7450 Ethernet Service Switch
7750 Service Router
7950 Extensible Routing System

Releases up to 24.3.R2

Interface Configuration Advanced Configuration Guide for Classic CLI

3HE 20790 AAAA TQZZA
Edition: 01
July 2024

Nokia is committed to diversity and inclusion. We are continuously reviewing our customer documentation and consulting with standards bodies to ensure that terminology is inclusive and aligned with the industry. Our future customer documentation will be updated accordingly.

Table of contents

List of figures....4

Preface....5

Multi-Chassis APS and Pseudowire Redundancy Interworking....6

Multi-Chassis LAG and Pseudowire Redundancy Interworking....23

Port Cross-Connect (PXC)....43

List of figures

Figure 1: MC-APS network topology....7

Figure 2: Access node and network resilience (part 1)....8

Figure 3: Access node and network resilience (part 2)....8

Figure 4: Association of SAPs/SDPs and endpoints....13

Figure 5: ICB spoke SDPs and association with the endpoints....16

Figure 6: Additional setup example 1 (part 1)....19

Figure 7: Additional setup example 1 (part 2)....19

Figure 8: Additional setup example 2 (part 1a)....20

Figure 9: Additional setup example 2 (part 1b)....20

Figure 10: Additional setup example 2 (part 2)....21

Figure 11: MC-LAG example topology....24

Figure 12: Network resiliency....25

Figure 13: Association of SAPs/SDPs and endpoints....32

Figure 14: ICB spoke SDPs and their association with the endpoints....36

Figure 15: Additional setup example 1....39

Figure 16: Additional setup example 2....40

Figure 17: Example topology....44

Figure 18: Non-redundant PXC....46

Figure 19: PXC redundant mode with LAG....49

Figure 20: AS mode with redundant FPE....59

Preface

About This Guide

Each Advanced Configuration Guide is organized alphabetically and provides feature and configuration explanations, CLI descriptions, and overall solutions. The Advanced Configuration Guide chapters are written for and based on several Releases, up to 24.7.R2. The Applicability section in each chapter specifies on which release the configuration is based.

The Advanced Configuration Guides supplement the user configuration guides listed in the 7450 ESS, 7750 SR, and 7950 XRS Guide to Documentation.

Audience

This manual is intended for network administrators who are responsible for configuring the routers. It is assumed that the network administrators have a detailed understanding of networking principles and configurations.

Multi-Chassis APS and Pseudowire Redundancy Interworking

This chapter describes multi-chassis APS and pseudowire redundancy interworking.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

This chapter was initially written for SR OS Release 6.0.R2, but the CLI in the current edition is based on SR OS Release 19.10.R2.The configuration in this chapter includes the use of the ATM ports. See the Release Notes for information about support of ATM (and other) MDAs on various platforms as well as MC-APS restrictions.

Overview

MC-APS

MC-APS is an extension to the APS feature to provide not only link redundancy but also node level redundancy. It can protect against nodal failure by configuring the working circuit of an APS group on one node while configuring the protect circuit of the same APS group on a different node.

The two nodes connect to each other with an IP link that is used to establish a signaling path between them. The relevant APS groups in both the working and protection routers must have the same group ID and working circuit, and the protect circuit must have compatible configurations (such as the same speed, framing, and port type). Signaling is provided using the direct connection between the two service routers. A heartbeat protocol can be used to add robustness to the interaction between the two routers.

Signaling functionality includes support for:

- APS group matching between service routers.
- Verification that one side is configured as a working circuit and the other side is configured as the protect circuit. In case of a mismatch, a trap (incompatible-neighbor) is generated.
- Change in working circuit status is sent from the working router to keep the protection router in sync.
- Protection router, based on K1/K2 byte data, member circuit status, and external request, selects the active circuit and informs the working router to activate or de-activate the working circuit.

Pseudowire redundancy

Pseudowire (PW) redundancy provides the ability to protect a pseudowire with a pre-provisioned pseudowire and to switch traffic over to the secondary standby pseudowire in case of a SAP and/or network failure condition. Normally, pseudowires are redundant by the virtue of the SDP redundancy mechanism. For instance, if the SDP is an RSVP LSP and is protected by a secondary standby path and/or by Fast-Reroute paths, the pseudowire is also protected.

However, there are a few applications in which SDP redundancy does not protect the end-to-end pseudowire path when there are two different destination SR-series PE nodes for the same VLL service. The main use case is the provisioning of dual-homing of a CPE or access node to two SR-series PE nodes located in different POPs. The other use case is the provisioning of a pair of active and standby BRAS nodes, or active and standby links to the same BRAS node, to provide service resiliency to broadband service subscribers.

Example topology

The setup in this section contains two access nodes and 4 PE nodes. The access nodes can be any ATM switches that support 1+1 bi-directional APS. Figure 1: MC-APS network topology shows the physical topology of the setup. Figure 3: Access node and network resilience (part 2) shows the use of both MC-APS in the access network and pseudowire redundancy in the core network to provide a resilient end-to-end VLL service.

Figure 1: MC-APS network topology

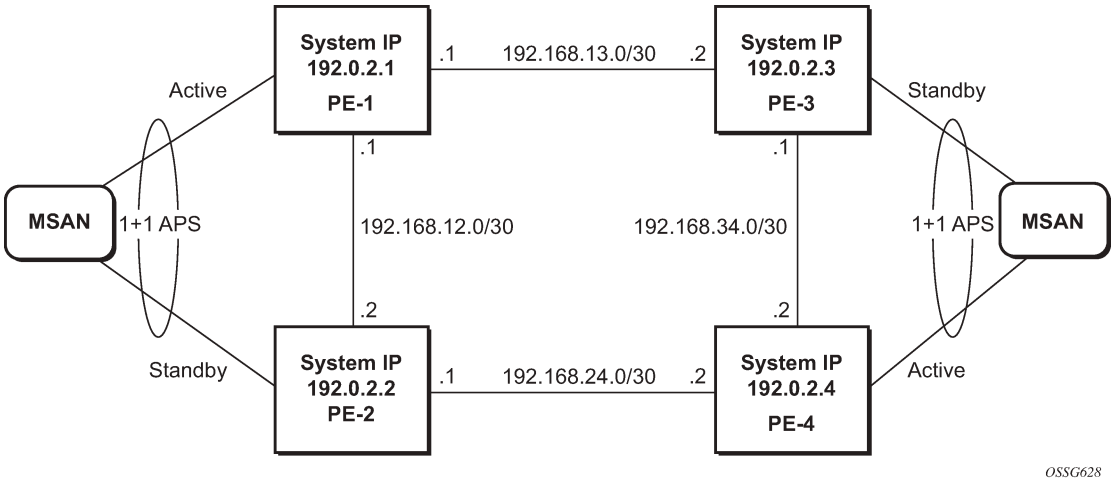


Figure 2: Access node and network resilience (part 1)

