draft-ietf-l2vpn-evpn-00.txt

A. Sajassi (Cisco), R. Aggarwal (Arktan), W. Henderickx (ALU), N. Bitar (Verizon), A. Issac (Bloomberg), J. Uttaro (ATT), F. Balus (ALU), R. Shekhar (Juniper), J. Drake (Juniper), S. Boutros (Cisco), K. Patel (Cisco)

March 29th, 2012

IETF Paris

Status Report

- Discussions among co-authors to simply E-VPN draft
- Align it with L3VPN as much as possible
  - If operation of two are similar, it will benefits operators who are familiar w/ L3VPN

How to Simplify ?

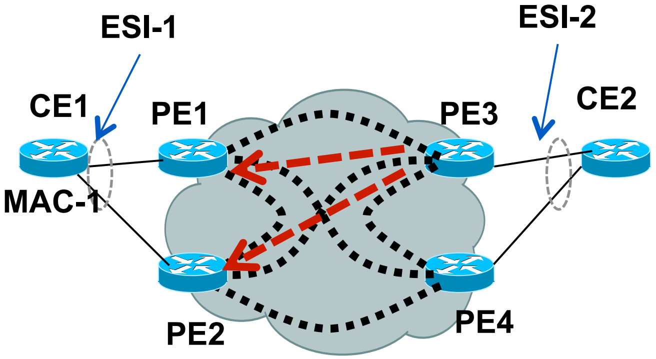
- Lot of the complexity of E-VPN is due to Ethernet AD route
- Reducing number of options will simplify E-VPN operation
- Specifying clearly what BGP attributes are allowed for what modes of this route, will help the vendor interop
- Ethernet AD Route has
  - Three forwarding modes
  - Six different route flavors

Ethernet AD Route

- The reason for having so many modes and so many flavors, is that too much of functionality has been built into this route (some of them due to the merge of MAC-VPN and R-VPLS drafts:
  1. Aliasing
  2. Multiple forwarding modes
  3. DF Election
  4. Split-horizon label advertisement
  5. Mass withdraw (upon ES link failure)
  6. Assistance in doing source-quenching flavor of ingress replication

1. Aliasing

- Sometimes CE1 may not hash its traffic for its MAC-SA to both PE1 & PE2.
- In such cases, the reverse traffic will end up on one of the Pes
- In order to ensure that the reverse traffic for that MAC-SA is shared by both PEs, Aliasing is used
- In order for PE3 to be able to perform this task, it needs to know that:
  a) ESI-1 sits behind both PE1 and PE2
  b) MAC-1 is associated with ESI-1
- PE1 and PE2 use Ethernet AD route to advertise ESI-1 sits behind them
- PE1 uses MAC route to advertise MAC1 sits behind ESI-1
- All the remote PEs (e.g. PE3) use these two routes in combination to associate
  a) MAC1 to ESI-1
  b) And subsequently MAC-1 to [PE1 and PE2]

