PIM Proxy in EVPN Networks
draft-skr-bess-evpn-pim-proxy-00

Jorge Rabadan (Nokia)
Jayant Kotalwar (Nokia)
Senthil Sathappan (Nokia)
Zhaohui Zhang (Juniper)
Ali Sajassi (Cisco)

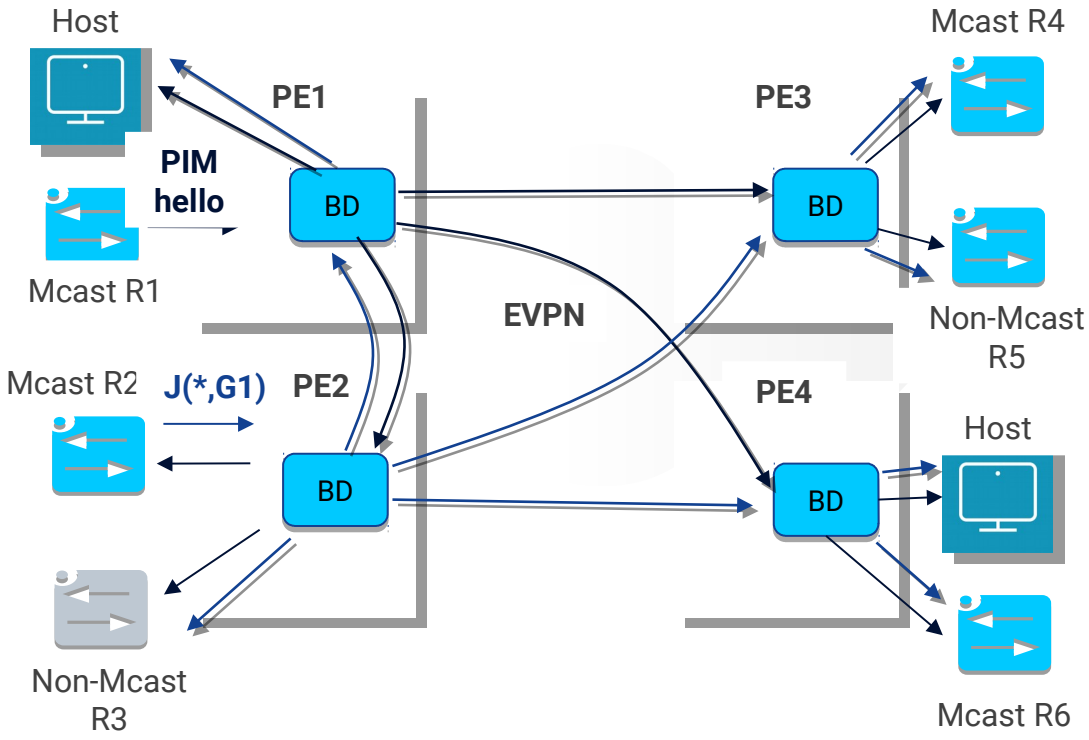
IETF99, July 2017
Prague

Background and objectives

Why PIM Proxy for EVPN networks

- Need to reduce flooding in an EVPN Broadcast Domains that are used as a "shared LANs" for PIM routers.
  - Similarly to proxy-ARP/ND and IGMP proxy in EVPN
  - VPLS also supports PIM Proxy

- Objectives
  1. Reduce/eliminate PIM message flooding in the core and to hosts/non-multicast routers. Focus on Hello and J/P messages.
  2. Forward IP multicast streams efficiently.
  3. Avoid IP multicast duplication and Assert procedures in the EVPN BD.
  4. Provide a fast failover multi-homing solution for PIM routers.



