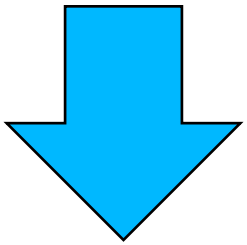
Bridge Model for ECMP Operation

Ali Sajassi

Created: December 9, 2010
Updated: January 6, 2010
IEEE 802.1 Weekly Call on ECMP

Frame Format

PBB I-tag Format



ECMP Tag Format

Service-ID/Flow-ID

- In order to avoid deep packet inspection in the core bridges, the concept of flow ID is introduced:
  - ➢The ingress BEB derives a 20-bit hash index from the five tuple representing a flow (C-MAC SA/DA, IP SA/DA, UDP port, etc.)
  - ➢The ingress BEB derives a flow ID by modulating this hash index on top of the E-SID (e.g., by simply XORing E-SID with hash index)
  - ➢The flow ID is used by all the BCBs for ECMP selection
  - ➢ On the egress PE, demodulation is performed by XORing hash index once again with the flow-ID thus retrieving E-SID

- NOTE: The use of homogenous ECMP algorithm (per B-VID) enables the egress PE of demodulating flow-ID and retrieving E-SID (another advantage of having homogenous ECMP)

ECMP Operation w/ only E-tag - Optional

Pros/Cons

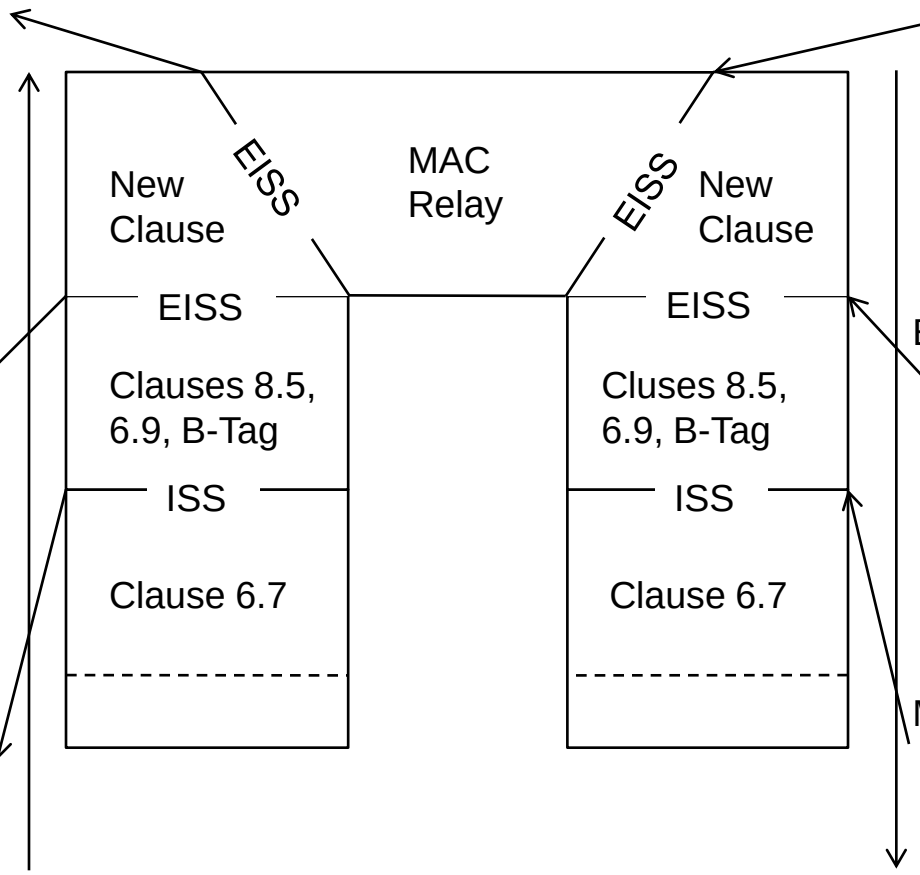
- Pros
  - Most efficient and compact encapsulation
  - 6-bit TTL accommodates both SP and DC applications
  - It is based on 802.1ah frame format that many vendors and providers are familiar with
  - Use of flow ID avoids any deep packet inspection in the BCBs
  - ECMP frames can optionally be sent w/o B-tag resulting in the most efficient encapsulation
- Cons
  - Reduces the Service ID field to 20 bits (but this doesn't create any inter-op issue as we will see)

Modified Baggy Pants Diagram for only TTL processing at BCB

EM_UNITDATA.indication (
- destination_address,
- source_address,
- mac_service_DU,
- priority,
- drop_eligible,
- vlan_identifier,
- flow_identifier,
- etc.)