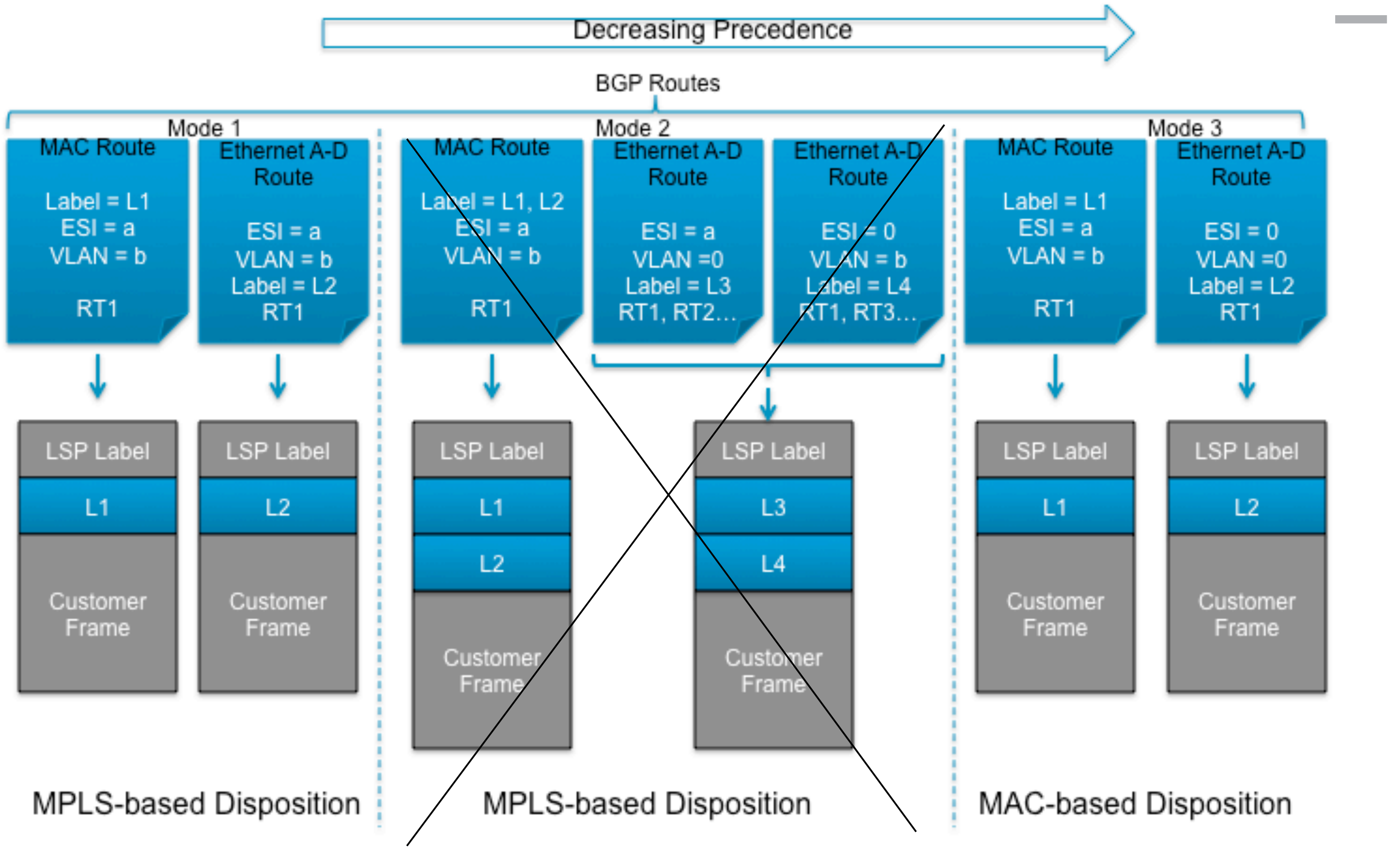


1. Aliasing – Cont.

- Aliasing doesn’t give enough bang for the buck
  - It only improves load-balancing to a given MAC in corner cases where random n-tuple hash for the same MAC-SA end up on the same PE (corner case)
  - Even in such corner cases, the aggregate traffic to that pair of PEs which serve many Multi-homed CEs can be very well balanced.
- This load-balancing is performed at the expense of additional local flooding at the PE which hasn’t learned MAC-SA of CE

- **=> should removed Ether AD options associated w/ Aliasing**

2. Multiple Forwarding Modes



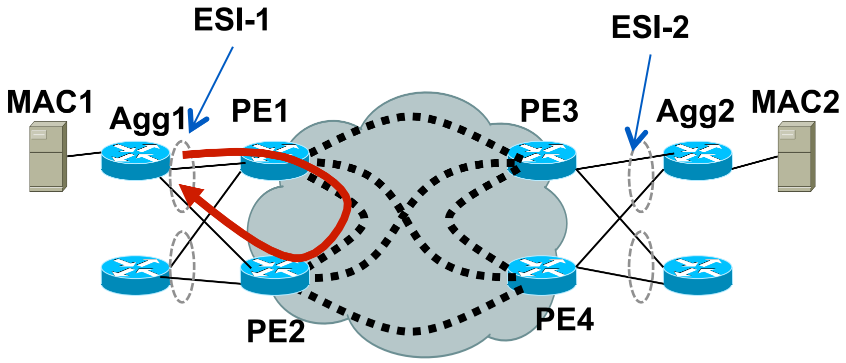
Note: All labels are downstream assigned.

3. DF Election

- The use of Ethernet AD route in DF election is inefficient because:
  - DF election needs to be performed among PEs participating in the LAG (typically 2 or 3)
  - Ethernet AD needs to be exchanged among all PEs participating in all VPNs for that segment (can be hundreds or thousands)
  - The exchanges among PEs for a given LAG can be substantial (mLACP state synchronization, etc.)
- **=> Use Segment Route for DF Election**