PIM Proxy for EVPN procedures

Multicast Router Discovery for PIM Proxy

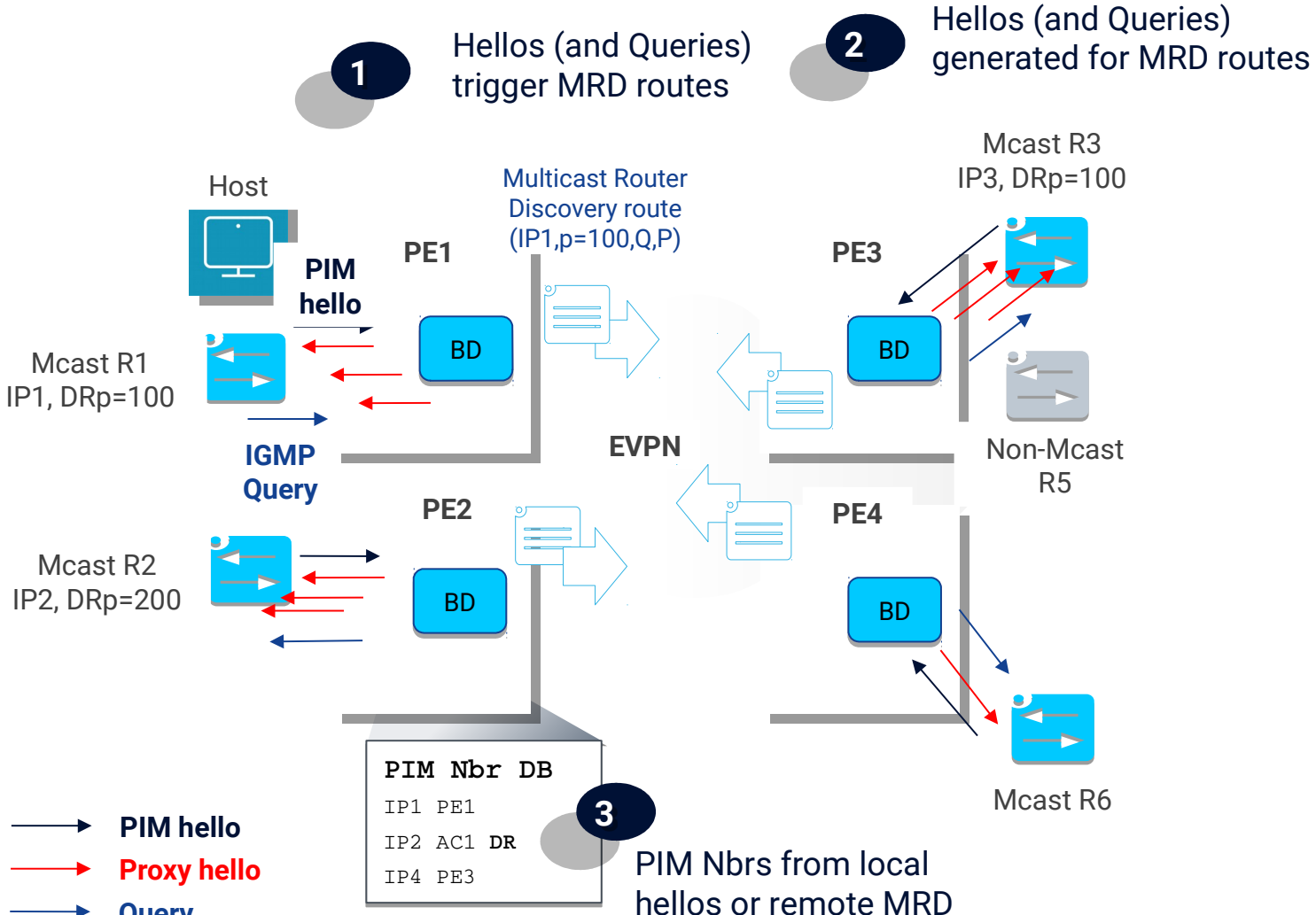
PIM Join/Prune Proxy Procedures

PIM Assert Optimization for EVPN BDs

EVPN multi-homing and PIM state synchronization

Interaction with IGMP hosts and sources in the same EVPN BD

PIM proxy for EVPN
Multicast Router Discovery routes



PIM Nbrs from local hellos or remote MRD routes
(timers, gen-id handled locally)

