draft-sajassi-bess-evpn-fast-df-recovery-00.txt

A. Sajassi (Cisco), B. G Badoni (Cisco),
Dhananjaya Rao (Cisco), P. Brissette (Cisco),
J. Drake (Juniper)

IETF 98, March 2017

Chicago

Problem Statement

- Baseline DF election procedure described in [RFC 7432] was a good starting point but there is room for improvement
- draft-ietf-bess-evpn-df-election-01 makes improvement on top of this baseline draft by using HRW algorithm to avoid VLAN (service ID) shuffling
- This draft describes additional incremental improvement on top of HRW for faster DF election upon PE recovery or insertion

Problem Statement

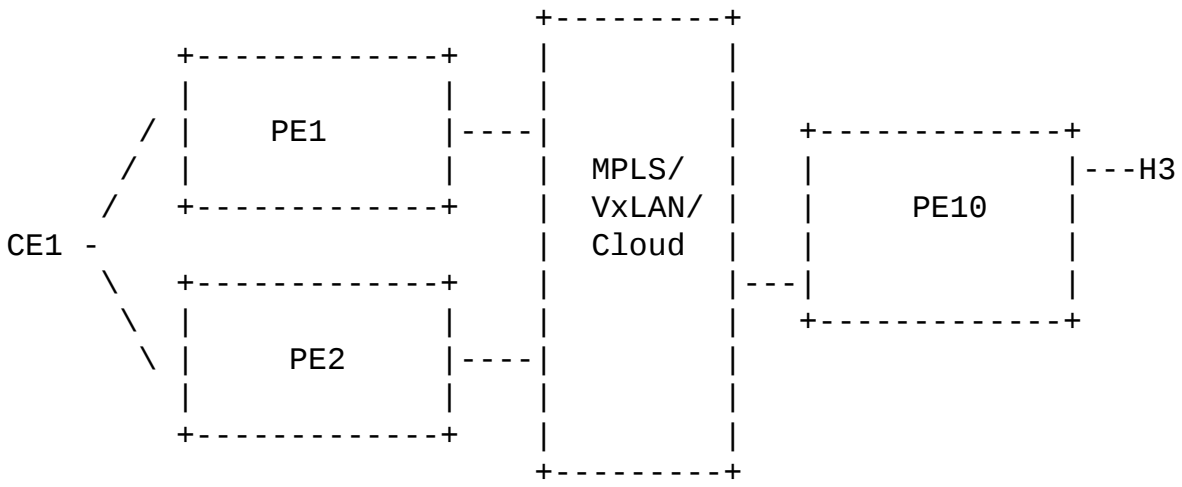


Figure 1: CE1 multi-homed to PE1 and PE2. Potential for duplicate DF.

- Currently, DF election errs on the side of transient black-holing over transient loop
- Recovered DF lets all other PEs know that it has joined the multi-homing group and starts a 3 sec. timer before doing DF election
- If timer is made too short, then there is a possibility of transient loop

Proposal

- This drat proposes two methods of reducing and even almost eliminating transient black-holing upon PE recovery or insertion
  1. Handshaking between recovered PE and other PEs in the redundancy group
  2. Time-synchronization and uni-direction signaling between recovered PE and other Pes in the redundancy group