EM_UNITDATA.indication (
- destination_address,
- source_address,
- mac_service_DU,
- priority,
- drop_eligible,
- vlan_identifier, etc.)

M_UNITDATA.indication (
- destination_address,
- source_address,
- mac_service_DU,
- priority,
- drop_eligible, etc.)



EM_UNITDATA.request (
- destination_address,
- source_address,
- mac_service_DU,
- priority,
- drop_eligible,
- vlan_identifier,
- flow_identifier,
- etc.)

EM_UNITDATA.request (
- destination_address,
- source_address,
- mac_service_DU,
- priority,
- drop_eligible,
- vlan_identifier, etc.)

M_UNITDATA.request (
- destination_address,
- source_address,
- mac_service_DU,
- priority,
- drop_eligible, etc.)

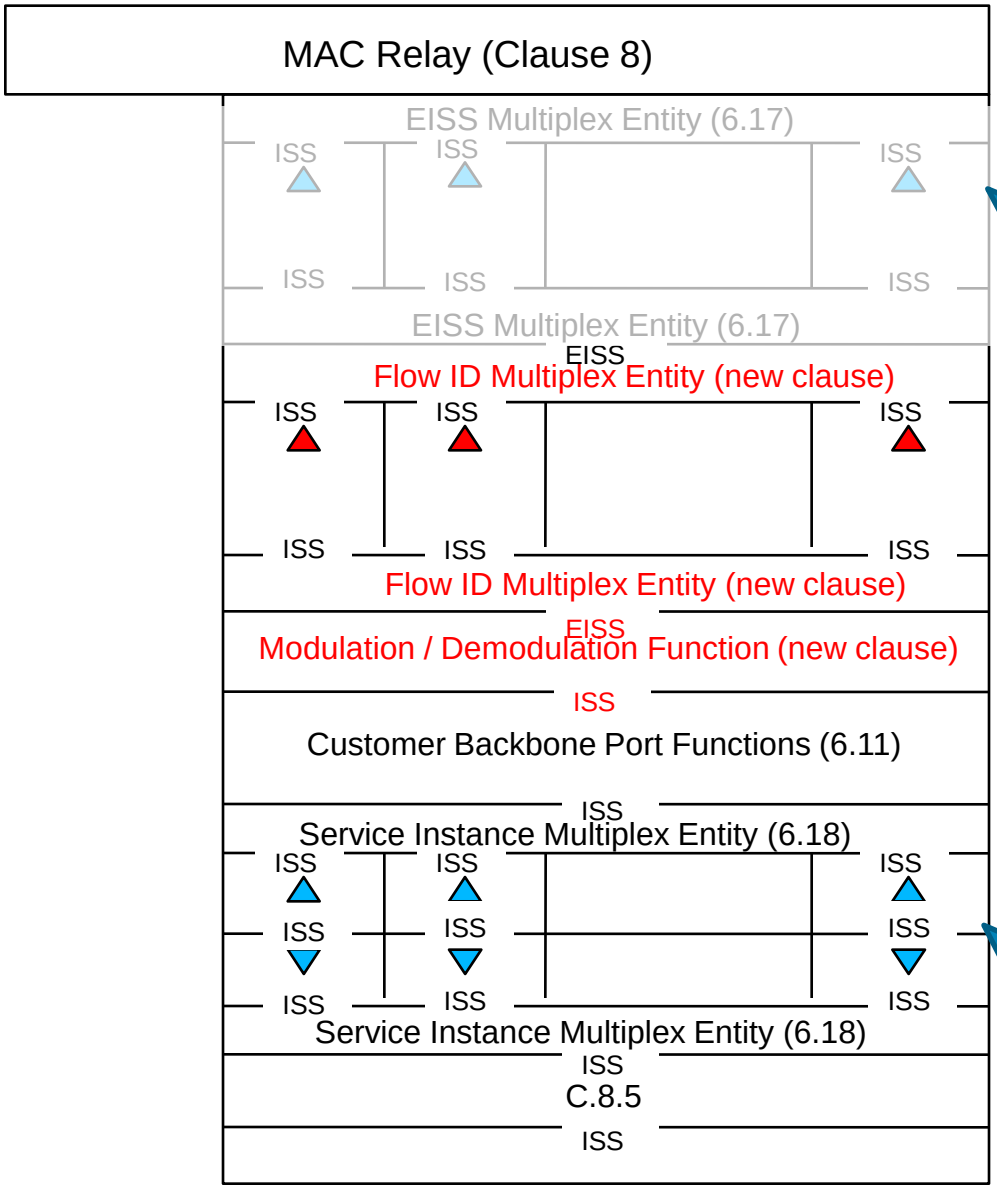
New Clause

- If next tag (after B-tag/S-tag) is E-tag, then extract TTL and flow_identifier and perform the following functions
- Use the flow_identifier in the MAC relay to select among ECMPs
- Use TTL to perform loop mitigation as follow:
  - Upon receiving TTL, if zero then discard the frame; otherwise, decrement TTL and process the frame
  - After decrementing TTL, if TTL==0 and UCA==0, then perform OAM processing
  - When setting TTL for unicast frames, it should be set to more than the min. required to accommodate re-forwarding during failure scenarios
  - When setting TTL for multicast frames, it should be set to the longest branch in the multicast tree plus a delta

New Clause – Cont.

- Flow-id is calculated and passed as a parameter of EISS API to MAC relay
- The MAC relay filtering database is enhanced so that for MAC addresses that correspond to ECMPs, it maintains several interface IDs for each MAC address since different ECMPs can take different interfaces.
- The MAC relay uses the flow-id to hash among different interface IDs for a given MAC address and select one of them

Baggy Pants Model for OAM operation at BEB



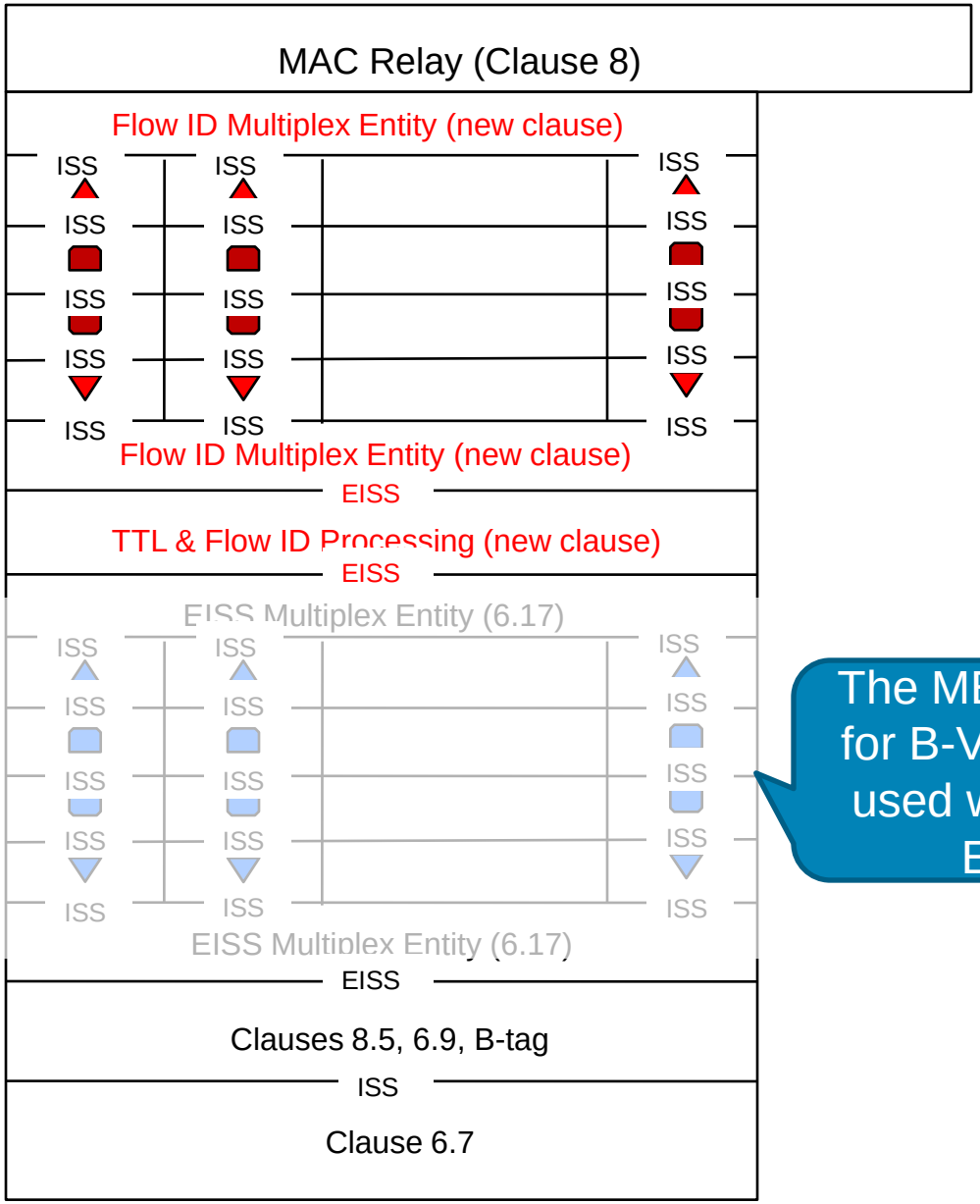
The MEPs for B-VIDs are not used when doing ECMP

