TOMORROW starts here.
Advanced VPLS attachment technology options

BRKRST-2008

Patrice Bellagamba
Distinguished SE
Cisco Europe
Problem statement

- Problem to solve:
  "Ethernet service Attachment-Circuit redundancy should not imply end to end Spanning-tree usage"

- STP is struggling with
  - Topology diameter
  - Fault isolation between sites
  - Resilience to WAN type connections

- Position Architecture: STP domain isolation thru dual homing
  - Carrier-Ethernet
  - Data-Center Interconnect
  - Ethernet Services
## Access Multi-homing Solutions Summary

<table>
<thead>
<tr>
<th>Highlights</th>
</tr>
</thead>
</table>
| **Multi-chassis LAG** | Simple solution for spoke-and-hub topology, works for both bridging and non-bridging access device.  
Standard based solution by using 802.3ad.  
Sub-second convergence.  
Phase 1 implement is active/standby mode.  
Phase 2 is per VLAN load balancing.  
Phase 3 is per flow load balancing with E-VPN. |
| **REP /REP access gateway** | Sub 200msec convergence.  
Good access ring isolation.  
Now standard based → G.8032 (XR4.1 release).  
Spoke-and-hub and ring topology, not works well for mesh network. |
| **MST/PVST access gateway** | Standard based solution as long as access network support MST/PVST.  
Works for any access network topology.  
Good access domain isolation.  
Work with 802.1ah PBB.  
Convergence time depends on access network STP. |
| **Node clustering** | VSS (Catalyst / Cisco 7600-Sup2T).  
Nv cluster (ASR9k).  
One control-plane for two chassis.  
Easiness, Active/Active. |
## L2 node peering solutions

Making usage of an “Inter-Chassis Communication Protocol”

<table>
<thead>
<tr>
<th>Service Type</th>
<th>L2VPN Transport Enabler</th>
<th>Access Redundancy</th>
<th>ICCP based Protocol / Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-LINE</td>
<td>VPWS</td>
<td>Hub and Spoke (Active / Backup)</td>
<td>mLACP + 2-way PW Red. (coupled mode)</td>
</tr>
<tr>
<td>E-LAN</td>
<td>VPLS</td>
<td>Hub and Spoke (Active / Backup)</td>
<td>mLACP + 2-way PW Red. (decoupled mode)</td>
</tr>
<tr>
<td>E-LAN</td>
<td>H-VPLS</td>
<td>Hub and Spoke (Active / Backup)</td>
<td>mLACP + 2-way PW Red. (decoupled mode)</td>
</tr>
<tr>
<td>DCI</td>
<td>VPLS</td>
<td>CPE service (Active per Service)</td>
<td>mLACP or pseudo-mLACP</td>
</tr>
<tr>
<td>DCI</td>
<td>EVPN</td>
<td>PE service (Active per Flow)</td>
<td>mLACP + BGP</td>
</tr>
</tbody>
</table>

*Don’t be scared of all acronyms, they will be explained along the presentation*
Agenda

- Technology introduction
  - Inter-Chassis Communication Protocol (ICCP)
  - Multi-Chassis Link Aggregation Group (MC-LAG)
- Implementing MC-LAG
  - E-LAN dual-homing
  - E-LINE dual-homing
  - Data-Center Interconnect dual-homing
- Active/Active MC-LAG or ICCP-SM
- Flow-based E-VPN
Inter-Chassis Communication Protocol for L2VPN PE Redundancy (ICCP)
draft-martini-pwe3-iccp-12
Inter-Chassis Communication Protocol

- ICCP allows two or more devices to form a ‘Redundancy Group’
- ICCP provides a control channel for synchronizing state between devices
- ICCP uses TCP/IP as the underlying transport
  - ICCP rides on targeted LDP session, but MPLS need not be enabled
- Various redundancy applications can use ICCP:
  - mLACP
  - Pseudowire redundancy
- Under standardization in IETF:
  - draft-ietf-pwe3-iccp-12.txt
Inter-Chassis Communication Protocol
draft-ietf-pwe3-iccp-12.txt

• ICC Protocol Transport Requirements
  – Reliable Message exchange
  – In-order Delivery
  – Sequence Numbers
  – Timeouts/Retransmissions
    ➔ use widely deployed LDP protocol.