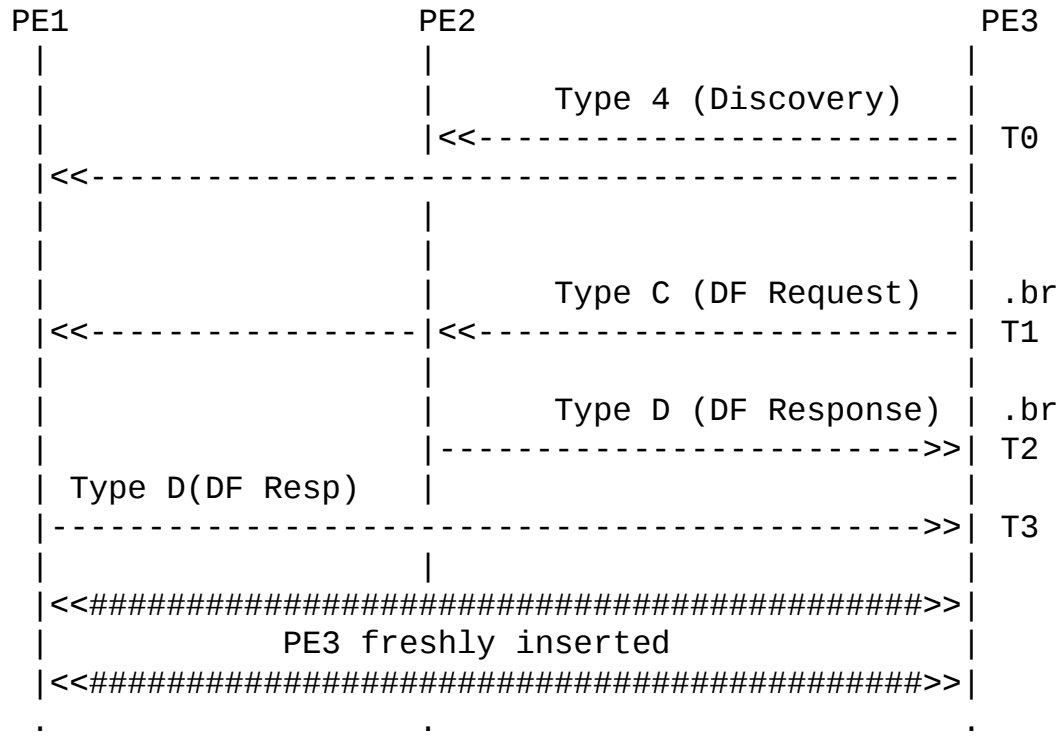
Handshake Mechanism

- Recovered/new PE advertises ES route and starts the wait timer as before
- Other PEs in the redundancy group upon receiving the ES route, run HRW algorithm for DF election as before
- If PEs in the redundancy group are capable of doing handshake, then they do the following:

Handshake Mechanism – Cont.

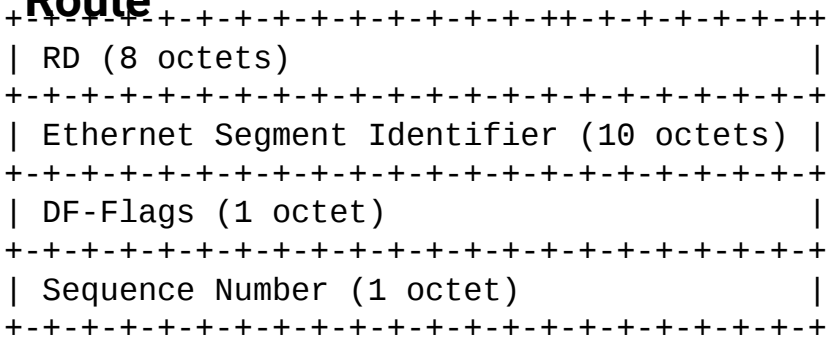
- Recovered/inserted PE sends the DF Request to previously inserted PEs with a new sequence no.
- Previously inserted PE(s) receives the DF Request and programs their hardware to block the VLANs that must be transferred to the newly inserted PE.
- Previously inserted PE(s) will send DF Response (e.g., ACK) to the newly inserted PE
- Newly inserted PE receives DF Response ACK and programs its hardware to assume DF state for the VLANs.
- *NOTE: handshaking is per PE and not per EVI/BD*

Handshake Mechanism - Cont.