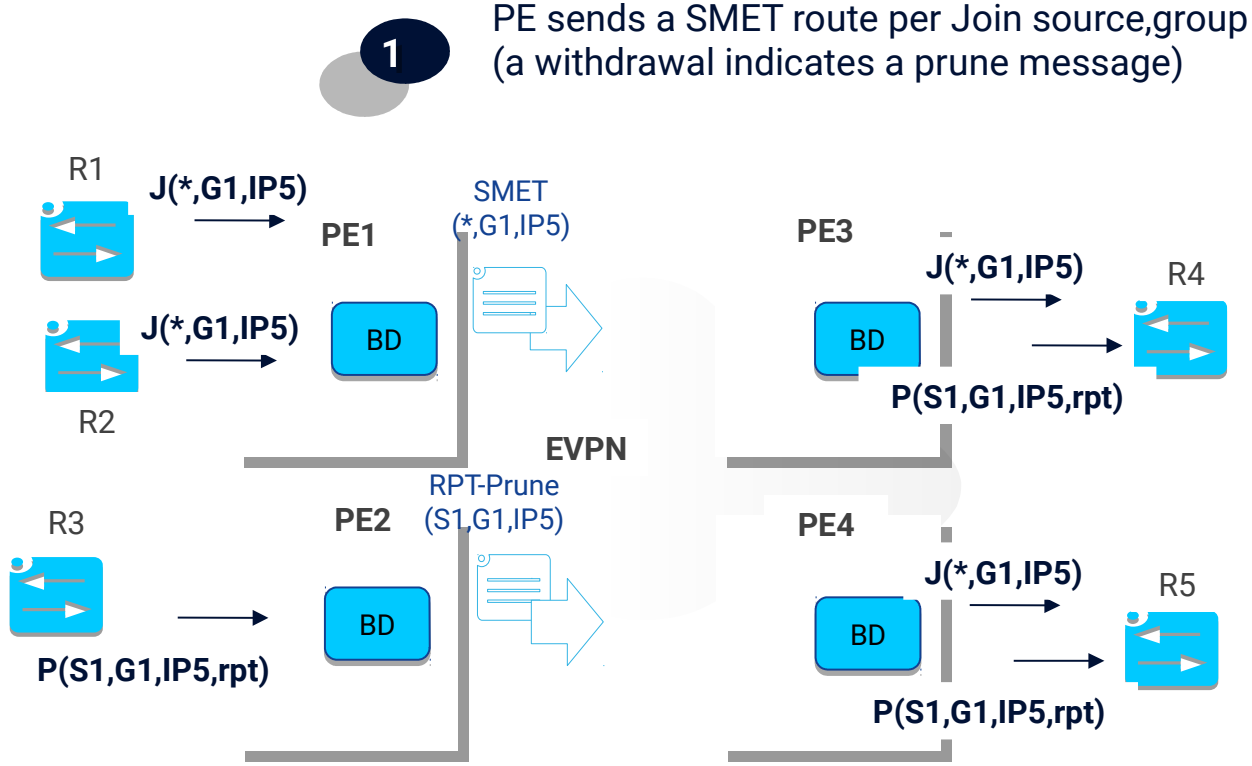
Multicast Router Discovery (MRD) route

New Route type that replaces soft-state hellos and queries

```
+-----------------------------------------------+
|  RD (8 octets)                                |
+-----------------------------------------------+
|  Ethernet Segment ID (10 octets)              |
+-----------------------------------------------+
|  Ethernet Tag ID (4 octets)                   |
+-----------------------------------------------+
|  Originator Router Length (1 octet)           |
+-----------------------------------------------+
|  Originator Router Address (Variable)         |
+-----------------------------------------------+
|  Mcast Router Length (1 octet)                |
+-----------------------------------------------+
|  Mcast Router Address 1 (variable)            |
+-----------------------------------------------+
|  Secondary Address List Length (1 octet)      |
+-----------------------------------------------+
|  Secondary Mcast Router Address 1 (variable)  |
+-----------------------------------------------+
|                .                              |
|                .                              |
|  Secondary Mcast Router Address n (variable)  |
+-----------------------------------------------+
|  DR Priority    (4 octets)                    |
+-----------------------------------------------+
|  Flags (1 octet)                              |
+-----------------------------------------------+
```

Flags:
Q: Querier flag. It indicates the encoded multicast router is a Querier.
P: PIM router flag. It indicates that the multicast router is a PIM router.
Q and P may be set simultaneously.

(modified) SMET route procedures and RPT-Prune routes



RPT-Prune route
For PIM Proxy

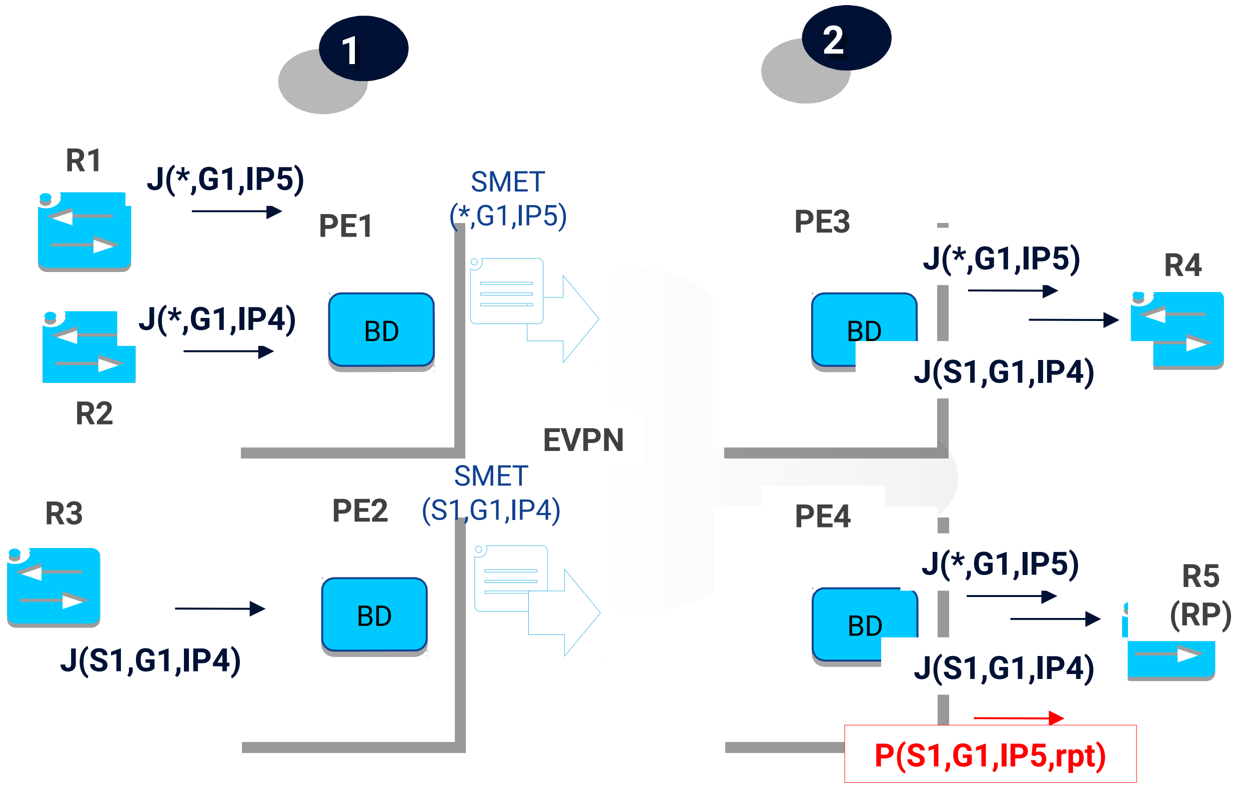
