

draft-rabadan-sajassi-bess-evpn-ipvpn-interworking-00

Jorge Rabadan (Nokia)

Ali Sajassi (Cisco)

Eric Rosen (Juniper)

John Drake (Juniper)

Wen Lin (Juniper)

Jim Uttaro (AT&T)

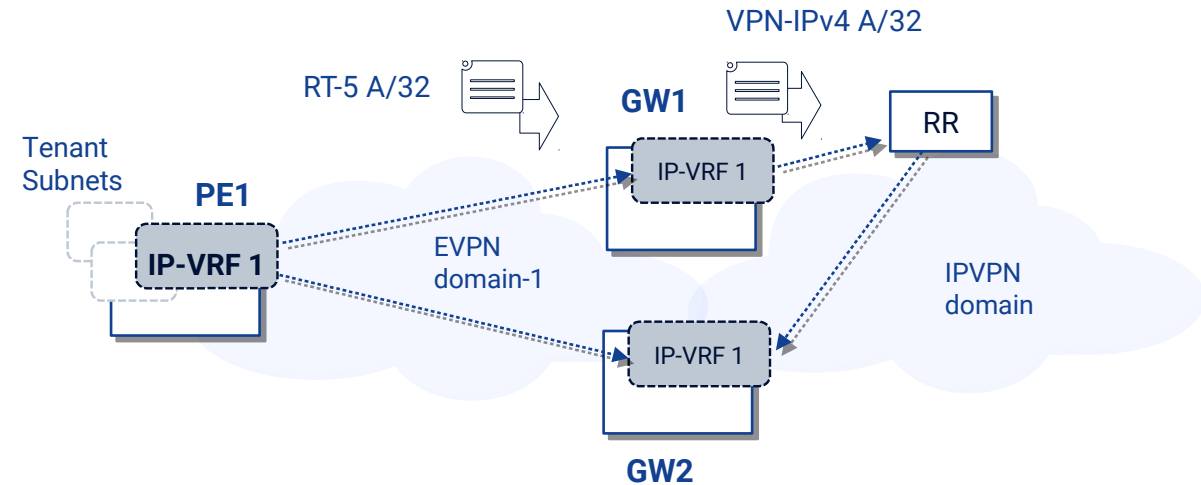
Adam Simpson (Nokia)

IETF100, Nov 2017

Singapore

Problem statement

A tenant domain main span multiple domains where different Inter-Subnet-Forwarding SAFIs (ISF-SAFIs) may provide inter-subnet connectivity.