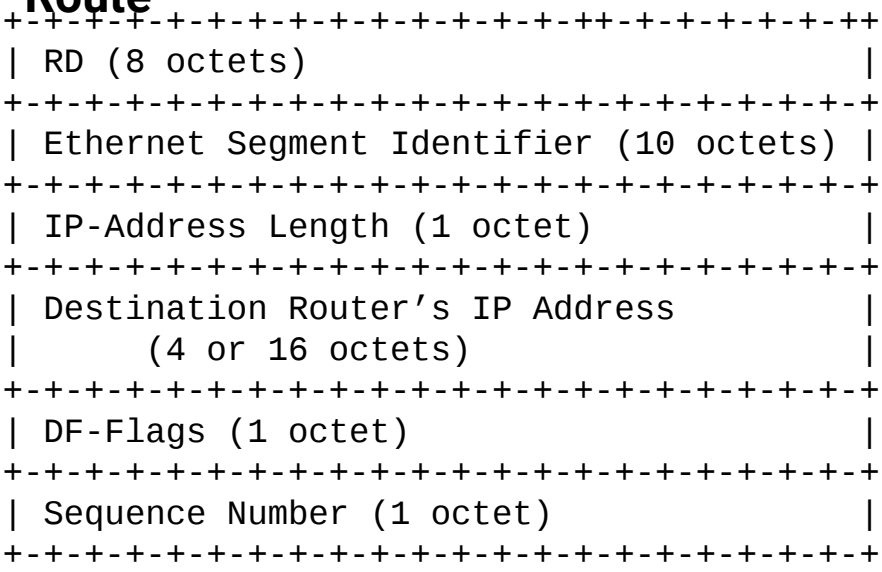


BGP Encoding

DF Election Handshake Request Route



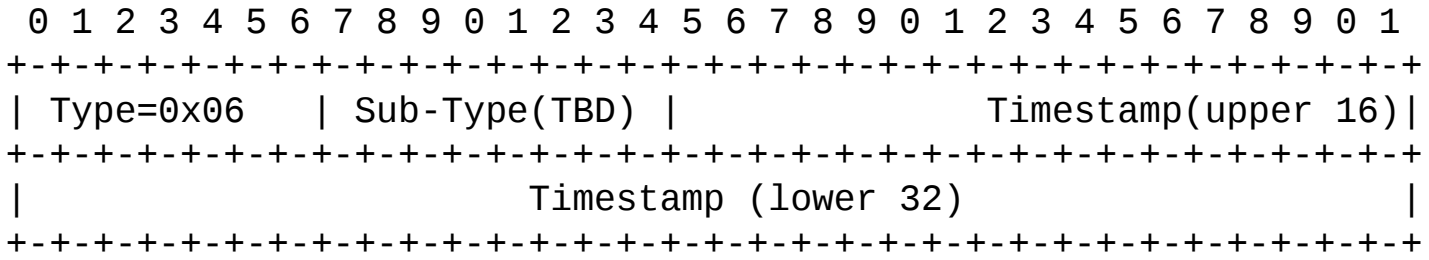
DF Election Handshake Response Route



Synchronization Mechanism

- If all PE devices attached to an Ethernet Segment are clock-synchronized with each other, then a simple uni-directional signaling can eliminate (almost) any transient black-holing and packet duplication for DF election
- Procedure:
  - A recovered/inserted PE simply signals to other multi-homing PE devices the time at which it will execute the DF election
  - All other multi-homing PE set themselves up to execute the DF election for that ES at that

Synchronization Mechanism – Cont.



This EC is advertised along with the ES route type (0x04)

Synchronization - Example

- Initial state: PE1 is in steady-state, PE2 is recovering
- PE2 recovers at (absolute) time t=99
- PE2 advertises RT-4 (sent at t=100) with target SCT value t=103 to partner PE1
- PE2 starts its 3sec peering timer as per RFC7432/HRW
- Both PE1 and PE2 carves at (absolute) time t=103; (PE1 should carve slightly before PE2 (skew))

Next Step

- More discussions among interested partitas
- Finalize the new routes
- Clarify that this approach is incremental on top of HRW draft – to avoid too many permutations
- Beef-up backward compatibility section for both mechanisms