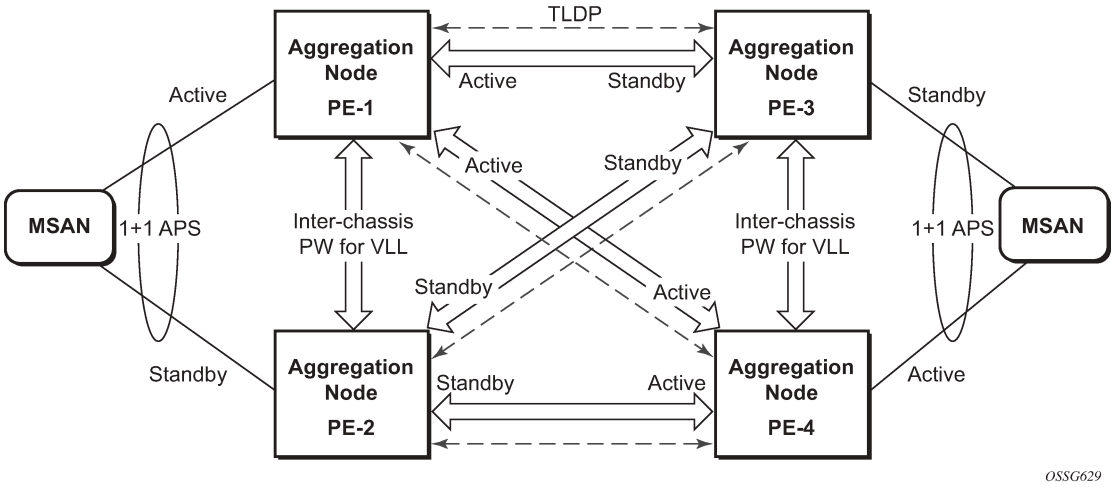
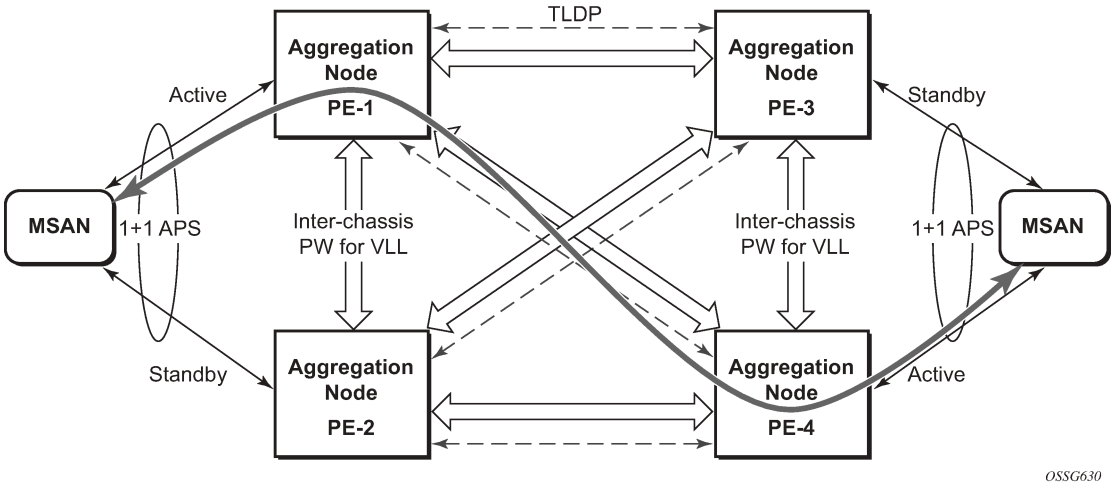


Figure 3: Access node and network resilience (part 2)



Configuration

The following configuration should be completed on the PEs before configuring MC-APS:

- Cards, MDAs and ports
- Router interfaces
- IGP configured and converged
- MPLS
- SDPs configured between all PE routers

For the IGP, OSPF or IS-IS can be used. MPLS or GRE can be used for the transport tunnels. Also, several protocols can be used for signaling MPLS labels. In this example, OSPF and LDP are used. The

following commands can be used to check if OSPF has converged and to make sure the SDPs are up (for example, on PE-1):

```
*A:PE-1# show router route-table
===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                              Type    Proto     Age        Pref
      Next Hop[Interface Name]                                      Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                    Local   Local     01h29m06s  0
       system                                                          0
192.0.2.2/32                                    Remote  OSPF      01h23m18s  10
       192.168.12.2                                                    100
192.0.2.3/32                                    Remote  OSPF      01h17m45s  10
       192.168.13.2                                                    100
192.0.2.4/32                                    Remote  OSPF      01h17m30s  10
       192.168.12.2                                                    200
192.168.12.0/30                                 Local   Local     01h29m06s  0
       int-PE-1-PE-2                                                   0
192.168.13.0/30                                 Local   Local     01h29m06s  0
       int-PE-1-PE-3                                                   0
192.168.24.0/30                                 Remote  OSPF      01h23m18s  10
       192.168.12.2                                                    200
192.168.34.0/30                                 Remote  OSPF      01h17m45s  10
       192.168.13.2                                                    200
-------------------------------------------------------------------------------
No. of Routes: 8
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

```
*A:PE-1# show service sdp
===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
12     0       1492    192.0.2.2        Up   Up          MPLS    L     TLDP
13     0       1492    192.0.2.3        Up   Up          MPLS    L     TLDP
14     0       1492    192.0.2.4        Up   Up          MPLS    L     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 3
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE, F = FPE
===============================================================================
```

Step 1. APS configuration on MSANs

The access nodes can be any ATM switches that support 1+1 bi-directional APS. Here is an example on 7670RSP (Routing Switching Platform).

```
RSP> configure
RSP> port 1-6-1-1
RSP> options protection type 1+1
RSP> options protection switching bidirect
RSP> options protection
                                                                    (Standby)
#                  Type        Status           Name
```

```
1-6-1-1           STM1_IR8    OK
Protection Group Contains:
  Protection Port     : 1-6-1-1     (Standby)
  Working Port        : 1-5-1-1
Protection Type       : 1+1
Switching Type        : Non-Revertive
Switching Mode        : Bi-directional
Wait-To-Restore Timer :  5 minute(s)
```

Step 2. MC-APS configuration on PE-1 and PE-2

Assuming the link between MSAN and PE-1 is working circuit and the link between MSAN and PE-2 is protection circuit.

Configure APS on the PE-1 port. Specify the system IP address of neighbor node (PE-2) and working-circuit.