• Extend LDP with a small set of new messages:
  – RG Connect Message
  – RG Disconnect Message
  – RG Notification Message
  – RG Application Data Message

• Use LDP Capability to bootstrap ICCP.
• Application layer specific TLVs.
Two mechanisms:

- **BFD – Bi-Directional Forward Detection**
  - Detection upon loss of BFD keepalives
  - Requires nodes to be co-located, with a direct link connection
  - No split-brain protection, mandates link to be port-channel dispatched over two different line cards

- **/32 IP Route-watch**
  - Detection upon loss of IP routing adjacency
  - Geo localization of nodes
  - Split-brain tie-break via MPLS network
  - Depends on IGP timers
  - OSPF/ISIS fast convergence tuning is required
  - Ensure not to have less specific route for detection to work
  - "ip routing protocol purge interface"
    - This is requiring to avoid route-watch flapping on peer link failure
      - Default in IOS-XR & IOS Rls15
Multi-Chassis Link Aggregation Group (ASR 9K - MC-LAG)
&
Multi-Chassis Link Aggregation control Protocol (Cisco 7600 - mLACP)
Background: Link Aggregation Control Protocol

- **System attributes:**
  - System MAC address: MAC address that uniquely identifies the switch
  - System priority: determines which switch’s Port Priority values win

- **Aggregator (bundle) attributes:**
  - Aggregator key: identifies a bundle within a switch (per node significance)
  - Maximum links per bundle: maximum number of forwarding links in bundle – used for Hot Standby configuration
  - Minimum links per bundle: minimum number of forwarding links in bundle, when threshold is crossed the bundle is disabled

- **Port attributes:**
  - Port key: defines which ports can be bundled together (per node significance)
  - Port priority: specifies which ports have precedence to join a bundle when the candidate ports exceed the Maximum Links per Bundle value
  - Port number: uniquely identifies a port in the switch (per node significance)
Extending LACP Across Multi-Chassis: mLACP

- mLACP uses ICCP to synchronize LACP configuration & operational state between PoAs, to provide DHD the perception of being connected to a single switch.
- All PoAs use the same System MAC Address & System Priority when communicating with DHD.
  - Configurable or automatically synchronized via ICCP.
- Every PoA in the RG is configured with a unique Node ID (value 0 to 7). Node ID + 8 forms the most significant nibble of the Port Number.
- For a given bundle, all links on the same PoA must have the same Port Priority.

DHD = Dual-Homed-Device

PoA = Point of Aggregation
Operational Variants
PoA-Based Control

- Each PoA is configured to limit the maximum number of links per bundle
  - Limit must be set to L, where L is the minimum number of links from DHD to any single PoA
- DHD max link should be set > L
  - In order to insure that it is slave of the POA
  - This will allow faster convergence
- Selection of active/standby links is the responsibility of the PoAs
- Advantages: Faster switchover times compared to other variants, and Minimum Link policy on PoA can be flexible
- Disadvantage: If ICCP transport is lost, Split Brain condition could occur

This is the most used variant
mLACP Offers Protection Against 5 Failure Points:

- A: DHD Port Failure
- B: DHD Uplink Failure
- C: Active PoA Port Failure
- D: Active PoA Node Failure
- E: Active PoA Isolation from Core Network
Failover Operation
Port/Link Failures

**Step 1** – For port/link failures (A,B,C), active PoA evaluates number of surviving in bundle:
- If >= M, then no action
- If < M, then trigger failover to standby PoA

**Step 2** – Active PoA signals failover to standby PoA over ICCP

**Step 3** – Failover is triggered on DHD by one of:
- **Dynamic Port Priority Mechanism:** real-time change of LACP Port Priority on active PoA to cause the standby PoA links to gain precedence
- **Links are either Hot-Standby or Up**
- **Brute-force Mechanism:** change the state of the surviving links on active PoA to admin down
- **Links are either Err-disabled or Up**

