CFM for ECMP

Ali Sajassi

November 11, 2010

IEEE 802.1 Plenary – Dallas, Tx

Objectives

- To extend 802.1ag to do per flow CFM (for ECMP w/ TTL)
  - In Scope: Fault Detection, Fault Verification, Fault Isolation
  - Out of Scope: Fault Notification (AIS), Fault Recovery

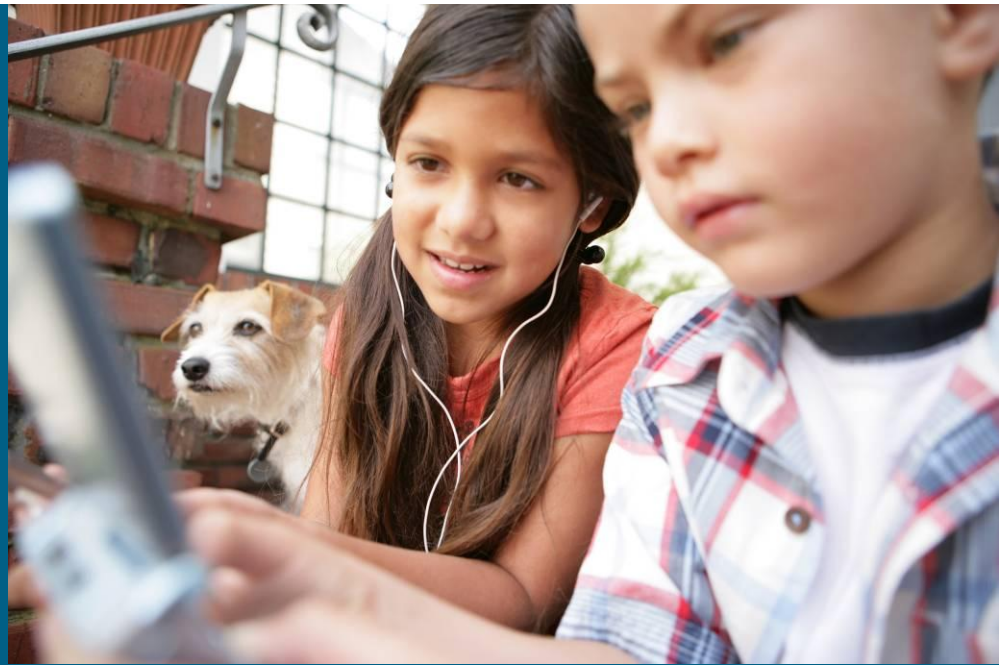
OAM Granularity: Network, Service & Flow

- 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.

Dealing with Multi-Pathing

- Load-balancing can be based on L2, L3 and/or L4 header.
- For OAM function to be truly in-band, OAM frames must exhibit the same set of L2 thru L4 fields - to be subjected to the same LB hashing function.

Fault Detection



Mechanism – 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)
- 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

Mechanism – 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 the service.
- On a per VLAN 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 N flows, and transmission rate of M flows/test-period, it takes N/M test-periods to retest a flow and a constant multiple of that to detect fault).
  - – 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)

Maintenance Point Identification, Flow Identification & Test Flows

- Flows are assigned local numeric IDs (FlowID), significant only per originating MEP.
- Test Flows:
  - Construct customer Test MACs via well-known (reserved) OUI that all bridges understand, and diversify the non-OUI portion of the MAC address (both DA and SA) to provide ECMP coverage.
  - Use the 127.0.0.0/8 IP address range for test IP addresses.
  - Source / Destination Port IDs can be wildcard.

Determining Test Flows

How to determine the required Test Flows on initiating MEP?

- Unicast:
  - –If the LB algorithm is homogeneous in the network: originating MEP knows topology & knows the hashing algorithm, hence can generate the Test Flows. In order to avoid polarization, LB algorithm would use System ID to create per-node variation. Originating MEP would know the system IDs of other bridges via the IS-IS CP.
- Multicast:
  - –Originating MEP determines flows from local LB algorithm & knowledge of shared trees

Identifying CCM Frames

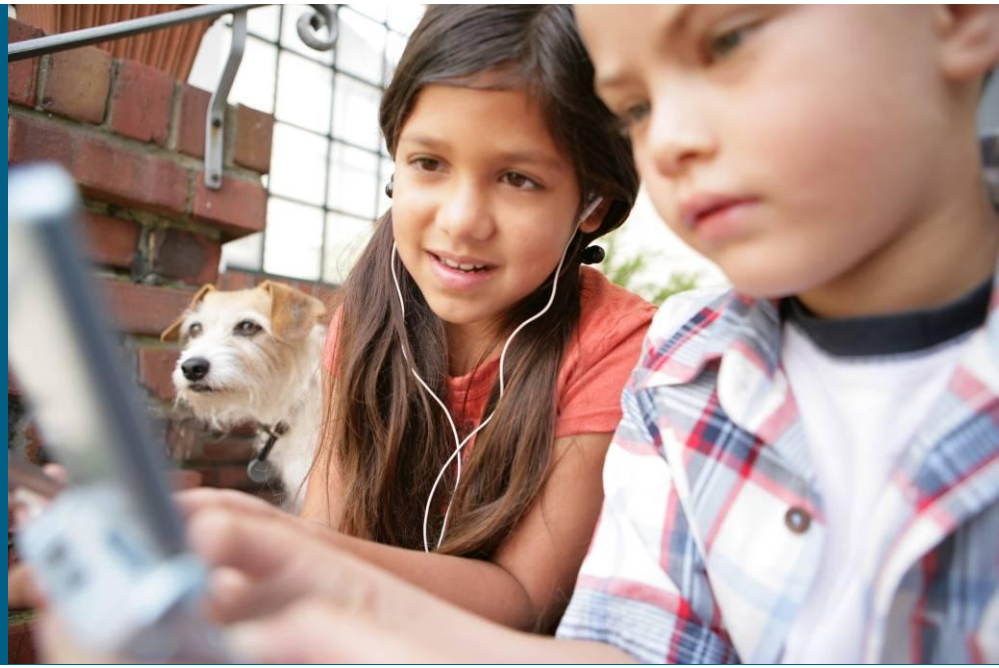
Unicast:

- CCM has [Test MAC, IP, Port] (service OAM) or same [L2, L3, L4] fields as data packet (flow OAM).
- CCM identified on egress bridge using OAM bit && Egress bridge MAC.

Multicast:

- CCM has [Test MAC, IP, Port] (service OAM) or same [L2, L3, L4] fields as data packet (flow OAM), including multicast MAC-DA.
- CCM identified on egress bridge using OAM bit && multicast MAC-DA

Fault Verification



LoopBack Mechanism

- LB to MIP and MEP are both required for fault verification.
- LB to a MP is only valid per flow, not per service because of ECMP.
  - –If MAC DA is that of MP (i.e. bridge brain MAC), then Ping may take different path from normal data. This makes the Ping useless, as it is not verifying connectivity inband. Hence, inner MAC DA must identify the flow being verified (i.e. test flow or customer MAC)
- MAC-DA should always be set to Egress bridge for flow (regardless of whether LB is to MEP or MIP).
  - –If proactive monitoring enabled, then edge bridges will likely have the MAC entry in FDB, and Ping can be issued from MEP located on the edge port.
  - –For reactive troubleshooting, MAC entry in FDB of edge bridges will most likely have expired. In that case, user must supply network operator with ingress/egress points of flow (i.e. ingress and egress bridges). Ping, in this case, is issued from MEP on the network-facing interface.
- LB to MIP/MEP is done by setting TTL to expire on the desired MIP/MEP.

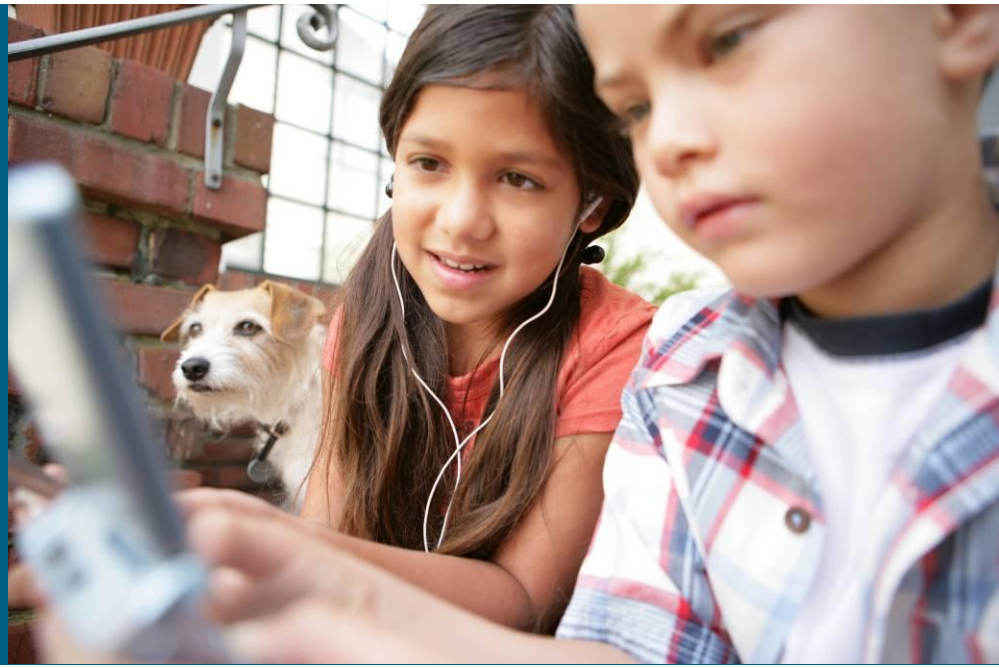
Identifying LB Frames

- Use a reserved bit in E-SID to identify OAM frames.
- Condition for punting to control-plane:
  - –OAM bit is set && TTL is expired

Multicast LB

- Only MEPs must reply. Transparent to MIPs.
  - –Performance management (delay measurement)
  - –E-Tree service (single-sided management from root perspective- can log onto root and know from there if service is healthy)
  - –In order to ping a transient node, it must be configured as an egress bridge (bud).
- Destination needs to delay response to avoid reply storms (some hysteresis timer)
  - –Optional (default no delay)

Fault Isolation



Fault Isolation

- To identify the location of a fault for a given flow that is previously detected by CCM
  - Send a LB message for that flow with TTL=1
  - Increment TTL, and send a LB message for that flow. Repeat this step till the location of fault is determined