```
# on PE-1
configure
    port 1/2/1
        sonet-sdh
        exit
        no shutdown
    exit
    port aps-1
        aps
            neighbor 192.0.2.2
            working-circuit 1/2/1
        exit
        sonet-sdh
            path sts3
                encap-type atm
                crc 32
                atm
                exit
                no shutdown
            exit
        exit
        no shutdown
    exit
```

Configure APS on the PE-2 port. Specify the system IP address of neighbor node (PE-1) and protect-circuit instead of working-circuit.

```
# on PE-2
configure
    port 1/2/1
        sonet-sdh
        exit
        no shutdown
    exit
    port aps-1
        aps
            neighbor 192.0.2.1
            protect-circuit 1/2/1
        exit
        sonet-sdh
            path sts3
                encap-type atm
                crc 32
                atm
                exit
```

```
            no shutdown
         exit
      exit
      no shutdown
```

The following parameters can optionally be configured under APS.

- advertise-interval — This command specifies the time interval, in 100s of milliseconds, between 'I am operational' messages sent by both protect and working circuits to their neighbor for multi-chassis APS.
- hold-time — This command specifies how much time can pass, in 100s of milliseconds, without receiving an advertise packet from the neighbor before the multi-chassis signaling link is considered not operational.
- revert-time — This command configures the revert-time timer to determine how long to wait before switching back to the working circuit after that circuit has been restored into service.
- switching-mode — This command configures the switching mode for the APS port which can be bi-directional or uni-directional.

Step 3. Verify the APS status on PE-1.

```
*A:PE-1# show port aps-1

===============================================================================
SONET/SDH Interface
===============================================================================
Description        : APS Group
Interface          : aps-1                 Speed               : oc3
Admin Status       : up                    Oper Status         : up
Physical Link      : Yes                   Loopback Mode       : none
Single Fiber Mode  : No
Clock Source       : loop                  Framing             : sonet
Last State Change  : 01/10/2020 09:38:21   Port IfIndex        : 1358987264
Configured Address : 04:0f:ff:00:02:49
Hardware Address   : 04:0f:ff:00:02:49
Last Cleared Time  : N/A
J0 String          : 0x01                  Section Trace Mode  : byte
Rx S1 Byte         : 0x00 (stu)            Rx K1/K2 Byte       : 0x00/0x00
Tx S1 Byte         : 0x0f (dus)            Tx DUS/DNU          : Disabled
Rx J0 String (Hex) : 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
Cfg Alarm          : loc lrdi lb2er-sf slof slos
Alarm Status       :
BER SD Threshold   : 6                     BER SF Threshold    : 3
Hold time up       : 500 milliseconds      Reset On Path Down  : Disabled
Hold time down     : 200 milliseconds

===============================================================================
Port Statistics
===============================================================================
                                                    Input                Output
-------------------------------------------------------------------------------
Packets                                                0                     0
Discards                                               0                     0
Unknown Proto Discards                                 0
===============================================================================
```

Step 4. Verify the MC-APS status and parameters on PE-1 and PE-2

Detailed parameters of the APS configuration on PE-1 can be verified, as follows. The admin/oper status of APS group 1 shows up/up. K1/K2 byte shows N/A as APS 1+1 exchanges that information through the protection circuit. The admin/oper status of the working circuit (the link between MSAN and PE-1) is up/up.

```
*A:PE-1# show aps detail

===============================================================================
APS Group: aps-1
===============================================================================
Description        : APS Group
Group Id           : 1                     Active Circuit       : 1/2/1
Admin Status       : Up                    Oper Status          : Up
Working Circuit    : 1/2/1                 Protection Circuit   : N/A
Switching-mode     : Bi-directional        Switching-arch       : 1+1(sig-only)
Annex B            : No
Revertive-mode     : Non-revertive         Revert-time (min)    :
Rx K1/K2 byte      : N/A
Tx K1/K2 byte      : N/A
Current APS Status : OK
Multi-Chassis APS  : Yes
Neighbor           : 192.0.2.2
Control link state : Up
Advertise Interval : 1000 msec             Hold Time            : 3000 msec
Mode mismatch Cnt  : 0                     Channel mismatch Cnt : 0
PSB failure Cnt    : 0                     FEPL failure Cnt     : 0
-------------------------------------------------------------------------------
 APS Working Circuit - 1/2/1
-------------------------------------------------------------------------------
Admin Status       : Up                    Oper Status          : Up
Current APS Status : OK                    No. of Switchovers   : 0
Last Switchover    : None                  Switchover seconds   : 0
Signal Degrade Cnt : 0                     Signal Failure Cnt   : 0
Last Switch Cmd    : N/A                   Last Exercise Result : N/A
Tx L-AIS           : None
===============================================================================
```

Detailed parameters of the APS configuration on PE-2 can be verified, as follows. The admin/oper status of APS group 1 shows up/up. Both Rx and Tx of the K1/K2 byte are in the status of 0x00/0x05 (No-Req on Protect) as there is no failure or force-switchover request. The admin/oper status of the protection circuit (the link between MSAN and PE-2) is up/up.

```
*A:PE-2# show aps detail

===============================================================================
APS Group: aps-1
===============================================================================
Description        : APS Group
Group Id           : 1                     Active Circuit       : N/A
Admin Status       : Up                    Oper Status          : Up
Working Circuit    : N/A                   Protection Circuit   : 1/2/1
Switching-mode     : Bi-directional        Switching-arch       : 1+1(sig-only)
Annex B            : No
Revertive-mode     : Non-revertive         Revert-time (min)    :
Rx K1/K2 byte      : 0x00/0x05 (No-Req on Protect)
Tx K1/K2 byte      : 0x00/0x05 (No-Req on Protect)
Current APS Status : OK
Multi-Chassis APS  : Yes
Neighbor           : 192.0.2.1
Control link state : Up
Advertise Interval : 1000 msec             Hold Time            : 3000 msec
Mode mismatch Cnt  : 0                     Channel mismatch Cnt : 0
PSB failure Cnt    : 0                     FEPL failure Cnt     : 1
```

