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

draft-ietf-bess-evpn-unequal-lb-02

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

IETF 105, July 2019
Montreal
Objective

Optimally handle scenarios with unequal PE-CE link bandwidth distribution within a multi-homed Ethernet Segment:

- Load-balance overlay unicast flows “unequally” in proportion to each PE’s link bandwidth share in a LAG
- Load-share DF role “unequally” in proportion to each PE’s link bandwidth share in a LAG

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

Hash 2/3<sup>rd</sup> flows to ES-1 via PE1

Hash 1/3<sup>rd</sup> flows to ES-1 via PE2

BGP-EVPN

Load-balance Distribution in proportion to ES link-bandwidth share

EVI-1, ESI-1

EVI-2, ESI-1

......

EVI-x, ESI-1

ESI-1 -> PE1

-> PE1

-> PE2

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

- PE1
- PE2
- PE-x
- CE1
- CE-x

Links:
- 200Gb from PE1 to CE1
- 100Gb from PE2 to CE1
- BGP-EVPN between PE1 and PE2
- L2 stretch
- ES-1 DF for 2/3rd EVIs
- ES-1 DF for 1/3rd EVIs

BUM Replication across all EVIs
Updates since last version

- Section 4.4 – added handling for new DF election algorithm proposed in Weighted HRW draft (draft-mohanty-bess-weighted-hrw-00) - BW weighted score computation for each PE minimizes reassignment.
- Updated references, minor editorial comments
Draft Status

• Stable & in good shape
• Pending implementation (in plans)
• ready for WGLC
Weighted Multi-Path Procedures for EVPN All-Active Multi-Homing
(draft-ietf-bess-evpn-unequal-lb-02)

Thank You

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