**Step 4** – Standby PoA and DHD bring up standby links per regular LACP procedures
Failover Operation

Node Failure

Step 1A – Standby PoA detects failure of Active PoA via one of:
- IP Route-watch: loss of IP routing adjacency
- BFD: loss of BFD keepalives

Step 1B – DHD detects failure of all its uplinks to previously active PoA

Step 2 – Both Standby PoA and DHD activate their Standby links per regular LACP procedures
Failover Operation
PoA Isolation from Core

Step 1 – Active PoA detects all designated core interfaces are down

interchassis group 21
backbone interface TenGigabitEthernet4/1
backbone interface TenGigabitEthernet1/4

Really useful if no direct connection between POA or using one only module toward core

Step 2A – Active PoA signals standby PoA over ICCP to trigger failover

Step 2B – Active PoA uses either Dynamic Port Priority or Brute-force Mechanism to signal DHD of failover

Step 3 – Standby PoA and DHD bring up standby links per regular LACP procedures
E-LAN Availability Models
Active/Backup Access Node Redundancy (mLACP)

Works similarly with H-VPLS (MPLS Access)
E-LAN availability model
Active / Backup Access Node Redundancy (mLACP)

- Port / Link Failures

- For VPLS Decoupled Mode, VFI’s PWs always advertised in Active state, regardless of AC state

<table>
<thead>
<tr>
<th>Events</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Initial state</td>
</tr>
<tr>
<td>F_{A-C}</td>
<td>Port / Link Failures</td>
</tr>
<tr>
<td>1_A</td>
<td>Active PoA detects failure and signals failover over ICCP</td>
</tr>
<tr>
<td>1_B</td>
<td>Failover triggered on DHD</td>
</tr>
<tr>
<td>2</td>
<td>Standby link brought up per LACP proc.</td>
</tr>
<tr>
<td>3</td>
<td>Standby PoA flushes MAC table and triggers LDP MAC add. withdrawal to remote peers</td>
</tr>
<tr>
<td>4</td>
<td>Remote PEs flush MAC addresses</td>
</tr>
</tbody>
</table>

PW - PseudoWire
VFI - Virtual Forwarding Instance
Port / Link Failures (cont.)

Events

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Remote PEs flush MAC addresses</td>
</tr>
<tr>
<td>E</td>
<td>End State</td>
</tr>
</tbody>
</table>

Non-Forwarding EoMPLS PW
Forwarding EoMPLS PW
E-LINE Availability Models
Active/Backup Access Node Redundancy (mLACP)
### Pseudowire Redundancy with LDP

#### Background

- Designate Pseudowires as either primary or backup
  - Primary Pseudowire used for traffic forwarding, and backup takes over in case of failure

- Signaling Redundant Pseudowires in targeted LDP session
  - Cold Redundancy: Backup PWs were not signaled until required to take over
  - Warm Redundancy: Backup PWs were signaled up in the control-plane but held down in the data-plane. Use AC Fault code-point in LDP Status Message to indicate a backup PW
  - Hot Redundancy: Use PW Preferential Forwarding Status Bits
Two-Way Pseudowire Redundancy

Overview

- Allows dual-homing of two local PEs to two remote PEs
- Four pseudowires: 1 primary & 3 backup provide redundancy for a dual-homed device
Pseudowire Redundancy with LDP

PW Status Signaling

- 0x00000000 - Pseudowire forwarding (clear all failures)
- 0x00000001 - Pseudowire Not Forwarding
- 0x00000002 - Local Attachment Circuit (ingress) Receive Fault
- 0x00000004 - Local Attachment Circuit (egress) Transmit Fault
- 0x00000008 - Local PSN-facing PW (ingress) Receive Fault
- 0x00000010 - Local PSN-facing PW (egress) Transmit Fault
- 0x00000020 - PW Forwarding Standby
- 0x00000040 – PW Request to go Active

Only this bit is required/used (with help of ICCP)

RFC 4447

draft-ietf-pwe3-redundancy-bit
Two-Way Pseudowire Redundancy
Independent Operation Mode

- Every PE decides the local status of the PW: Active or Standby
- A PW is selected as primary for forwarding if it is active on both local & remote PEs
- A PW is considered as backup if it is declared as Backup by either local or remote PE
Two-Way Pseudowire Redundancy