```
-------------------------------------------------------------------------------
 APS Working Circuit - Neighbor
-------------------------------------------------------------------------------
Admin Status       : N/A                  Oper Status          : N/A
Current APS Status : OK                   No. of Switchovers   : 0
Last Switchover    : None                 Switchover seconds   : 0
Signal Degrade Cnt : 0                    Signal Failure Cnt   : 1
Last Switch Cmd    : No Cmd               Last Exercise Result : Unknown
Tx L-AIS           : None
-------------------------------------------------------------------------------
 APS Protection Circuit - 1/2/1
-------------------------------------------------------------------------------
Admin Status       : Up                   Oper Status          : Up
Current APS Status : OK                   No. of Switchovers   : 0
Last Switchover    : None                 Switchover seconds   : 0
Signal Degrade Cnt : 0                    Signal Failure Cnt   : 0
Last Switch Cmd    : No Cmd               Last Exercise Result : Unknown
Tx L-AIS           : None
===============================================================================
```

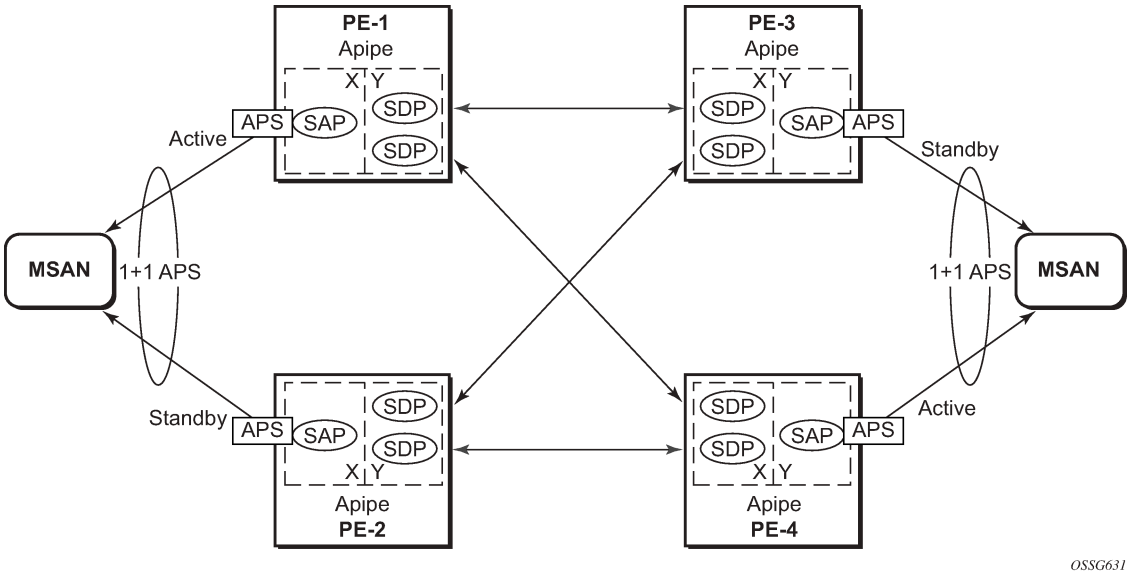
Step 5. MC-APS configuration on PE-3 and PE-4

The MC-APS configuration on PE-3 and PE-4 is similar to the configuration on PE-1 and PE-2. Configure the working circuit on PE-4 and the protection circuit on PE-3.

Step 6. Pseudowire configuration

Configure an Apipe service on every PE and create endpoints X and Y. Associate the SAPs and spoke SDPs with the endpoints, as shown in Figure 4: Association of SAPs/SDPs and endpoints.

Figure 4: Association of SAPs/SDPs and endpoints




```
# on PE-1
configure
    service
        apipe 1 name name "Apipe 1" customer 1 create
            endpoint "X" create
            exit
            endpoint "Y" create
```

```
            exit
            service-mtu 1400
            sap aps-1:0/32 endpoint "X" create
            exit
            spoke-sdp 13:1 endpoint "Y" create
            exit
            spoke-sdp 14:1 endpoint "Y" create
            exit
            no shutdown
        exit
```

Syntax aps-1:0/32 specifies the APS group and VPI/VCI of the ATM circuit using the *aps-id:vpi/vci* format.

Likewise, an Apipe service, with endpoints, SAPs and spoke SDPs must be configured on the other PE routers.

Step 7. Pseudowire verification

The Apipe service is up in PE-1 (MC-APS working circuit), as follows:

```
*A:PE-1# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Apipe     Up   Up   1          1
---snip---
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

The Apipe service is down in PE-2 (MC-APS protect circuit), as follows:

```
*A:PE-2# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Apipe     Up   Down 1          1
---snip---
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

The Apipe service is down in PE-3 (MC-APS protect circuit), as follows:

```
*A:PE-3# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Apipe     Up   Down 1          1
---snip---
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-3#
```

The Apipe service is up in PE-4 (MC-APS working circuit), as follows:

```
*A:PE-4# show service service-using
===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Apipe     Up   Up   1          1
---snip---
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

Note: After configuring ICB spoke-SDPs, the Apipe will be up on all PEs.

Step 8. Verify SDP status

The status of SDP 23:1 on PE-2 can be verified as follows.

Peer Pw Bits shows the status of the pseudowire on the peer node. In this example, both the local node (PE-2) as the remote node (PE-3) are sending the *lacIngressFault*, *lacEgressFault* and *pwFwdingStandby* flags. This is because the Apipe service on these nodes is down because the MC-APS is in protection status.

```
*A:PE-2# show service id 1 sdp 23:1 detail

===============================================================================
Service Destination Point (Sdp Id : 23:1) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 23:1  -(192.0.2.3)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id             : 23:1                    Type              : Spoke
Spoke Descr     : (Not Specified)
Split Horiz Grp    : (Not Specified)
VC Type            : AAL5SDU                 VC Tag            : 0
Admin Path MTU     : 0                       Oper Path MTU     : 1492
Delivery           : MPLS
Far End            : 192.0.2.3               Tunnel Far End    :
Oper Tunnel Far End: 192.0.2.3
LSP Types          : LDP

Admin State        : Up                      Oper State        : Up
MinReqd SdpOperMTU : 1404
Acct. Pol          : None                    Collect Stats     : Disabled
Ingress Label      : 524283                  Egress Label      : 524282
Ingr Mac Fltr-Id   : n/a                     Egr Mac Fltr-Id   : n/a
Ingr IP Fltr-Id    : n/a                     Egr IP Fltr-Id    : n/a
Admin ControlWord  : Preferred               Oper ControlWord  : True
Admin BW(Kbps)     : 0                       Oper BW(Kbps)     : 0
BFD Template       : None
BFD-Enabled        : no                      BFD-Encap         : ipv4
Last Status Change : 01/10/2020 09:46:58     Signaling         : TLDP
Last Mgmt Change   : 01/10/2020 09:46:52
Endpoint           : Y                       Precedence        : 4
PW Status Sig      : Enabled
Class Fwding State : Down
Flags              : None
```

```
Local Pw Bits      : lacIngressFault lacEgressFault pwFwdingStandby
Peer Pw Bits       : lacIngressFault lacEgressFault pwFwdingStandby
Peer Fault Ip      : None
Peer Vccv CV Bits  : lspPing bfdFaultDet
Peer Vccv CC Bits  : pwe3ControlWord mplsRouterAlertLabel

---snip---

===============================================================================
```

In case of failure, the access link can be protected by MC-APS. An MPLS network failure can be protected by pseudowire redundancy. Node failure can be protected by the combination of MC-APS and pseudowire redundancy.