E-SID MEPs requires enhancement to perform round-robin of CCs among different flows for a given E-SID

Baggy Pants Model for OAM at BEB – Cont.

- The reason for having the flow MEPs at CBP instead of PIP is:
  a) CBP needs to receive un-modulated E-SID in order to perform functions of clause 6.11
  b) But if E-SID is un-modulated, then there is no flow-id and thus there can be no MIPs functions on interim BCBs
- There is no need to configure MEPs for B-VIDs because for ECMP operation, a given B-VID identifies a specific ECMP algorithm and not a broadcast domain!!
  a) A traditional MEP on B-VID can only monitor a single path among many possible paths for that B-VID at the presence of ECMP
  b) If using a single ECMP algorithm network wide, then the use of B-VID is optional
- E-SID MEPs require additional enhancement to transmit CC messages on a round robin among different flows for a given E-SID

Baggy Pants Diagram for OAM operation at BCB



OAM Granularity: Network, Service & Flow

- Network OAM: OAM functions performed on a Test VLAN. Test Flows are chosen to exercise all ECMPs for the Test VLAN.
- Service OAM: OAM functions performed on the user VLAN itself. Test Flows are chosen to exercise all the ECMPs.
- Flow OAM: OAM functions performed on the user Flows.

Flow OAM (reactive)

- User supplies flow information, including one or more of:
  - •MAC SA and/or DA
  - •IP Src and/or Dst
  - •Src and/or Dst Port (TCP or UDP)
- Flow parameters are converted to a flow ID (e.g., NMS can query platform using flow parameters and get back flow ID)
- MEP monitors the flow by sending periodic CCMs for that flow.
  - •Monitoring of unicast flows uses unicast CCMs
  - •Monitoring of multicast flows uses multicast CCMs