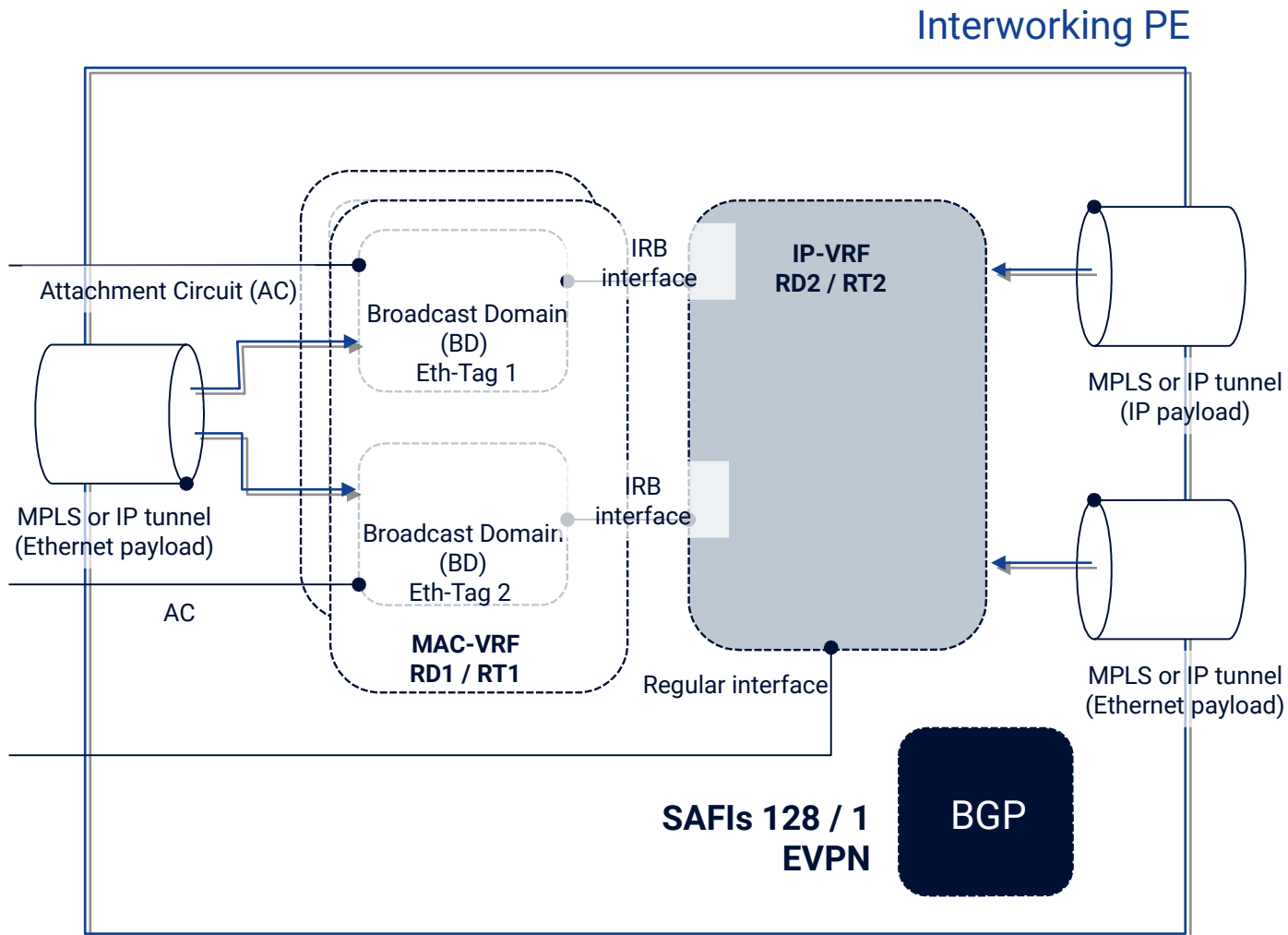
This document addresses the interworking aspects between ISF-SAFIs when one of them is EVPN:

- BGP route selection

- Loop Prevention

- Path Attribute Propagation

Terminology



Interworking PE

- a PE that may advertise a given prefix in both an EVPN route (RT-2 or RT-5) and in a route of another ISF SAFI.
- An Interworking PE has one IP-VRF per tenant, and one or multiple MAC-VRFs per tenant.
- Each MAC-VRF may contain one or more BTs, where each BT may be attached to that IP-VRF via IRB.

MAC-VRF

Defined as per RFC7432

Properties:

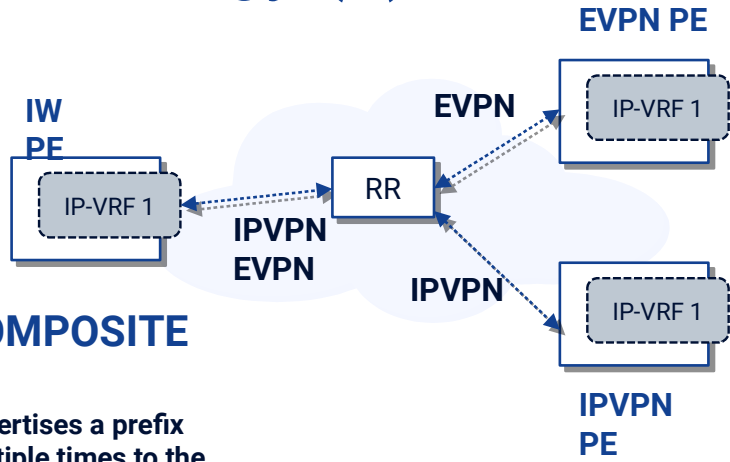
- RD and RT(s)
- N BDs per MAC-VRF, each one different Eth-tag
 - If N=1 \Rightarrow Eth-tag= 0
- Each BD may be linked to an IRB if it requires inter-subnet communication
- MPLS or IP tunnels with Ethernet payload
- Each BD has a MAC table populated by local MACs, MACs learned on ACs or via BGP EVPN

