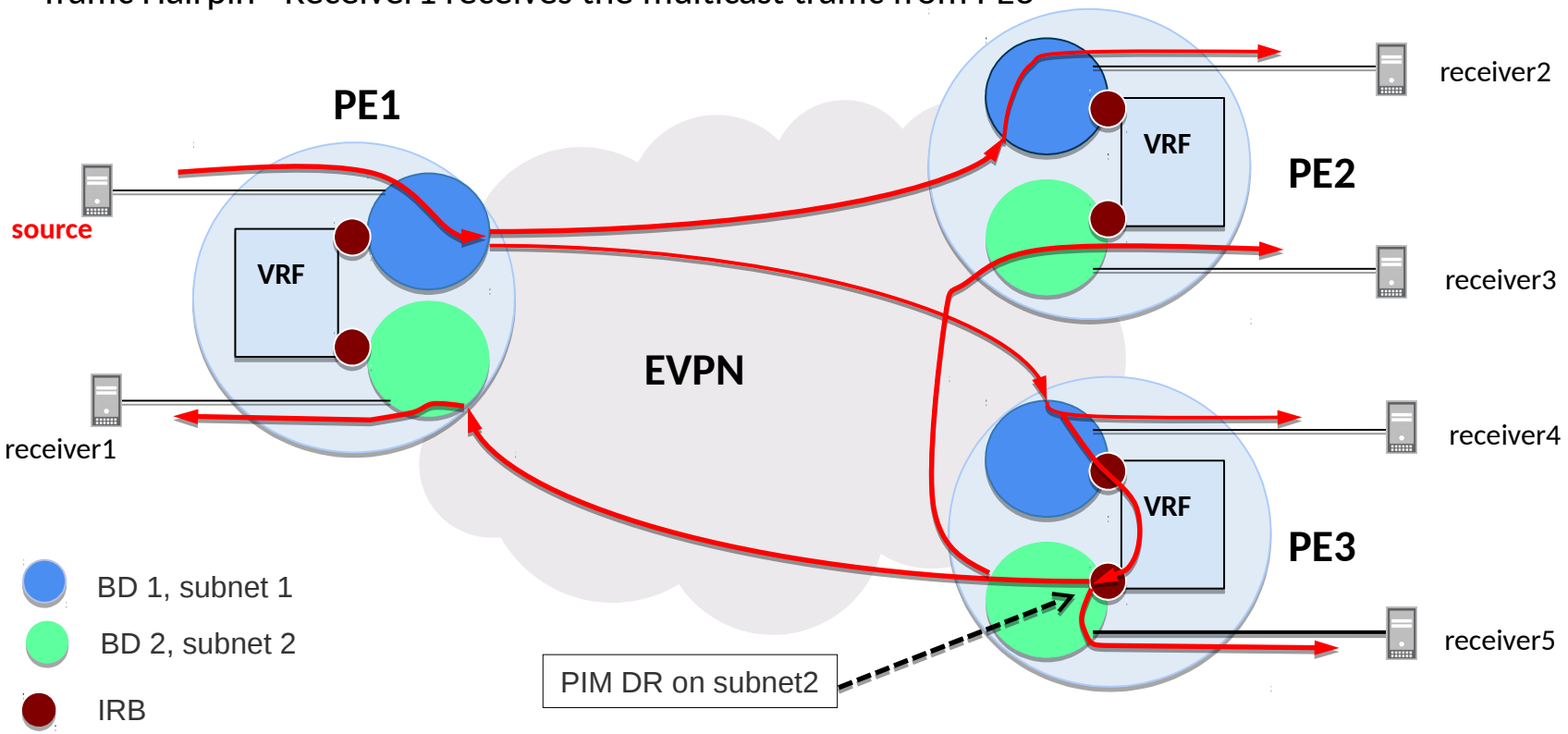
Agenda

- Background and base solution in previous revisions
- Advanced topics covered in this revision
- Plan