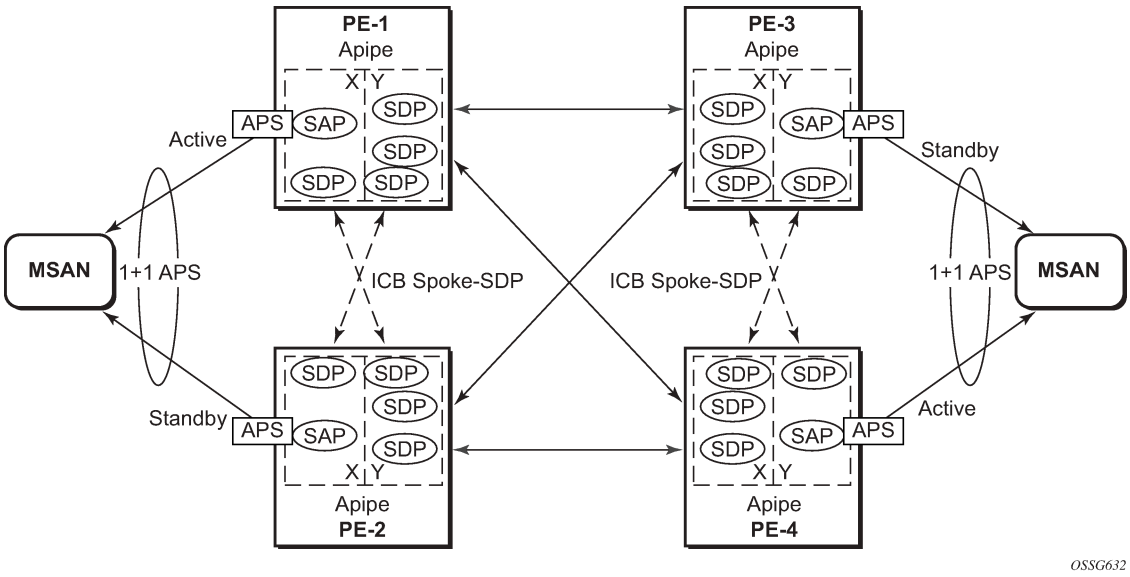
Step 9. Inter-Chassis Backup (ICB) pseudowire configuration.

Configuring Inter-Chassis Backup (ICB) is optional. It can reduce traffic impact by forwarding traffic on ICB spoke SDPs during MC-APS switchover. The ICB spoke SDP cannot be added to the endpoint if the SAP is not part of an MC-APS (or MC-LAG) instance. Conversely, a SAP which is not part of a MC-APS (or MC-LAG) instance cannot be added to an endpoint which already has an ICB spoke SDP. Forwarding between ICBs is blocked on the same node. The user has to explicitly indicate the spoke SDP is actually an ICB at creation time. Figure 5 shows some setup examples where ICBs are required.

After configuring ICB spoke SDPs, the Apipe will be in admin/oper up/up status on all PE routers.

ICB SDPs are configured and associated to endpoints, as shown in Figure 5: ICB spoke SDPs and association with the endpoints.

Figure 5: ICB spoke SDPs and association with the endpoints



Two ICB spoke SDPs must be configured in the Apipe service on each PE router, one in each endpoint. The same SDP IDs can be used for the ICBs since the far-end will be the same. However, the VC-id must be different. The ICB spoke SDPs must cross, meaning one end should be associated with endpoint X and the other end (on the other PE) should be associated with endpoint Y.

An ICB is always the last forwarding resort. Only one spoke SDP will be forwarding. If there is an ICB and an MC-APS SAP in an endpoint, the ICB will only forward if the SAP goes down. If an ICB resides in an endpoint together with other spoke SDPs the ICB will only forward if there is no other active spoke SDP.

The following shows the additional configuration for ICB on each PE:

```
# on PE-1
configure
    service
        apipe 1
            spoke-sdp 12:1 endpoint "X" icb create
            exit
            spoke-sdp 12:2 endpoint "Y" icb create
            exit
        exit
    exit
exit
```

```
# on PE-2
configure
    service
        apipe 1
            spoke-sdp 21:1 endpoint "Y" icb create
            exit
            spoke-sdp 21:2 endpoint "X" icb create
            exit
        exit
    exit
exit
```

```
# on PE-3
configure
    service
        apipe 1
            spoke-sdp 34:1 endpoint "X" icb create
            exit
            spoke-sdp 34:2 endpoint "Y" icb create
            exit
        exit
    exit
exit
```

```
# on PE-4
configure
    service
        apipe 1
            spoke-sdp 43:1 endpoint "Y" icb create
            exit
            spoke-sdp 43:2 endpoint "X" icb create
            exit
        exit
    exit
exit
```

Step 10. Verification of active objects for each endpoint

The following command shows which objects are configured for each endpoint and which is the active object at this moment:

```
*A:PE-1# show service id 1 endpoint

===============================================================================
Service 1 endpoints
===============================================================================
```

```
Endpoint name                : X
Description                  : (Not Specified)
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Tx Active                    : aps-1:0/32
Tx Active Up Time            : 0d 00:38:31
Revert Time Count Down       : never
Tx Active Change Count       : 1
Last Tx Active Change        : 01/10/2020 09:46:45
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
SAP      : aps-1:0/32                           Oper Status: Up
Spoke-sdp: 12:1 Prec:4 (icb)                    Oper Status: Up
===============================================================================
Endpoint name                : Y
Description                  : (Not Specified)
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Tx Active (SDP)              : 14:1
Tx Active Up Time            : 0d 00:35:40
Revert Time Count Down       : never
Tx Active Change Count       : 2
Last Tx Active Change        : 01/10/2020 10:03:31
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 12:2 Prec:4 (icb)                    Oper Status: Up
Spoke-sdp: 13:1 Prec:4                          Oper Status: Up
Spoke-sdp: 14:1 Prec:4                          Oper Status: Up
===============================================================================
===============================================================================
```

On PE-1, both the SAP and the spoke SDP 14:1 are active. The other objects do not forward traffic.

Step 11. Other types of setups

The following figures show other setups that combine MC-APS and pseudowire redundancy.