Terminology (2)

1

COMPOSITE PE

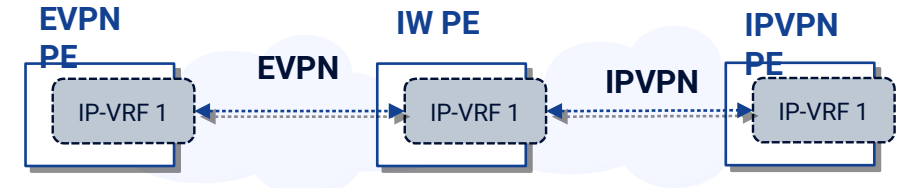
Advertises a prefix multiple times to the same peer



2

GATEWAY PE

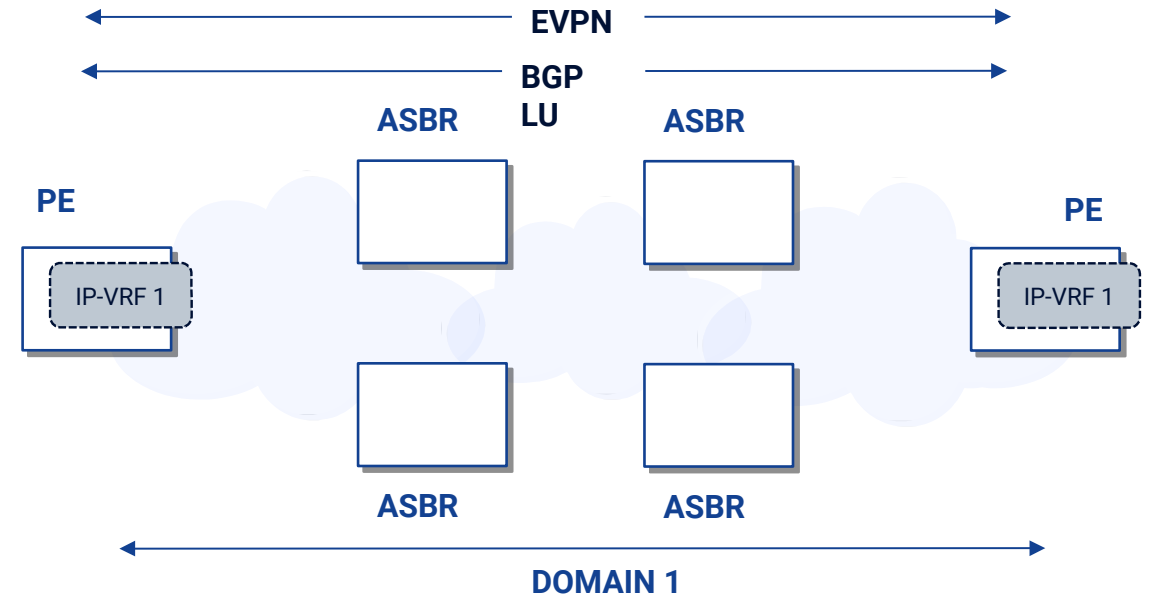
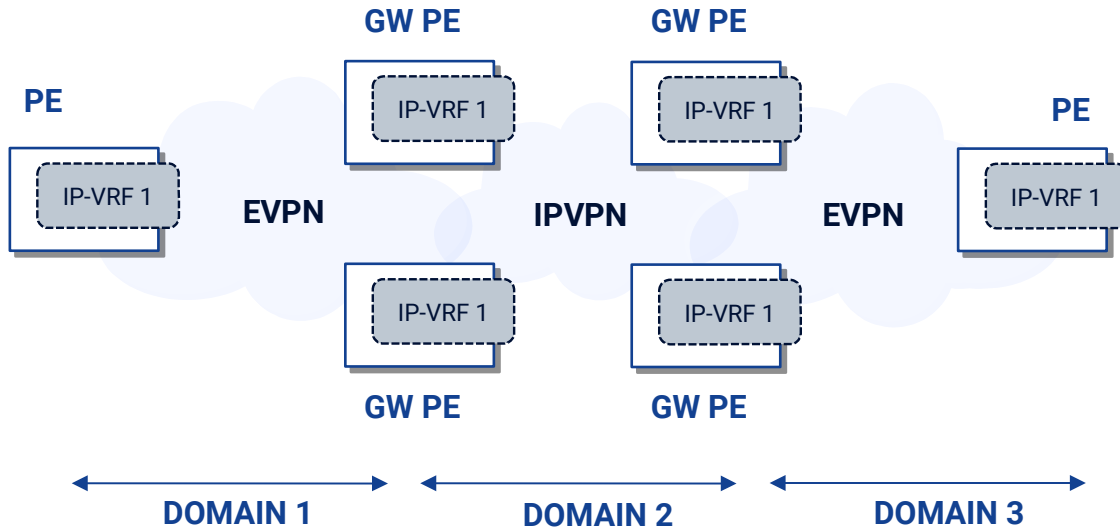
Advertises a prefix to different peer