Determining Pseudowire State

- **VPWS / H-VPLS** – two-way coupled:
  - When AC changes state to Active\(^1\), both PWs will advertise Active
  - When AC changes state to Standby\(^1\), both PWs will advertise Standby

- **H-VPLS** – two-way decoupled:
  - Regardless from AC state, Primary PW and Backup PWs will advertise Active state
  - ASR9K implementation

- For H-VPLS, all PWs in VFI (at nPE) are Active simultaneously, for both access & core PWs

(1) Active / Standby AC states determined for example by mLACP
E-LINE availability model
Active / Backup Access Node Redundancy (mLACP)

- Port / Link Failures

- For VPWS Coupled Mode, attachment circuit (AC) state (Active/Standby) drives PW state advertised to remote peers

<table>
<thead>
<tr>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Initial state</td>
</tr>
<tr>
<td>1A Port / Link Failures</td>
</tr>
<tr>
<td>1B Active PoA detects failure and signals failover over ICCP</td>
</tr>
<tr>
<td>2 Standby link brought up per LACP proc.</td>
</tr>
<tr>
<td>3 Standby PoA advertises “Standby” state on its PWs</td>
</tr>
<tr>
<td>4 Standby PoA advertises “Active” state on its PWs</td>
</tr>
</tbody>
</table>

- For VPWS Coupled Mode, attachment circuit (AC) state (Active/Standby) drives PW state advertised to remote peers
E-LINE availability model
Active / Backup Access Node Redundancy (mLACP)

- Port / Link Failures (cont.)

- Local site access failure does not trigger LACP failover at remote site (i.e. control-plane separation between sites)

**Events**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Initial state</td>
</tr>
<tr>
<td>F&lt;sub&gt;A-C&lt;/sub&gt;</td>
<td>Port / Link Failures</td>
</tr>
<tr>
<td>1&lt;sub&gt;A&lt;/sub&gt;</td>
<td>Active PoA detects failure and signals failover over ICCP</td>
</tr>
<tr>
<td>1&lt;sub&gt;B&lt;/sub&gt;</td>
<td>Failover triggered on DHD</td>
</tr>
<tr>
<td>2</td>
<td>Standby link brought up per LACP proc.</td>
</tr>
<tr>
<td>3</td>
<td>Active PoA advertises “Standby” state on its PWs</td>
</tr>
<tr>
<td>4</td>
<td>Standby PoA advertises “Active” state on its PWs</td>
</tr>
<tr>
<td>E</td>
<td>End State</td>
</tr>
</tbody>
</table>
Implementing MC-LAG
Attachment Circuit Using mLACP/MC-LAG

Multi-Chassis LACP synchronization:
- LACP BPDUs (01:80:C2:00:00:00) are exchanged on each Link
- System Attributes: Priority + bundle MAC Address
- Port Attributes: Key + Priority + Number + State

Terminology:
- mLACP: Multi-Chassis Link Aggregation Control Protocol
- MC-LAG: Multi-Chassis Link Aggregation Group
- DHD: Dual Homed Device (Customer Edge)
- DHN: Dual Homed Network (Customer Edge)
- POA: Point of Attachment (Provider Edge)
Additional Features for IOS-XR

- Configurable ‘wait-while’ timer
  - Timer used to bulk member aggregation by delaying initial link state transitions within a bundle (standards set a time of 2 s)

- Configurable flap suppression timer
  - Timer used to allow a bundle time to move links from Standby into Active state (i.e. bring in replacement links) before determining whether the bundle has gone down

- Configurable delay for revertive switchbacks
  - Allows other applications time to complete a sync of their data
  - Useful for ‘Revertive’ behavior which would normally make the highest priority device Active as soon as possible

- LACP level bundle shutdown
  - Method for costing out a bundle while allowing a quick bring up
  - LACP continues to be performed but...
    - Links are kept in Standby state
    - The bundle is Down to other applications
Active/Standby Deployment Options

- Traffic may take sub-optimal path via the DHD inter chassis link
- Optimal path by bundling 4 links in vPC
- Direct path exist between each N7K and active PE
Only error 2/3/4 are leading to ICCP convergence
Rem: 2 & 4 are dual errors

- 500 VLAN Unicast: *Link error sub-1s & Node error sub-2s*
- 1200 VLAN unicast: *Link error sub-2s & Node error sub-4s*
Local Connect

- Local connect service between bundle port and bundle port, or between bundle port and regular port are configurable. And it works in the normal condition.