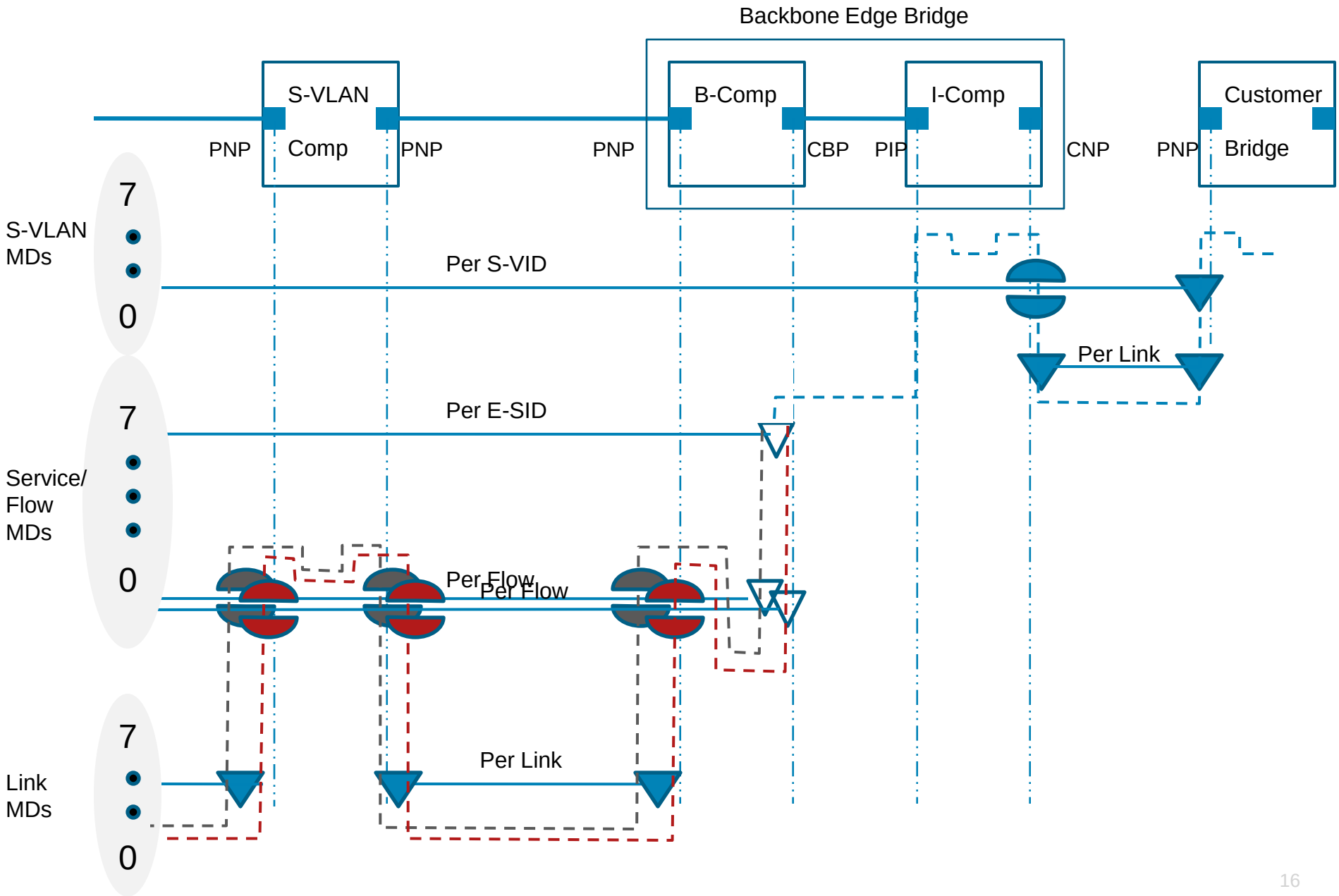
Service OAM (proactive)

- A MEP, knowing the topology and how to exercise the ECMPs, first calculates the necessary Test Flows for full coverage of all paths in a given service instance.
- On a per service instance basis, MEPs perform monitoring of all unicast and multicast paths using the Test Flows.
- MEPs follow a ‘round-robin of Test Flows’ scheme to verify connectivity over all ECMP paths (unicast) and shared trees (multicast).
  - –Round-robin scheduling reduces processing burden on nodes, and modulates the volume of OAM messaging over the network.
  - – Comes at the expense of relatively longer fault detection time
  - – For critical flows, it is possible to schedule their connectivity check continuously.
  - – MEP CCDB will track every flow independently (timer per flow per remote MEP rather than per remote MEP in CFM)

Network OAM (Proactive)

- Network OAM is a degenerate case of service OAM where a single default E-SID can be configured on all BEBs and the CFM is performed for that default E-SID just as described above for service-level OAM
  - This default E-SID is per B-VID – e.g., per ECMP algorithm. If there are multiple ECMP algorithms in the network and the E-SIDs are divided among these algorithms, then one default E-SID is needed per E-SID group (e.g., per B-VID).
  - Typically there is only a single ECMP algorithm

CFM Flow


Backbone Edge Bridge
S-VLAN
Comp
B-Comp
I-Comp
Customer
Bridge
PNP
PNP
PNP
CBP
PIP
CNP
PNP
S-VLAN
MDs
7
0
Per S-VID
Per Link
Service/
Flow
MDs
7
0
Per E-SID
Per Flow
Per Flow
Link
MDs
7
0
Per Link


CISCO