draft-ietf-bess-evpn-df-election-framework-00

S. Mohanty, Ed. (Cisco)
J. Rabadan, Ed. (Nokia)
A. Sajassi (Cisco)
J. Drake (Juniper)
K. Nagaraj (Nokia)
S. Sathappan (Nokia)

IETF101, Mar 2018
London

draft-ietf-bess-evpn-df-election & draft-ietf-bess-evpn-ac-df

A bit of History

- RFC7432 default Designated Forwarder (DF) Election procedure:
  - Process of discovering PEs in the ES, building the candidate list and choosing who the DF is - or **DF ALGORITHM**
  - RFC7432 DF ALGORITHM is based on (V mod N) function
- df-election and ac-df are 3+ year old drafts that improve different aspects of the DF Election:
  - df-election **DEFINES** a new DF Election **ALGORITHM** (HRW) and **CLARIFIES** the DF Election procedure state machine
  - ac-df **DEFINES** a new **CAPABILITY** or modification of the DF Election procedure
  - Both may work **TOGETHER**
- draft-ietf-bess-evpn-df-election-framework:
  - **MERGES** df-election and ac-df as requested by ac-df shepherd and BESS WG chair
  - **REDEFINES** DF Election extended community (initially in df-election) and sets up an IANA **REGISTRY** for DF types and capabilities
  - Includes some **IMPROVEMENTS** for HRW and ac-df

Highest Random Weight (HRW) Based DF-Election
https://tools.ietf.org/html/draft-ietf-bess-evpn-df-election

- Every PE computes hash ***H(***Pe_i, v_j), for every Pe_i which is a DF participant
- Pe_k corresponding to highest value of ***H*** is the DF for vlan v_j

Suggested hash function
H = (1103515245 * ((1103515245 * Si + 12345) XOR CRC32(D(v))) + 12345)
Computed in modulo 0x7FFFFFFF arithmetic
Where
Si = IP address of PE
D(v) = 31-bit Digest (CRC-32) of the Ethernet Tag after discarding the MSB

Important property that ensures DF for a vlan does not move among unchanged PEs:

- The hash does not depend on the number of PEs

Highest Random Weight (HRW) Based DF-Election

https://tools.ietf.org/html/draft-ietf-bess-evpn-df-election-framework-00

- Every PE computes hash ***H(***Pe_i, v_j, Es), for every Pe_i which is a DF participant
- Pe_k corresponding to highest value of ***H*** is the DF for vlan v_j

Suggested hash function

H = (1103515245 * ((1103515245 * Si + 12345) XOR CRC32 (D(v,Es))) + 12345)

Computed in modulo 0x7FFFFFFF arithmetic

Where

Si = IP address of PE

D(v,Es) = 31-bit Digest (CRC-32) of the Ethernet Tag **and** Ethernet Segment Identifier treated as a 14-byte stream (after discarding the MSB)

Important property that ensures DF for a vlan does not move among unchanged PEs:

- The hash does not depend on the number of PEs

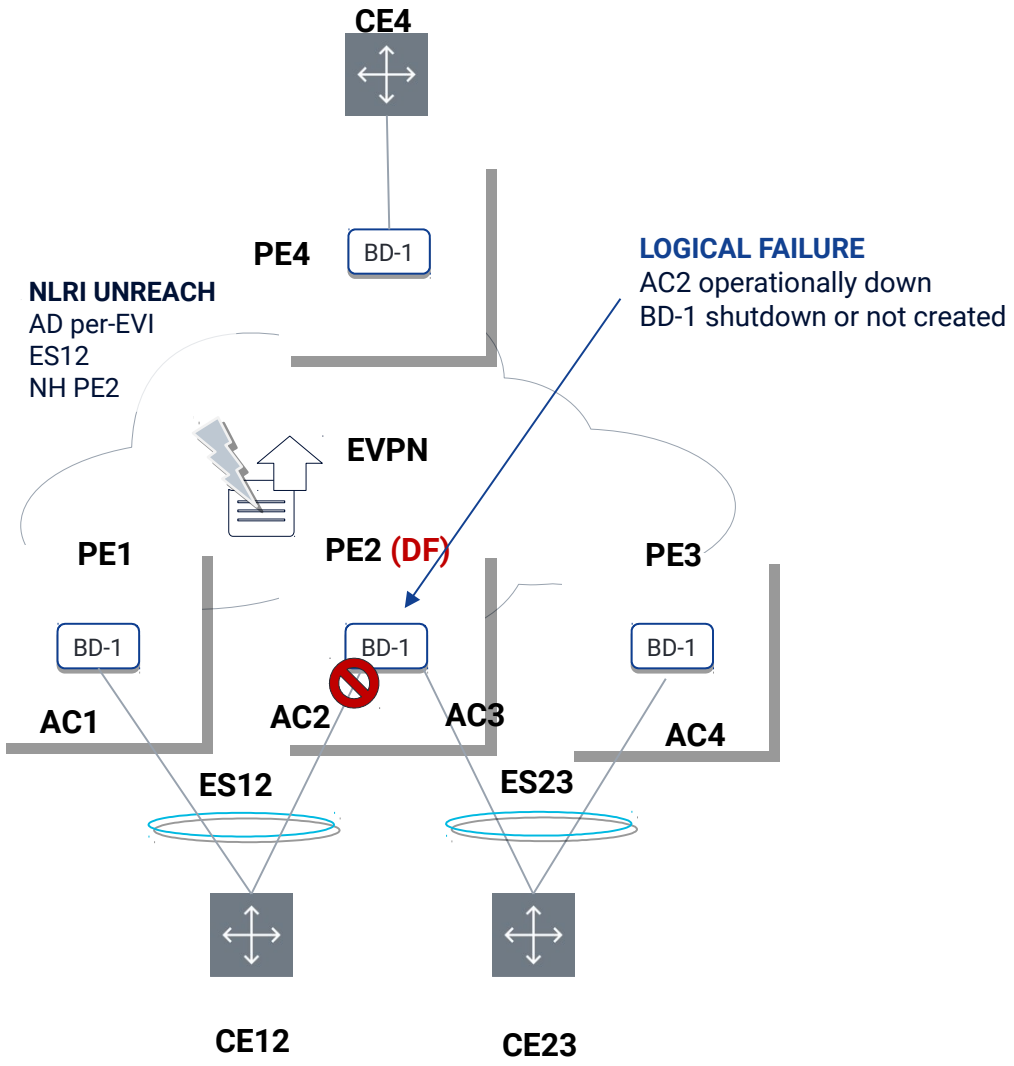
https://tools.ietf.org/html/draft-ietf-bess-evpn-df-election-framework-00 Highest Random Weight (HRW) Based DF Election

Advantages:

- If the same set of PEs are multihomed to the same set of ESes, then the DF election algorithm used in [RFC7432] would result in the same PE being elected DF for the same set of broadcast domains on each ES
- This can have adverse side-effects on both load balancing and redundancy.
- Including ESI in the DF election algorithm introduces additional entropy which significantly reduces the probability of the same PE being elected DF for the same set of broadcast domains on each ES.

AC-influenced DF Election (AC-DF capability)

Avoiding black holes due to "logical" failures



- AC-DF refresher:
  - RFC7432 mandates AD per-EVI withdrawal upon AC2 or BD-1 failures (but no influence in DF Election)
  - AC-DF prunes the candidate list based on AD per-EVI routes
- New:
  - AC-DF modifies the DF Election procedure for VLAN-aware bundle services – now per <ES,VLAN>
  - AC-DF capability is signaled to the rest of the PEs in the ES for backwards compatibility

Signaling DF algorithms and DF capabilities
DF Election extended community and request for IANA registry

DF ELECTION Extended Community

0	8	16	24
Type 0x06	Sub-Type 0x06	DF Type	Bitmap
Reserved			

Ext Community advertised/processed with ES route

- Types 0 and 1 compatible with AC-DF bit
- Inconsistent types in the ES ␣ fall back to default procedure/algorithm
- The reserved field value is specific to the DF Type (e.g. if Type=2 ␣ Preference is encoded)

DF Type (algorithm) values

- Value 0 == Default type (modulo)
- Value 1 == HRW
- Value 255 == Experimental

Others (in different specs)

- Value 2 == Preference-based DF
- Value 3 == BW-based DF
- Value 4 == Per-mcast flow DF

Bitmap

- Bit 25 == AC-DF capability

Others (in different specs)

- Bit 24 == Don't Preempt me (non-revertive DF in Pref-based)
- Bit 25 == Time

Conclusions and next steps

- Improves RFC7432’s DF Election and creates a framework for DF extensibility
- Ensures consistency and backwards compatibility in the Ethernet Segment
  - DF Types and capabilities are signaled in the DF Election extended community
- NEXT STEPs
  - Authors request immediate WG Last Call
  - PLEASE READ and COMMENT