3

DOMAIN

Set of IPVRFs with the same Domain-ID. Two PEs are in the same Domain if there are no (tenant) IP lookups in intermediate routers



Domain Path Attribute (D-PATH)

A new attribute for loop protection and ISF_SAFI visibility

- Length field followed by a sequence of Domain segments, where each Domain segment is represented by <DOMAIN-ID:ISF_SAFI_TYPE>
 - Length (1B) contains the number of Domain Segments
 - DOMAIN-ID (6B) identifies the domain, 4B Global Admin field + 2 Local Admin field
 - ISF_SAFI_TYPE (1B) identifies the SAFI (1, 70 or 128)
- Domain segments:
 - The leftmost entry in the list is the <DOMAIN-ID:ISF_SAFI_TYPE> that the Gateway PE added when sending the prefix into the local DOMAIN.
 - The rightmost entry in the list is the originating <DOMAIN-ID:ISF_SAFI_TYPE> for the prefix, that was added by the first Gateway PE propagating the update between Domains.
 - Intermediate entries are transit <DOMAIN-ID:ISF_SAFI_TYPE> that the update has passed through on its way. The number of segments reflects the number of GW PEs, irrespective of the BGP speakers.
 - E.g. {<6500:2:IPVPN>,<6500:1:EVPN>} ⇒ the Prefix was originally advertised in EVPN within Domain 6500:1, re- advertised by a first Gateway PE using an IPVPN route in Domain 6500:2 and re-advertised by a second Gateway PE into the local Domain.
- D-PATH used for Loop Prevention
 - Any Interworking PE that imports a Prefix route MUST flag the route as "looped" if its D-PATH contains a <DOMAIN-ID:ISF_SAFI_TYPE> segment, where DOMAIN-ID matches a local DOMAIN-ID in the tenant IP-VRF.

Route Selection

Between EVPN and other ISF-SAFIs

Removes from consideration the routes following the rules and the order defined in [RFC4271], with the following exceptions and in the following order:

1. After removing from consideration lowest Local Preference, any routes that do not have the shortest D-PATH are also removed from consideration. Routes with no D-PATH are considered to have a zero-length D-PATH.
2. Then regular [RFC4271] selection criteria is followed.
3. If at least one route still under consideration is an RT-2 route, remove from consideration any RT-5 routes.
4. Steps 1-3 could possibly leave Equal Cost Multi-Path (ECMP) between IP and EVPN paths.
 - A. By default, the EVPN path is considered
 - B. If ECMP across ISF SAFIs is enabled by policy, and an "IP path" and an "EVPN path" remain at the end of step 3, both path types will be used.