- However, if bundle failover happen, then it won’t work. For example, for CE2, if it failover to the bottom POA, then there is no local connection to CE1 or CE3.

- In summary, although you can configure the local connect service for the MC-LAG bundle port, but it could cause problem. So it’s not supported.

- Local bridging is supported assume there is L2 link between two POA.

Local Bridging
### mLACP Platform Support

<table>
<thead>
<tr>
<th>Feature</th>
<th>Cisco Catalyst 3750-ME</th>
<th>Cisco ME 3600X / 3800X</th>
<th>Cisco ASR 901/903</th>
<th>Cisco 7600</th>
<th>Cisco ASR 9000</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-Way PW Redundancy</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Two-Way PW Redundancy</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>LDP MAC Address Withdrawal</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>LACP</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>mLACP</td>
<td></td>
<td></td>
<td>X (903)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ICCP</td>
<td></td>
<td></td>
<td>X (903)</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Pseudo mLACP

ICCP-SM - Service Multi-Homing
Pseudo MLACP (P-mLACP) concept

- P-mLACP provides an Active / Active dual homing redundancy mechanism with per-VLAN load balancing
  - higher bandwidth utilization
- Backup time does not require link state changes
  - Improve scale and convergence time
- Supports today Dual Homed Device (DHD)
  - Will support Dual Homed Network (DHN) in future
- New extensions to Inter-Chassis Communication Protocol (ICCP) used for communication to control the failover process
- Available in Cisco 7600 with 15.1(3)S release mandates ES / ES+ linecards
- Available in ASR 9000 with 4.3.1 released in Q2 CY13
Pseudo MLACP (P-mLACP) concept (cont.)

P-mLACP provides VLAN based redundancy by allowing one to configure one primary and one secondary interface pair for each member VLAN.

Dual Homed Device (DHD) configured with two separate port-channels aggregating to one LAG on Point of Attachment (PoAs).
- Pseudo-LACP application on PoAs ensure that two port-channels on DHD will get bundled to one logical port-channel on PoA.
Attachment Circuit options

Mono-Device Dual Homed Device (DHD)

Multi-Devices Dual Homed Device (DHD)
⇒ The one analyzed in this presentation

Dual Homed Network (DHN)
• 7600 & ASR9K do support MVRP
• it would require DHN to support MVRP to flush local MAC
• ASR9K supports STP TCN to flush edge MAC
• 4K VLAN Scale (no multi-chassis etherchannel)
Active path failures

A = Nexus port failure (shut)
B = Attachment-Circuit failure
C = Dual AC failure
D = PE crash
E = MPLS link failure
N = DHD member crash

A = Nexus port failure (shut)
B = Attachment-Circuit failure
C = Dual AC failure
D = PE crash
E = MPLS link failure
N = DHD member crash
Failure condition and MAC flush consideration
MAC learning on nominal mode

LAG  PW
Forwarding  Blocking
Failure condition and MAC flush consideration
MAC learning on PE failure
Failure condition and MAC flush consideration

MAC learning on PE up time

No MVRP on Nexus 7000
Solved with ASR9K using a TCN instead

Automatic reversion
- Manual procedure
## Active path failures measurement

<table>
<thead>
<tr>
<th>Failure type</th>
<th>Traffic ➔</th>
<th></th>
<th>Traffic ←</th>
<th></th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Down</td>
<td>Up</td>
<td>Down</td>
<td>Up</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>DHD link shut</td>
<td>0.5s-1s</td>
<td>1.5s</td>
<td>&lt;0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>C</td>
<td>PE link shut</td>
<td>0.8s</td>
<td>1.8s</td>
<td>&lt;0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>D</td>
<td>PE crash</td>
<td>0.8s</td>
<td>1.7s</td>
<td>0.8s</td>
<td>1.7s</td>
</tr>
<tr>
<td>E</td>
<td>MPLS link down</td>
<td>0.1s</td>
<td>&lt;0.1s</td>
<td>0.1s</td>
<td>&lt;0.1s</td>
</tr>
<tr>
<td>N</td>
<td>DHD crash</td>
<td>0.3s</td>
<td>&lt;0.1s</td>
<td>0.3s</td>
<td>&lt;0.1s</td>
</tr>
</tbody>
</table>