Figure 6: Additional setup example 1 (part 1)

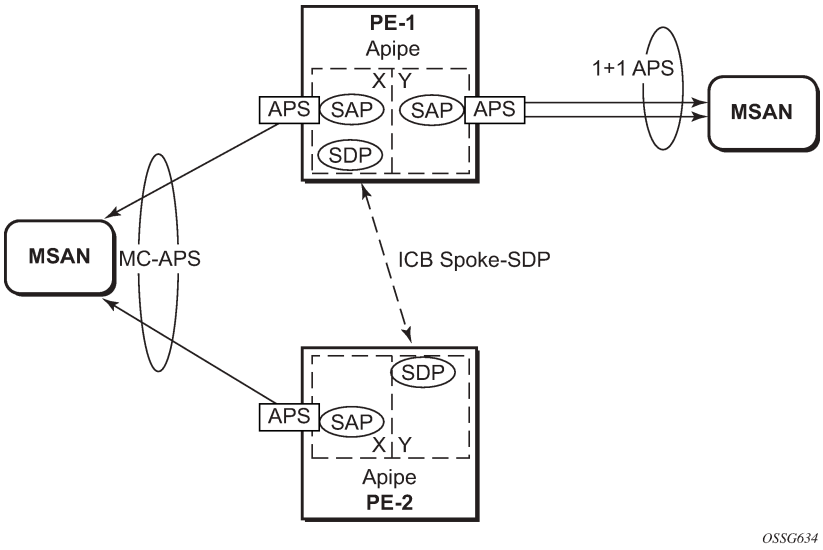


Figure 7: Additional setup example 1 (part 2)

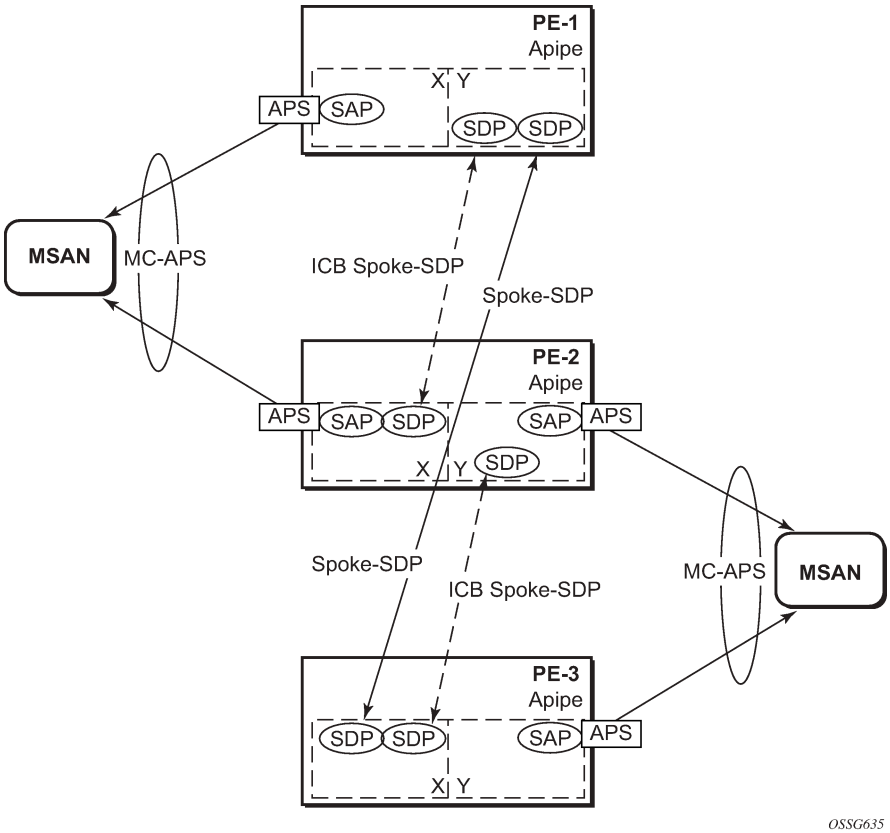


Figure 8: Additional setup example 2 (part 1a)

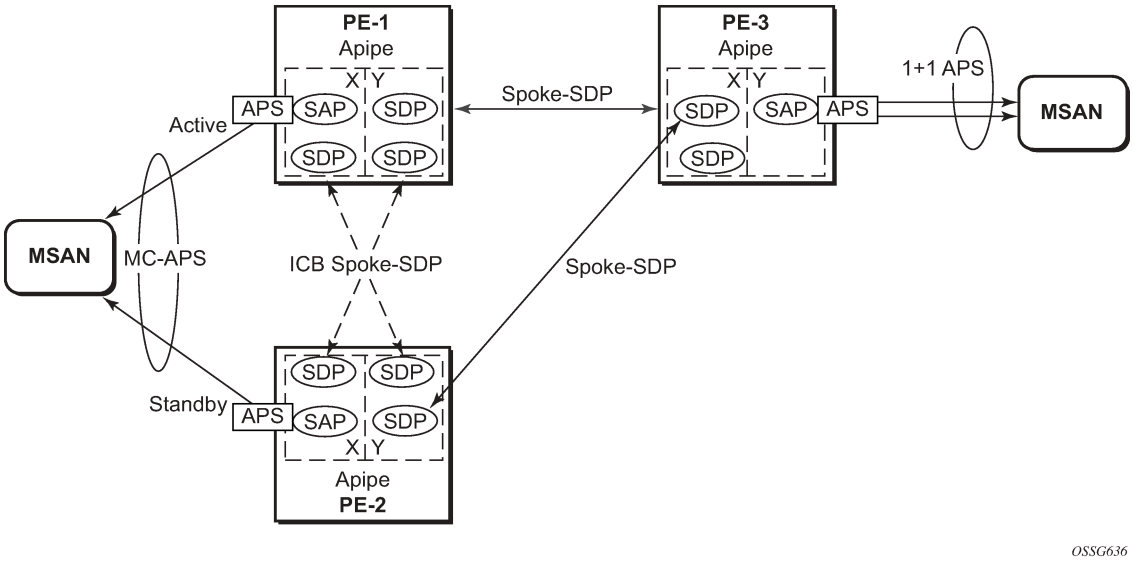


Figure 9: Additional setup example 2 (part 1b)