(modified) SMET route
For PIM Proxy

```
+---------------------------------------+
|  RD (8 octets)                        |
+---------------------------------------+
|  Ethernet Tag ID (4 octets)           |
+---------------------------------------+
|  Multicast Source Length (1 octet)    |
+---------------------------------------+---------------------------+
|  RD (8 octets)                        | ource Address (variable)  |
+---------------------------------------+---------------------------+
|  Ethernet Tag ID (4 octets)           | oup Length (1 octet)      |
+---------------------------------------+---------------------------+
|  Multicast Source Length (1 octet)    | oup Address (Variable)    |
+---------------------------------------+---------------------------+
|  Multicast Source Address (variable)  | outer Length (1 octet)    |
+---------------------------------------+---------------------------+
|  Multicast Group Length (1 octet)     | outer Address (variable)  |
+---------------------------------------+---------------------------+
|  Multicast Group Address (Variable)   | ter Length (1B)           |
+---------------------------------------+---------------------------+
|  Originator Router Length (1 octet)   | ter Addr (variable)       |
+---------------------------------------+---------------------------+
|  Originator Router Address (variable) |
+---------------------------------------+
|  Flags (1 octets) (optional)          |
+---------------------------------------+
|  Upstream Router Length (1B)(optional)|
+---------------------------------------+
|  Upstream Router Addr (variable)(opt) |
+---------------------------------------+
Flags:
 0  1  2  3  4  5  6  7
+--+--+--+--+--+--+--+--+
|     |  | P|IE|v3|v2|v1|
+--+--+--+--+--+--+--+--+
```

