EVPN Preference-based DF Election

draft-rabadan-bess-evpn-pref-df-00

Jorge Rabadan
Senthil Sathappan
Tony Przygienda
Wen Lin
Tapraj Singh
Ali Sajassi
Satya Mohanty

IETF 95, April 2016
Buenos Aires, Argentina
The need to improve RFC7432’s DF Election

RFC7432 Service-carving does not meet the Service Provider operational requirements

The user must be able to control the Designated Forwarder (DF) Election with an admin Preference value per EVI/ISID

The user must be able to preempt the DF at any moment without changing the configuration in all the PEs

The user must be able to configure a given ES with a “revertive” or “non-revertive” operation. Non-revertive avoids service impact when an ES comes back up.

The solution must work for:
• SA and AA multi-homing
• EVPN and PBB-EVPN
• Virtual and non-virtual Ethernet-Segments
What new BGP attributes does Pref DF Election use?

DF Types:
- Type 0 - Default, mod based DF election as per RFC7432.
- Type 1 - HRW algorithm as per [EVPN-HRW-DF]
- **Type 2 - Preference algorithm (this document)**

DF Preference – 0..65535; default 32767

Candidate PEs will be ordered based on the advertised Pref and DP bit

Pref DF Election uses type 2 in the DF Election extended community defined in draft-mohanty-bess-evpn-df-election
The Preference algorithm

**DF Pref exchange**
PEs exchange ES routes including the DF Election ext-community

**ES provisioning**
The user provisions a [Pref, Preempt option] per ES
If multiple EVI/ISIDs are associated to the ES, the user will configure EVI/ISID ranges, e.g.:
- EVI/ISID range-1 ⇒ use highest-pref
- EVI/ISID range-2 ⇒ use lowest-pref

**DF Election for the ES**
DF Election type (2) must be consistent across the PEs in the ES (otherwise fall back to service-carving)
Candidate DF list ordered by Pref, DP bit and PE-address
After the DF timer, the PEs run the DF election per PE/EVI
The “Non-revertive” option avoids service-impact on failure recovery

Don’t Preempt (DP) bit exchange
- Optional Non-Revertive config option per ES
- If configured with the NR option, after the DF timer and DF Election each PE sends an update with DP=1
- The DP bit is used as tie-breaker (it does not change the DF Election result unless the same Pref exists in another PE)

Former DF’s failure recovery
- PE3’s ES comes back up
- After a boot-timer/hold-timer PE3 compares its [Pref, DP] with the other PEs’ [Pref, DP]
  - If PE3’s Pref IS NOT the highest → PE3 sends an update with its admin [Pref, DP]
  - If PE3’s Pref IS the highest → PE3 sends an update with an ‘in-use’ [Pref, DP] matching the second highest but DP=0, e.g. Pref=200, DP=0
- PE does not take over as long as PE2 (current DF) is active.
Conclusions and next steps

• Current RFC7432’s DF Election does not meet some of the operational requirements needed by some Service Providers
  
  Preference based DF Election
  
  Manual preemption of the DF on-the-fly for maintenance operations
  
  Non-revertive behavior

• This document provides a solution to satisfy the above requirements

• The authors request more feedback from the WG and solicit WG adoption given the interest expressed by multiple Service Providers