BGP Path Propagation Across ISF-SAFIs

Two modes of operation

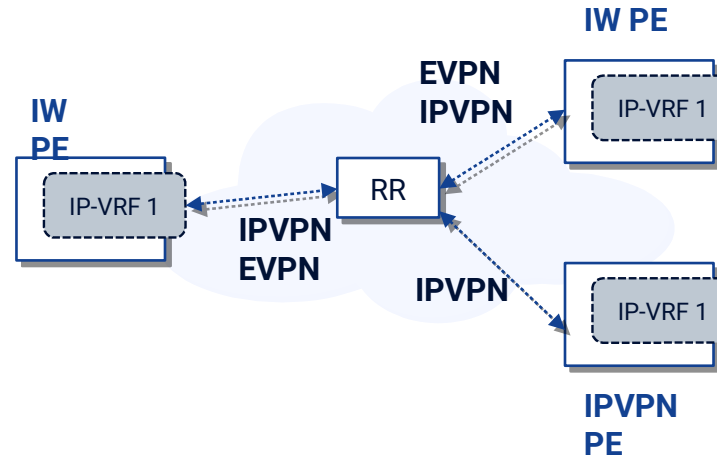
- No Propagation mode (default)
- Uniform Propagation mode

Uniform Propagation:

- AS_PATH
- D-PATH
- IBGP-only Path Attributes: LOCAL_PREF, ORIGINATOR_ID, CLUSTER_ID
- MED
- AIGP
- Communities, (non-EVPN) Extended Communities and Large Communities

Operational Procedures

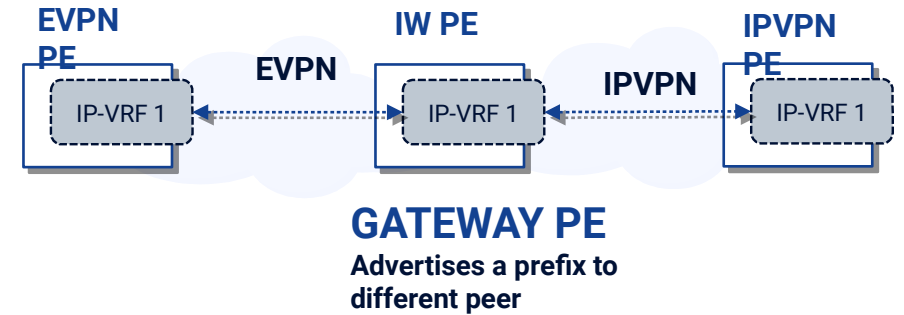
Composite PEs



- Advertise the same IP Prefixes in each ISF SAFI to the RR.
- The RR does not forward EVPN routes to non-EVPN routes.
- On Composite PEs, upon receiving the same prefix in two separate ISF-SAFIs, the route selection follows (section 4).
- If the IPVPN path is selected instead of the EVPN path, the operator should be aware that the EVPN advanced forwarding features, e.g. recursive resolution to overlay indexes, will be lost for the PE.
- When a given route has been selected as the route for a particular packet, the transmission of the packet is done according to the rules for that route's AFI/SAFI.

Operational Procedures

GW PEs



- A Gateway PE that imports an ISF SAFI-x route to prefix P in an IP- VRF, MUST export P in ISF SAFI-y if:
 - a) P is installed in the IP-VRF
 - b) PE has a BGP peer for SAFI-y (enabled for the same IP-VRF) and
 - c) Either x or y is EVPN.
- ISF SAFI routes advertised by a Gateway PE MUST include a D-PATH attribute for loop avoidance
 - When a Gateway PE re-advertises an IP Prefix between EVPN and another ISF SAFI, it MUST prepend a <DOMAIN-ID:ISF_SAFI_TYPE> to the received D-PATH attribute.
 - The DOMAIN-ID and ISF_SAFI_TYPE fields refer to the Domain over which the Gateway PE received the IP Prefix.
 - If the received IP Prefix route did not include any D- PATH attribute, the Gateway IP MUST add the D-PATH when re-advertising.

Next steps

The Authors would like to request feedback and discussion from the Working Group

Thank you