Flags:
P: PIM router flag. It indicates that the multicast router is a PIM router.

PIM Assert Optimization procedures

Avoiding multicast duplication in a shared BD



1

DOWNSTREAM PEs

- If two Joins for same (*,G) different Nbr are received, a SMET route is issued with the highest IP.
- Same for (S,G)
- If two Joins with (*,G) and (S,G) for different Nbr are received, two SMET routes are issued.

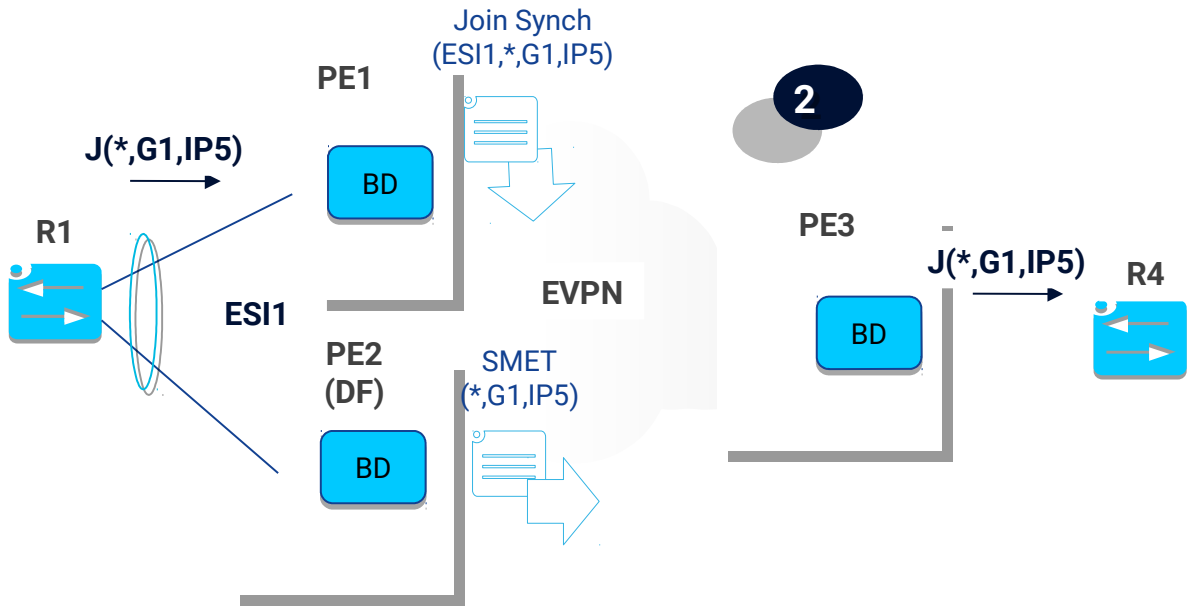
2

UPSTREAM PEs

- A single Upstream Nbr is selected per group (IP4)
- Tie-breaking rules (in order):
    1. Nbr in (S,G) SMET is preferred over (*,G) SMET
    2. Highest Upstream Nbr is preferred
- PE instructs data path to discard multicast on an interface connected to non-selected Nbr (mcast for G1 from R5 is discarded on PE4)
- PE4 issues a P(S1,G1,IP5,rpt)

PIM proxy state synchronization

Avoid proxy state synchronization in a shared BD



1

Join/RPT-Prune synch routes to synch PIM state

- Following the procedures of IGMP/NLD proxy draft

2

MRD with non-zero ESI to synch PIM Nbr DB

- All PEs in the ES will add R1 to their PIM Nbr DB and the DF will generate hellos upon receiving remote MRD routes.

(modified) Join Synch route

For PIM Proxy

RPT-Prune Synch route

For PIM Proxy

```
+-----------------------------------------------+
|  RD (8 octets)                                |
+-----------------------------------------------+
+-----------------------------------------------+t Identifier (10 octets)      |
|  RD (8 octets)                                |---------------------------+
+-----------------------------------------------+D  (4 octets)               |
| Ethernet Segment Identifier (10 octets)       |---------------------------+
+-----------------------------------------------+ce Length (1 octet)         |
|  Ethernet Tag ID  (4 octets)                  |---------------------------+