Medium scale = 1000 VLAN per vPC *2
ICCP Based Service Multihoming

(ICCP-SM) = pseudo mlacp for ASR9000

- 5-tuple hashing is used for load balancing among bundle member links, using the following parameters:
  - IP source address
  - IP destination address
  - Router ID
  - Layer 4 source port
  - Layer 4 destination port

- Applications running on the two POAs (mLACP, IGMP snooping, DHCP snooping or ANCP (Access Node Control Protocol)) synchronize their state using ICCP.

- Ethernet CFM on Link Aggregation Group (LAG) interfaces (Ethernet bundle interfaces), Ethernet and bundle subinterfaces, and LAG member (bundle member) interfaces.
PoA1 Config

iccp group 100
redundancy
iccp
group 100
member
neighbor 222.222.222.222
!
backbone
interface GigabitEthernet0/0/0/10
interface GigabitEthernet0/0/0/11
interface GigabitEthernet0/0/0/12
!
isolation recovery-delay 30
!

PoA2 Config

iccp group 100
redundancy
iccp
group 100
member
neighbor 111.111.111.111
!
backbone
interface GigabitEthernet0/0/0/10
interface GigabitEthernet0/0/0/11
interface GigabitEthernet0/0/0/12
!
isolation recovery-delay 30
!

Scale Numbers - 30 ICCP Groups, 240 Bundle Interfaces, 4K VLANs, 4K Sub-Interfaces per Bundle
PoA1 Config

l2vpn
nsr
redundancy
  iccp group 100
  multi-homing node-id 1
interface Bundle-Ether100
  primary vlan 1-100
  secondary vlan 101-200
  recovery delay 60
bridge group bg100
bridge-domain bd100-1
mac
  withdraw state-down
  limit
    maximum 512000
!
interface Bundle-Ether100.1
!
vfi v100-1
  neighbor 55.55.55.55 pw-id 1001
!
neighbor 66.66.66.66 pw-id 1001
!

PoA2 Config

l2vpn
nsr
redundancy
  iccp group 100
  multi-homing node-id 2
interface Bundle-Ether100
  primary vlan 101-200
  secondary vlan 1-100
  recovery delay 60
bridge group bg100
bridge-domain bd100-1
mac
  withdraw state-down
  limit
    maximum 512000
!
interface Bundle-Ether100.1
!
vfi v100-1
  neighbor 55.55.55.55 pw-id 1001
!
neighbor 66.66.66.66 pw-id 1001
!
Normal Operation – No Failures

RP/0/RSP0/CPU0:ICCPSM1#show l2vpn iccp-sm group 100

ICCP-based Service Multi-Homing

Group ID: 100, State: Synchronized with Peer
Local Node ID: 1, Remote Node ID: 2,
   ICCP State: Transport Up, Member Up

Interface Name: Bundle-Ether100
MAC flushing: MVRP
Recovery Delay: 60 (Timer not running)
   Local State: Operational
Remote State: Operational

<table>
<thead>
<tr>
<th>Local</th>
<th>VLAN IDs</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary:</td>
<td>1-100</td>
<td>Forwarding</td>
</tr>
<tr>
<td>Secondary:</td>
<td>101-200</td>
<td>Blocked</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Remote</th>
<th>VLAN IDs</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary:</td>
<td>101-200</td>
<td>Forwarding</td>
</tr>
<tr>
<td>Secondary:</td>
<td>1-100</td>
<td>Blocked</td>
</tr>
</tbody>
</table>
Remote PoA is DOWN (D) Failure