Transparent PIM Routing Method

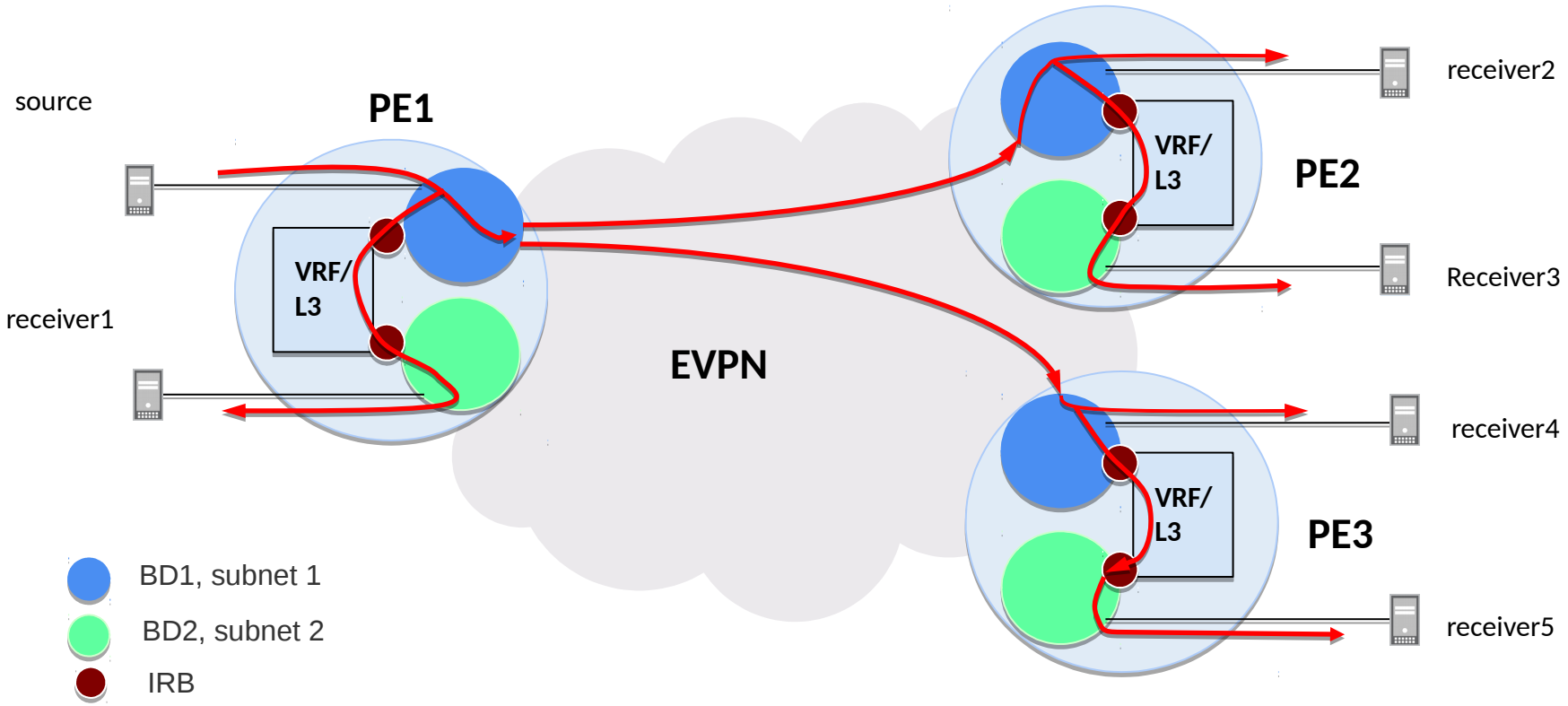
- Multicast Traffic is switched to receivers in source subnet1 following EVPN BUM procedure
- A copy is routed to other subnets via IRB following regular PIM procedure
  - On PE3, multicast traffic is routed to IRB in the subnet2
  - The routed traffic is then switched in the subnet2 following EVPN BUM procedure

- Traffic sent across core multiple times - once in every subnet/BD
  - PE2 receives the same multicast flow from the core in EACH subnet
- Traffic Hairpin - Receiver1 receives the multicast traffic from PE3



Optimized Inter-Subnet Multicast (OISM)

- Traffic is L2 switched following existing EVPN procedures to all NVEs in the source subnet
- A copy of the packet received in the source subnet, whether on an AC or from the core, is routed into other subnets that have receivers, w/o requiring PIM or regardless of PIM DR status
  - Via IRB interfaces
- Multicast data traffic routed down an IRB interface is only sent to local receivers and not across the core
- If PIM runs on an NVE for the tenant, joins are sent for local receivers even if the NVE is not a PIM DR