+-----------------------------------------------+ce Address (variable)       |
|  Multicast Source Length (1 octet)            |---------------------------+
+-----------------------------------------------+p Length (1 octet)         |
|  Multicast Source Address (variable)          |---------------------------+
+-----------------------------------------------+p Address (Variable)       |
|  Multicast Group Length (1 octet)             |---------------------------+
+-----------------------------------------------+ter Length (1 octet)        |
|  Multicast Group Address (Variable)           |---------------------------+
+-----------------------------------------------+ter Address (variable)      |
|  Originator Router Length (1 octet)           |---------------------------+
+-----------------------------------------------+r Length (1B)(optional)    |
|  Originator Router Address (variable)         |---------------------------+
+-----------------------------------------------+r Addr (variable)(opt)     |
|  Flags (1 octet)                              |---------------------------+
+-----------------------------------------------+
|  Upstream Router Length (1B)(optional)        |
+-----------------------------------------------+
|  Upstream Router Addr (variable)(opt)         |
+-----------------------------------------------+

 Flags:
  0   1   2   3   4   5   6   7
+---+---+---+---+---+---+---+---+
|   |   |   | P|IE|v3|v2|v1|
+---+---+---+---+---+---+---+---+
```

Conclusions and next steps

PIM proxy for EVPN completes the set of multicast optimizations for EVPN BDs

Need to agree on the new route types to be supported

- SMET routes for PIM proxy can be reused from IGMP proxy draft OR new route types can be asked for
- Same thing for PIM Join synch route

Need to agree whether other PIM procedures should be covered

- PIM Bootstrap and RP Discovery?
- PIM-DM?

We need feedback / comments from the WG

Thank you