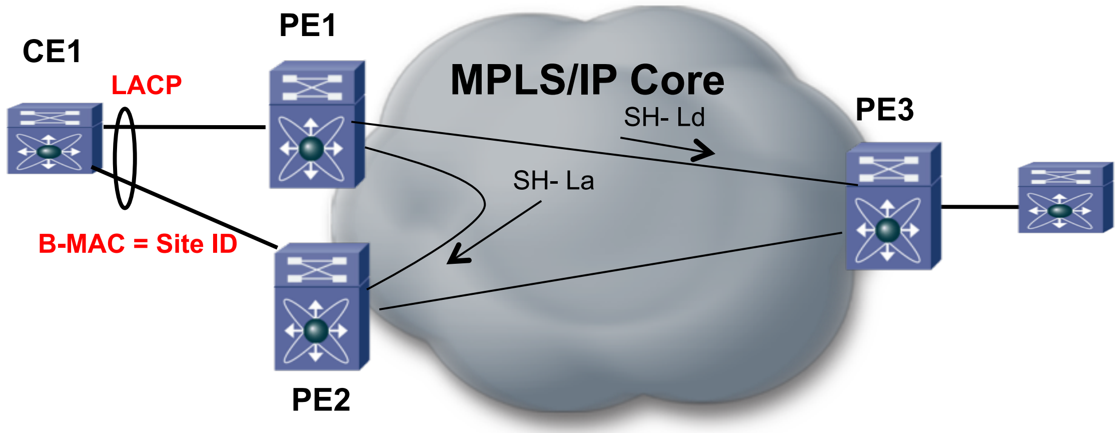
4. Split-Horizon Filtering

- PE1 advertises in BGP a split-horizon label associated with the ESI-1 (in the Ethernet AD route)
- Split-horizon label is only used for multi-destination frames (unknown unicast, mcast, bcast)
- When PE1 wants to forward a multi-destination frame, it appends this SH label to the packet
- PE2 uses this label to perform split-horizon filtering for frames destined to ESI-1
  - - e.g., a frame originated by a segment must not be received by the same segment



- For BUM traffic using P2MP LSP, we need to advertise SH label to all the PEs associated with all the VPNs for that segment
- **=> we need SH advertisement using Ethernet AD route**

4. SH – Cont. (need for don’t-care label)



- When PE1 does ingress replication (w/o source quenching), it needs to use SH downstream-assigned label of PE2 and PE3 for its source ES identification
- PE1 knows the downstream label of PE2 (SH-La) because PE2 belongs to the same ES
- However PE3 is not part of the same segment and thus PE1 never received a label from PE3
- That’s why PE3 needs to send a don’t care label (per PE) to all other PEs participating in the same set of VPNs
- **=> We can use Ether AD route for SH don’t care label advertisement**
- Note: If PE3 doesn’t send don’t care label, then it needs to be able to do its MPLS processing/forwarding decision based on depth of MPLS stack and sometimes even that is not possible – e.g., for ingress replication when a PE1 uses flow-label and PE2 doesn’t use flow-label

Ethernet AD Route

- Ethernet AD Route is a multi-personality route that it can be used to advertise:

  A. A route per <Ethernet Segment, VLAN> for MPLS forwarding

  B. A route per <VPN> for MPLS forwarding of unicast data

  C. A route per <Ethernet Segment> for advertising split-horizon label

  D. A route per <PE> for advertising don’t care split-horizon label

  E. A route per <Ethernet Segment> for MPLS forwarding (ES w/ MPLS label stack)

  F. A route per <VLAN> for MPLS forwarding (ES w/ MPLS label stack)

Ethernet AD Route: Different Flavors

Flavor	A	B	C	D	E	F
RD	~~VPN RD~~	~~VPN RD~~	Segment RD	Segment RD	~~Segment RD~~	~~VPN RD~~
Ethernet Segment ID	~~VALID~~	~~NULL~~	VALID	NULL	~~VALID~~	~~NULL~~
Ethernet Tag ID	~~VALID~~	~~NULL~~	NULL	NULL	~~NULL~~	~~VALID~~
MPLS Label	~~VALID~~	~~VALID~~	NULL	NULL	~~VALID~~	~~VALID~~
RT	~~Single~~	~~Single~~	Multiple (corresponding to all VPNs on Segment)	Multiple (corresponding to all VPN instances enabled on PE)	~~Multiple (corresponding to all VPNs on Segment)~~	~~Single~~
ESI MPLS Label Extended Community	~~Not used~~	~~Not used~~	Contains the SH Label	Contains the SH Label	~~Contains the SH Label~~	~~Not used~~
Use	~~Advertise forwarding label per (ESI, Tag) for MPLS-based disposition~~	~~Advertise forwarding label per VPN for MAC-based disposition.~~	1. Advertise SH Label for an Ethernet Segment. 2. Mass Mac withdraw upon a ES link failure	1. Advertise the special 'Don't Care' SH Label for ingress replication w/o source quenching 2. Keep MPLS label stack consistent specially w/ flow label	~~1. Advertise forwarding label per ESI for MPLS-based disposition with label stack.~~ ~~2. Advertise SH Label for an Ethernet Segment.~~	~~Advertise forwarding label per Tag for MPLS-based disposition with label stack.~~