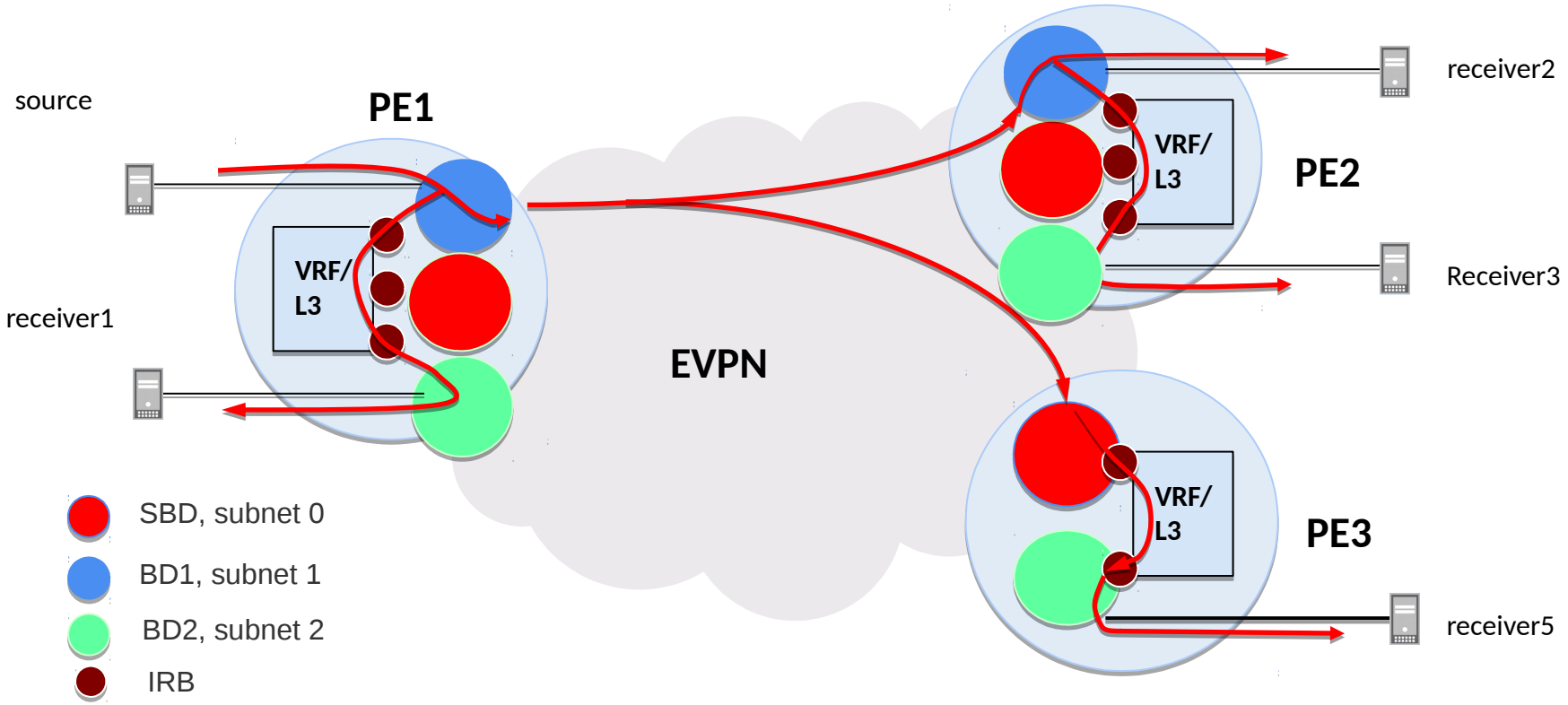


Advanced topics in -03

- **Subnets not presented on every NVE**
- **Selective Multicast**
- **Sources/receivers outside of EVPN domain**
- MVPN integration
- NVEs not supporting OISM
- Tenant routers

Source subnet not presented everywhere

- Configure a Supplemental BD (SBD) on all NVEs of a tenant
  - Either vlan-based or in a vlan-aware bundle
  - Have IRB but no ACs
- Traffic switched in the source subnet will reach all NVEs and then switched/routed accordingly
  - On receiving NVEs on the source subnet, traffic is associated with the source subnet
  - On receiving NVEs not on the source subnet, traffic is associated with the SBD



