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

- 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

$$H = (1103515245 \times ((1103515245 \times S_i + 12345) \text{ XOR CRC32}(D(v))) + 12345)$$

Computed in modulo $0x7FFFFFFF$ arithmetic

Where

- $S_i = IP$ 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
Every PE computes hash $H(\text{Pe}_i, v_j, Es)$, for every $\text{Pe}_i$ which is a DF participant.

$\text{Pe}_k$ corresponding to highest value of $H$ is the DF for vlan $v_j$.

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

- The hash does not depend on the number of PEs.
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.
Avoiding black holes due to “logical” failures

AC-influenced DF Election (AC-DF capability)

**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

<table>
<thead>
<tr>
<th>Type</th>
<th>Sub-Type</th>
<th>DF Type</th>
<th>Bitmap</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x06</td>
<td>0x06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>8</td>
<td>16</td>
<td>24</td>
</tr>
</tbody>
</table>

**DF ELECTION Extended Community**

<table>
<thead>
<tr>
<th>Ext Community advertised/processed with ES route</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Types 0 and 1 compatible with AC-DF bit</td>
</tr>
<tr>
<td>• Inconsistent types in the ES fall back to default procedure/algorithm</td>
</tr>
<tr>
<td>• The reserved field value is specific to the DF Type (e.g. if Type=2 Preference is encoded)</td>
</tr>
</tbody>
</table>

**DF Type (algorithm) values**
• Value 0 == Default type (modulo)
• Value 1 == HRW
• Value 255 == Experimental

**Bitmap**
• Bit 25 == AC-DF capability

**Others (in different specs)**
• Value 2 == Preference-based DF
• Value 3 == BW-based DF
• Value 4 == Per-mcast flow DF

**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