Preference-based EVPN DF Election
draft-rabadan-bess-evpn-pref-df-02

Jorge Rabadan (Nokia)
Senthil Sathappan (Nokia)
Tony Przygienda (Juniper)
Wen Lin (Juniper)
John Drake (Juniper)
Ali Sajassi (Cisco)
Satya Mohanty (Cisco)

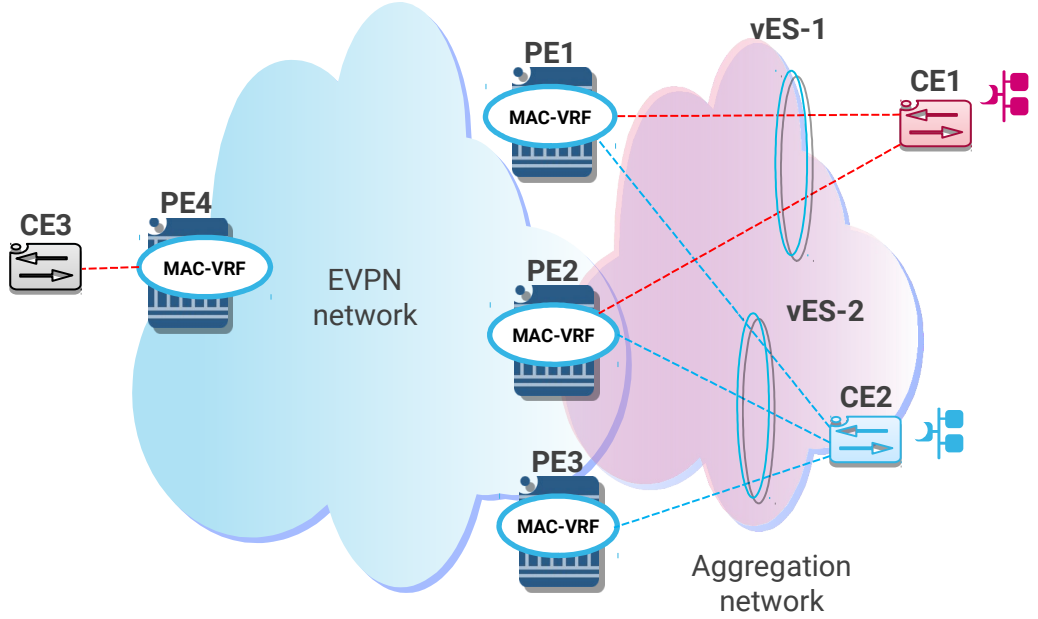
IETF98, March 2017
Chicago

draft-rabadan-bess-evpn-pref-df-03

Adding Deterministic DF Election and non-revertive behavior

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

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

Changes in rev 02

Adding non-revertive behavior on ES with Lowest and Highest Pref EVI/ISIDs

The Preference-based DF Election selects the best PE for a given ES

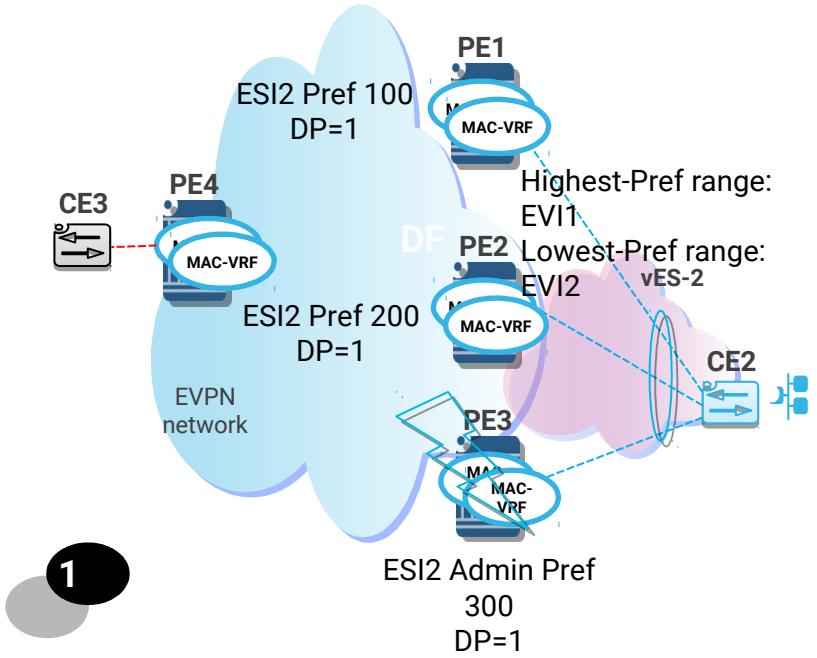
- Based on the exchange of a 2-byte Preference value in ES routes.
- By default, the highest Pref PE will become the DF for the ES.