IMET Routes related to SBD

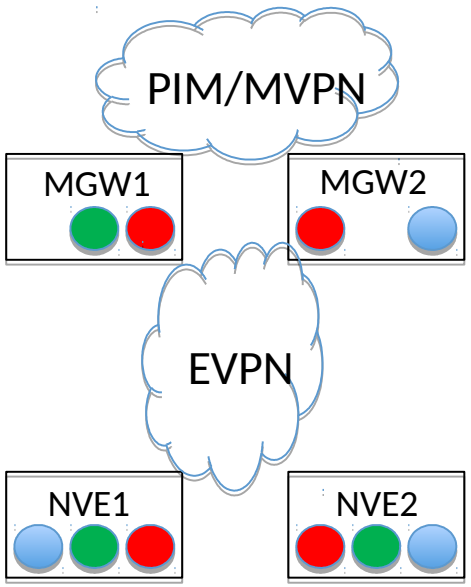
- Ingress Replication
  - Each NVE advertises an IMET route for the SBD
    - With RT for the SBD – routes imported by all NVEs
    - Downstream allocated label in the route maps to the SBD
  - Source NVE builds tunnel for a source subnet by checking SBD IMET routes
    - If a remote NVE also advertises an IMET route for the source subnet, that source subnet IMET route's label is used
    - Otherwise, the remote NVE's SBD IMET route's label is used
- mLDP Tunnel
  - For the source subnet, an NVE advertises IMET route with two RTs
    - One RT for the source subnet, and one RT for the SBD
      - The SBD RT causes the route to be imported by all NVEs
  - A receiving NVE joins the announce the tunnel, and bind it to:
    - the source subnet if it is present on the NVE
    - the SBD otherwise
- Other tunnels – details in the spec

Selective Multicast

- With Selective Multicast Ethernet Tag (SMET) routes, and S-PMSI routes in case of P2MP/PIM tunnels
- If an NVE has snooped local (s,g) or (*,g) IGMP/MLD membership, it proxies the state from tenant BDs into the SBD, which may trigger corresponding (s,g)/(*,g) SMET routes in the SBD
  - SMET route is not needed in tenant BDs
    - Unless there are tenant routers connected in those BDs
      - They'd need to receive IGMP/MLD reports/leaves in the tenant BDs converted from SMET routes
- In case of Ingress Replication
  - A source NVE builds a selective tunnel in a source subnet to NVEs from which it has received corresponding SMET routes in the SBD
  - If the remote NVE also advertises an IMET route for the source subnet, the label in that IMET route is used. Otherwise the label in the SBD IMET route is used
  - The only difference from inclusive tunnel case is that SMET route is used when determining the remote NVEs
- Other tunnel types: details in the spec

Connecting to external world

- Multicast GWs (MGWs) connect EVPN subnets to external network/MVPN
  - MGWs are also NVEs
- SBD and some tenant BDs are present on the MGWs, with IRB interfaces
  - MGWs run PIM on the IRB interfaces and run PIM/MVPN externally
- MGWs behave as PIM FHRs and LHRs on the IRB interfaces
  - Send PIM join or MVPN C-Multicast route towards external sources/RPs for receivers in BDs
    - Learnt via IGMP/MLD messages or SMET routes in the SBD or tenant BDs
  - Send PIM register messages toward external RPs for ASM sources in BDs



Sources/receivers outside EVPN

- For traffic from external sources
  - The PIM joins pull traffic to the MGWs, who will route into BDs
    - Local receivers attached to MGWs receive in the tenant BDs directly
    - Remote NVEs receive traffic routed through the SBD from the PIM DR MGW and then route into their attached BDs to their local receivers
      - ***Traffic routed down an SBD IRB interface is sent across the core***
- For traffic from internal sources to external receivers
  - ASM: DR MGW receives all traffic and sends register towards external RP
    - Register-stop will come or one of the MGWs will receive source tree joins
  - SSM: one of the MGMs will receive source tree joins
  - In all cases, corresponding SMET routes are advertised into EVPN to pull/prune corresponding traffic from internal sources

Next Steps

Seek and address comments from the WG

Update procedures for interop with NVEs not supporting the optimization

Add/update details on connecting tenant routers

Add text about Assisted Replication

Will seek WG adoption afterwards