RP/0/RSP0/CPU0:ICCPSM1#show l2vpn iccp-sm group 100

ICCP-based Service Multi-Homing

Group ID: 100, State: Provisioned
Local Node ID: 1, Remote Node ID: 2,
   ICCP State: Transport Up, Member Down

Interface Name: Bundle-Ether100
   MAC flushing: MVRP
Recovery Delay: 60 (Timer not running)
   Local State: Operational
   Remote State: unknown

<table>
<thead>
<tr>
<th>Local VLAN IDs</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary: 1-100</td>
<td>Forwarding</td>
</tr>
<tr>
<td>Secondary: 101-200</td>
<td>Forwarding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Remote VLAN IDs</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary: 101-200</td>
<td>Blocked</td>
</tr>
<tr>
<td>Secondary: 1-100</td>
<td>Blocked</td>
</tr>
</tbody>
</table>

BRKRST-2008  © 2014 Cisco and/or its affiliates. All rights reserved.  Cisco Public
Troubleshooting Information

Trouble Shooting - Info to collect:

Running-config
"show l2vpn iccp-sm private"
"show iccp group <>"
"show iccp trace"
"show l2vpn trace"
“show l2vpn forwarding mstp detail location 0/x/CPU0”
“show uidd data location 0/x/CPU0 bundle-ether x.y ingress | incSTP”
Attachment Circuit A/A for Ethernet-VPN
E-VPN and PBB-EVPN (aka Routed VPLS)
Main Principles

- Next generation L2 VPN to extend Layer 2 connectivity across MPLS
- Standard based approach. IETF drafts by L2VPN Work Group
  E-VPN – draft-ietf-l2vpn-evpn
  PBB-EVPN – draft-ietf-l2vpn-pbb-evpn
- E-VPN improves upon VPLS for high scalability, and high availability
  Full-Mesh of PW no longer required
- PBB-EVPN combines Provider Backbone Bridge (PBB) 802.1ah with E-VPN to simplify core transport and enhance scalability.

Control-Plane Distribution of Customer MAC-Addresses using BGP
Supported Access Topologies

**Single Home Device (SHD)**
- Single Home Network (SHN)

**Dual Home Device (DHD)**
- Active / Active Per-Flow LB
- Dual Home Device (DHD)
- Active / Active Per-Service LB

**Dual Home Network (DHN)**
- Active / Active Per-Service LB

- Bundle Interface
- Physical Interfaces
- Bundle Interfaces (shown)
- Bundle Interfaces (shown)
- Bundle Interfaces (shown)
AAPF – Imposition
Active / Active Per Flow load balancing

- CE1 does load balancing of flows based on 5 tuples
- (SMAC/DMAC/VLAN/SIP/DIP/TCP ports / UDP ports)
- CE1 will split the customer traffic between members of same bundle
- Both PE1-A and PE1-B forward all imposition traffic
AAPF – Disposition
Active / Active Per Flow load balancing

- PE1-A (DF)
- PE1-B (non DF)
- PE2
- CE1
- Mpls Core
- CE2

FWD Unicast

DROP FLOOD to avoid Duplicates
ASR9k Internal PBB Topology

- Customer Network Port (CNP)
- Provider backbone Network Port (PNP)
- EVPN
- PW
- EFP-x
- Core BD
- Edge BD-1
- Edge BD-2
- Edge BD-n
- VIP
- System internal virtual port
- I-component
- B-component
- PW component
- B-component
- CBP

VIP

VIP

VIP

VIP

VIP

VIP
Configuration: EdgeBD / I-component

l2vpn
  bridge group PBB
  bridge-domain PBB-EDGE
  interface GigabitEthernet0/0/0/38.100
  interface Bundle-ether10.100
  neighbor 10.10.10.1 pw-id 1010
  !
  pbb edge i-sid 1200 core-bridge PBB-CORE
  !
Configuration: Core BD/ B-Component

```
l2vpn
  bridge group PBB
  bridge-domain PBB-CORE
  interface G0/5/0/10.100

pbb core
  evpn
  vpn-id 100

router bgp 200
  nsr
timers bgp 20 60
  bgp router-id 10.10.10.10
  address-family l2vpn evpn

neighbor 20.20.20.20
  remote-as 200
  update-source Loopback0
  address-family l2vpn evpn
```
Active / Active per-FLOW Load Balancing