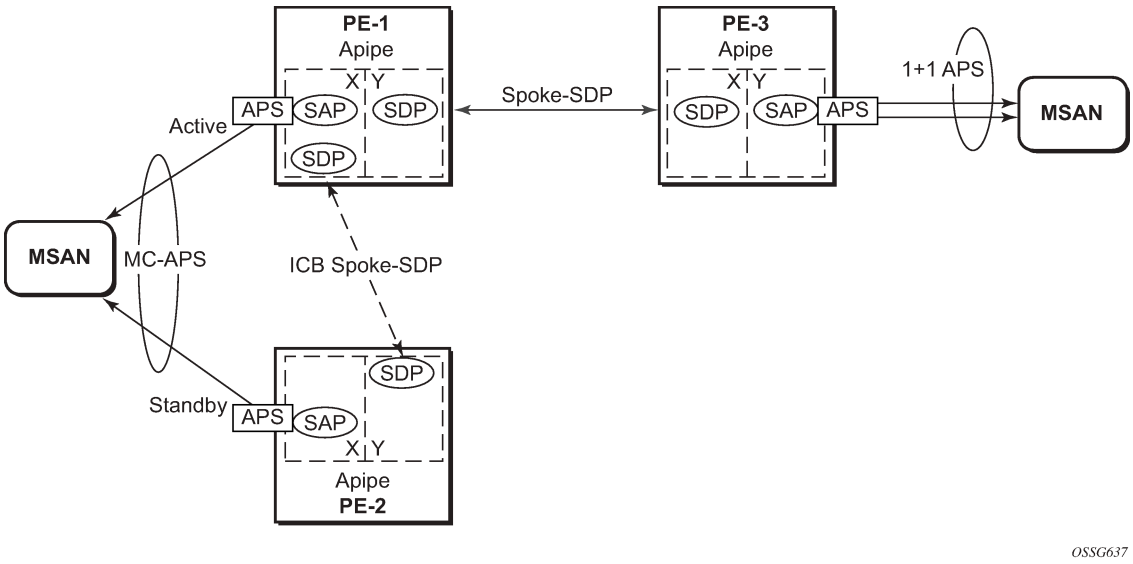
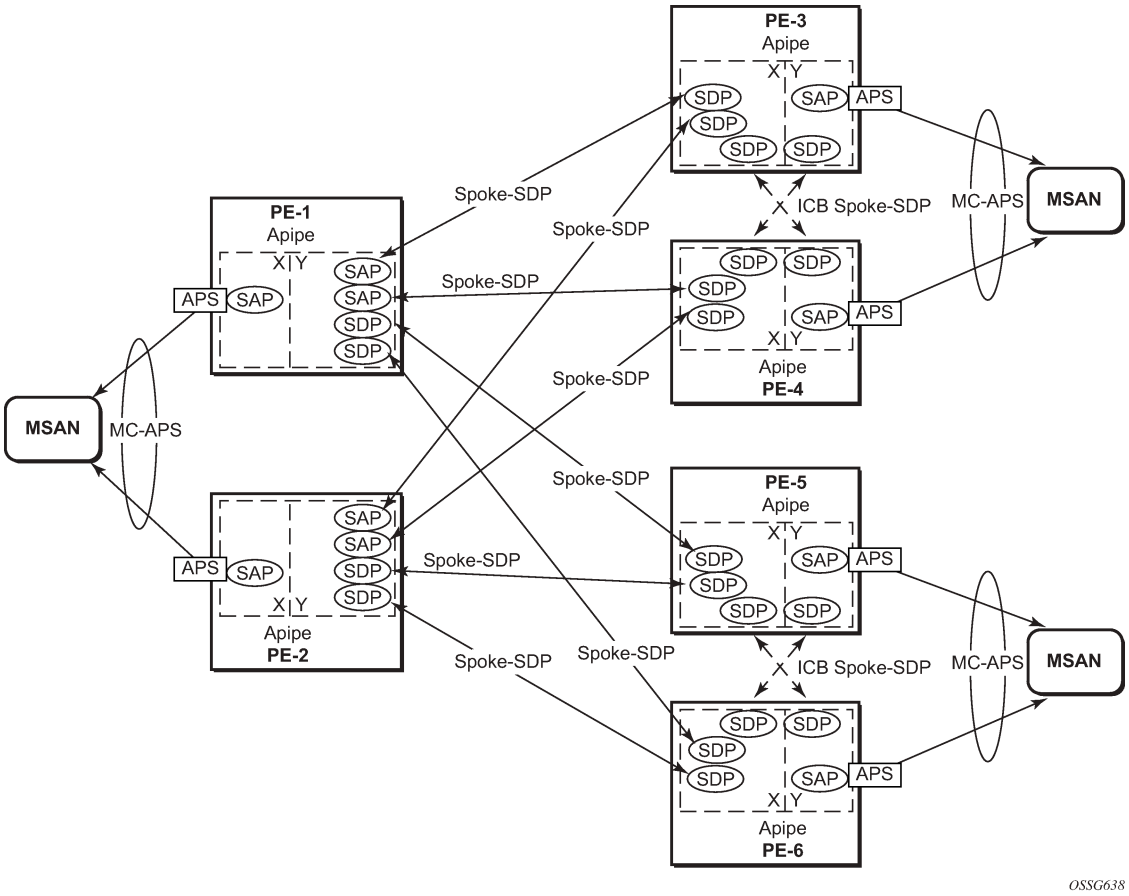


Figure 10: Additional setup example 2 (part 2)



Forced switchover

MC-APS convergence can be forced with the **tools perform aps** command:

```
*A:PE-1# tools perform aps force
  - force <aps-id> {protect|working} [number <number>]

 <aps-id>             : aps-<group-id>
                        aps            - keyword
                        group-id       - [1..128]
 <protect|working>    : keyword
 <number>             : [1-2]
```

After the forced switchover, it is important to clear the forced switchover:

```
*A:PE-1# tools perform aps clear
  - clear <aps-id> {protect|working} [number <number>]

 <aps-id>             : aps-<group-id>
                        aps            - keyword
                        group-id       - [1..128]
 <protect|working>    : protect|working
```

```
<number>            : [1-2]
```

Conclusion

In addition to Multi-Chassis LAG, Multi-Chassis APS provides a solution for both network redundancy and access node redundancy. It supports ATM VLL and Ethernet VLL with ATM SAP. Access links and PE nodes are protected by APS and the MPLS network is protected by pseudowire redundancy/FRR. With this feature, Nokia can provide resilient end-to-end solutions.

Multi-Chassis LAG and Pseudowire Redundancy Interworking

This chapter provides information about Multi-Chassis Link Aggregation (MC-LAG) and pseudowire redundancy interworking.

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

MC-LAG is supported only on Ethernet MDAs, and this only for access ports, because the LAG group must be in access mode.

This chapter was initially written for SR OS Release 7.0.R5. However, the CLI in the current edition is based on SR OS Release 19.10.R2.

Overview

MC-LAG

MC-LAG is an extension to the LAG feature to provide not only link redundancy but also node-level redundancy. This feature provides a Nokia added value solution which is not defined in any IEEE standard.