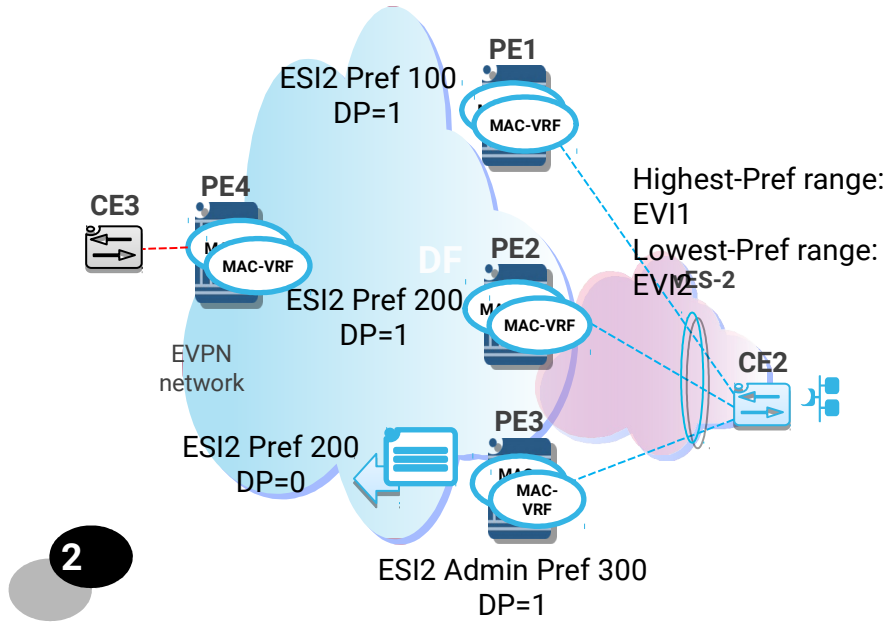
In order to provide load balancing on RFC7432 Ethernet Segments (where there are multiple EVI/ISIDs per ES) the PEs are consistently configured with:

- EVI/ISID-ranges using low Pref algorithm (the rest will use high Pref algorithm)
- E.g. : (EVI 1-2000, high_Pref), (EVI 2001-4000, low_Pref)

The non-revertive behavior must avoid any traffic hit, irrespective of the high or low Pref algorithm chosen for the EVI/ISID

Non-revertive option Example
Using high and low Pref ranges





Don't Preempt (DP) bit exchange and PE failure

- Optional Non-Revertive config option per ES
- If configured with the NR option, 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)

PE recovery

- PE3's ES comes back up and waits for boot-timer/hold-timer
- PE3 selects highest_PE (PE2) and lowest_PE (PE1)
- PE3 compares its admin [Pref, DP] with highest and lowest PEs
  - If PE3's Pref > highest_PE ␣ PE3 sends 'oper' highest_PE's Pref, DP=0
  - If PE3's Pref < lowest_PE ␣ PE3 sends 'oper' lowest_PE's Pref, DP=0
  - If lowest_PE < PE3's Pref < highest_PE ␣ PE3 sends admin Pref
  - Example: PE3's Oper Pref=200, DP=0
- PE does not take over as long as PE2 (current DF) is active.

Conclusions and next steps

This draft has vendor implementations and it is being deployed in Service Provider networks

The authors request Call for WG Adoption

Thank you