PE1
redundancy
iccp
group 66
mlacp node 1
mlacp system mac 0aaa.0bbb.0ccc
mlacp system priority 1
mode singleton

interface Bundle-Ether25
mlacp iccp-group 66

interface Bundle-Ether25.1 l2transport
encapsulation dot1q 777

Note: MPLS / LDP configuration required on core-facing interfaces (not shown)
C-MAC table

C-MAC table is learnt on Linecard hardware when customer traffic flows
Similar to VPLS/ PBB-VPLS.

1) Edge-BD
RP/0/RSP0/CPU0:PE2#sh l2vpn forwarding bridge-domain PBB:PBB_EDGE_1011  mac-address location
(CMAC Entry learnt from edge-bd)
00bb.0100.00aa dynamic Gi0/6/0/19.101 0/6/CPU0 0d 0h 0m 17s N/A

(CMAC Entry learnt from core-bd)
00aa.0100.0000 dynamic BD id: 3 0/6/CPU0 0d 0h 0m 8s 0011.1111.0000
B-MAC table

B-MAC table is learnt by BGP, via control plane. Static Entries created in MAC table.
This is new in PBB-EVPN.

2) Core-BD

RP/0/RSP0/CPU0:PE2# sh l2vpn forwarding bridge-domain PBB:PBB_CORE1  mac-address

<table>
<thead>
<tr>
<th>Mac Address</th>
<th>Type</th>
<th>Learned from/Filtered on</th>
<th>LC learned</th>
<th>Resync Age</th>
<th>Mapped to</th>
</tr>
</thead>
<tbody>
<tr>
<td>0011.0000.0000 BMAC</td>
<td>BD id: 0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>STATIC ENTRIES (Built by control plane, does not age out)</td>
</tr>
<tr>
<td>0011.1111.0000 BMAC</td>
<td>BD id: 0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>0011.1111.1111 BMAC</td>
<td>BD id: 0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>0022.0000.0000 BMAC</td>
<td>BD id: 0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>
PBB EVPN Scale Limits

- Max 128K Bundle EFP per system
- Max 4K EFP per bundle

- Max 64K Edge-BD per system
- Max 200 Core-BD per system

- Max 2K BMACs per system
- Max 200 PE per core-BD
Conclusion
Service Multi-Homing

Building blocks

- Point to point L2 Service → PW redundancy
- Multi-point L2 Service → MC-LAG
Changing the Redundancy Paradigm

Virtual Cluster and Satellite for BNG

- Geo Redundant Dual Homing
- High Availability
- Single-Chassis-like look&feel and Management of cluster Members
- Stateful Failover between chassis
- Active/active LAG user facing

+  
- Huge 1GE Fan-out
- Satellites appear like ASR 9000 Linecards
- Simplified topology, No Spanning tree or other L2 redundancy protocols needed
Call to Action…

- Visit the **Cisco Campus** at the World of Solutions to experience the following demos/solutions in action:
  Speaker to add relevant demos/areas to visit from the campus demos list

- Get hands-on experience with the following **Walk-in Labs**
  Speaker to add the relevant walk in labs from the list

- **Meet the Engineer**
  Speaker to specify when will they be available for ad-hoc meetings at the MTE village, and provide other recommended names…

- Discuss your project’s challenges at the **Technical Solutions Clinics**

- Attend one of the **Lunch Time Table Topics**, held in the main Catering Hall

- **Recommended Reading**: For reading material and further resources for this session, please visit [www.pearson-books.com/CLMilan2014](http://www.pearson-books.com/CLMilan2014)

- **CL365** - Visit us online after the event for updated PDFs and on-demand session videos. [www.CiscoLiveEU.com](http://www.CiscoLiveEU.com)
Complete Your Online Session Evaluation

- Complete your online session evaluation
- Complete four session evaluations and the overall conference evaluation to receive your Cisco Live T-shirt