

Weighted Multi-Path Procedures for EVPN All-Active Multi-Homing

draft-ietf-bess-evpn-unequal-lb-05

Neeraj Malhotra (Cisco)

Ali Sajassi (Cisco)

Jorge Rabadan (Nokia)

John Drake (Juniper)

Samir Thoria (Cisco)

Avinash Lingala (AT&T)

IETF 108, July 2020

Online

Recap

Unequal PE-CE link bandwidth distribution within a multi-homed Ethernet Segment:

- Procedures for unequal load-balancing of flows from remote PEs.
- Procedures for unequal load-sharing of DF role across PEs in an ES.

Both overlay unicast and BUM flows load-balanced in proportion to PE-CE link bandwidth share in a LAG.

Status

- WGLC – March, 2020
- WGLC comments addressed in latest revision, except one outstanding issue
- Outstanding issue - Link BW extended community referenced from <https://tools.ietf.org/html/draft-ietf-idr-link-bandwidth-07>.

Problem with Link Bandwidth Extended Community reference

Problem

- Extended community defined in this reference is Non-Transitive
- EVPN needs this to be conditionally transitive:
 - Pass it across eBGP session when next-hop is not rewritten.
 - Drop it across eBGP session when next-hop is rewritten.
- Existing implementations and deployments make it difficult to redefine existing link BW ext-comm.

Proposal

- Define a new link bandwidth extended community for EVPN, with the above conditional transitive behavior:
 - Type: 06 (EVPN)
 - Sub-Type: 10
- Published in latest revision yesterday – 06.
- Ready to progress further.

Weighted Multi-Path Procedures for EVPN All-Active Multi-Homing

(draft-ietf-bess-evpn-unequal-lb-05)

Thank You

Neeraj Malhotra (Cisco) , Ali Sajassi (Cisco)
Jorge Rabadan (Nokia), John Drake (Juniper)
Samir Thoria (Cisco), Avinash Lingala (AT&T)

BACKUP

Solution Summary

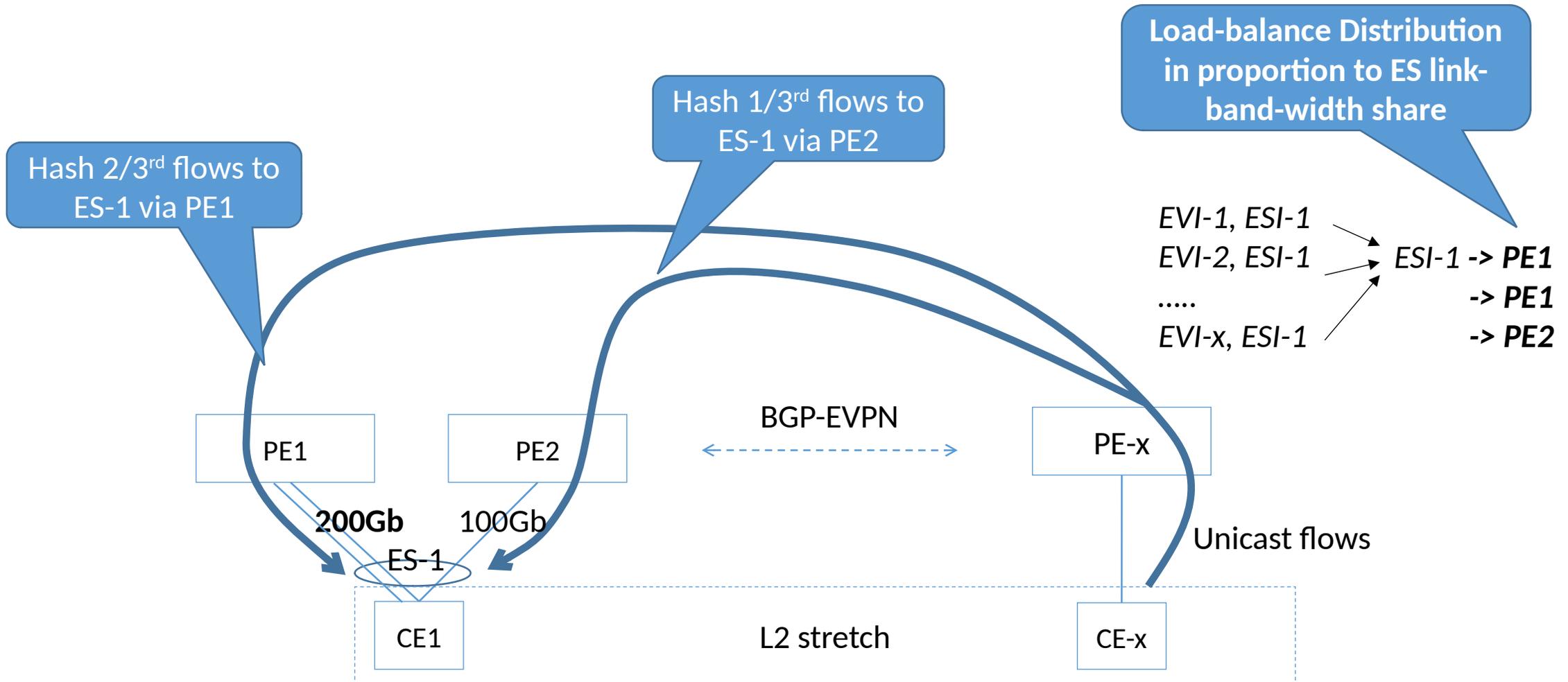
Unicast Traffic Load-Balancing

- Local PE
 - Advertises per-ESI link-band-width attribute as part of per-ESI EAD RT-1
- Remote PE
 - ESI Path-list is computed in proportion to received link-band-width attribute from each PE

DF Election

- New “BW” capability bit (28) in DF Election Extended-Community indicates desire to augment specified DF election algorithm to be “BW aware” as specified in section 4 of this draft
- Local PE
 - Advertises additional per-ES link-band-width attribute with per-ES RT-4
- Remote PE
 - Type 0 (service carving): Candidate PE list computed in proportion to bandwidth share
 - Type 1 and 4 (HRW): Candidate hash computations for each PE in proportion to it’s bandwidth share
 - Weighted HRW (Type TBD): BW weighted score computation for each PE
 - Type 2 (Preference): additional link-band-width tie-breaker based on PE’s bandwidth share

Overlay Load Balancing in proportion to PE-CE link bandwidth share in a LAG



DF Role Load Sharing in proportion to PE-CE link bandwidth share in a LAG