A proprietary messaging system between redundant-pair nodes supports coordinating the LAG switchover.

Multi-chassis LAG supports LAG switchover coordination: one node connected to two redundant-pair peer nodes with the LAG. During the LACP negotiation, the redundant-pair peer nodes act like a single node using active/stand-by signaling to ensure that only links of one peer node are used at a time.

Pseudowire redundancy

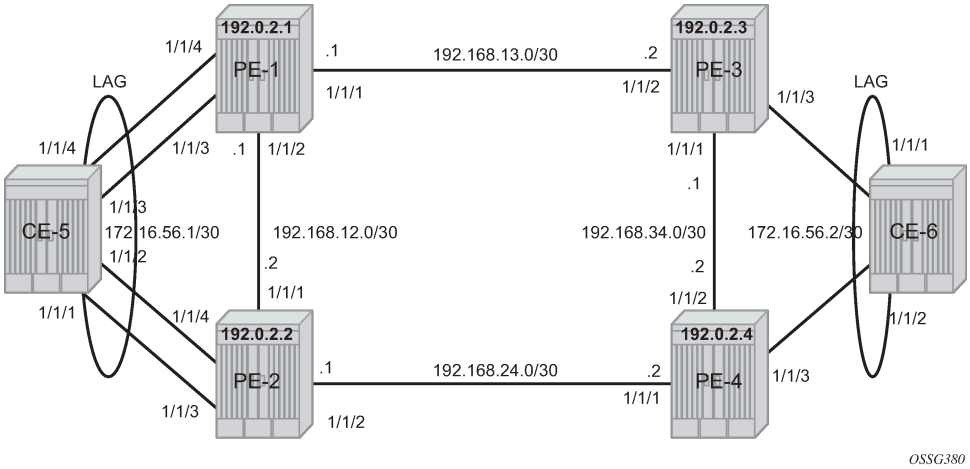
Pseudowire (PW) redundancy provides the ability to protect a pseudowire with a secondary pre-provisioned pseudowire and to switch traffic over to the secondary standby pseudowire in case of a SAP and/or network failure condition. Normally, pseudowires are redundant by the virtue of the SDP redundancy mechanism. For instance, if the SDP relies on an RSVP LSP that is protected by a secondary standby path and/or by Fast-Reroute paths, the pseudowire is also protected.

However, there are a few applications in which SDP redundancy does not protect the end-to-end pseudowire path, for example when there are two different destination SR-series PE nodes for the same VLL service.

The main use case for PW redundancy is a scenario where dual homed CPEs or access nodes connected to two SR-series PE nodes are located in different POPs. The other use case is the scenario where service resiliency for broadband service subscribers is required, for example when a pair of active and standby BRAS nodes are provisioned, or where active and standby links to the same BRAS node are provisioned.

Example topology

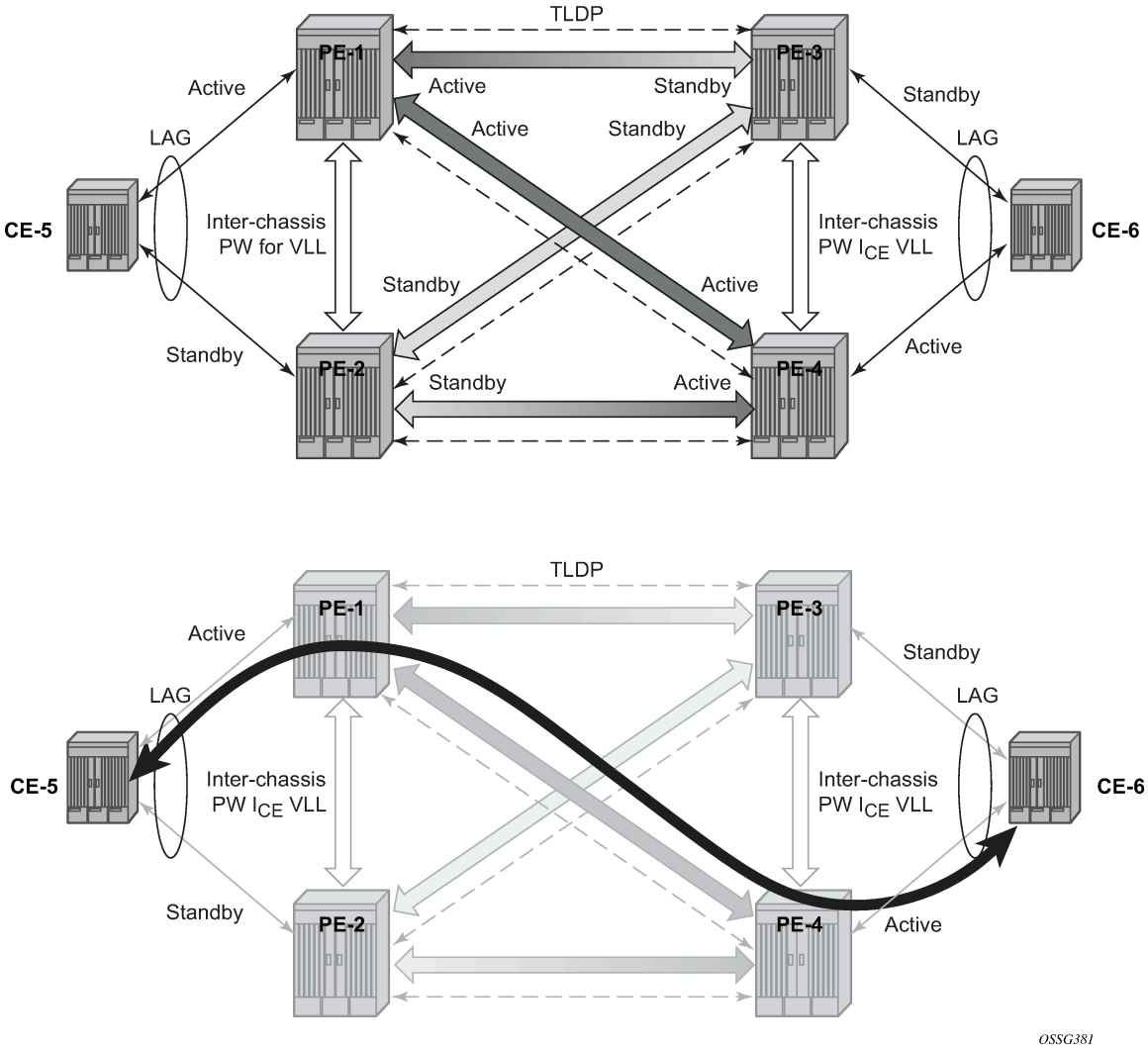
Figure 11: MC-LAG example topology



This section describes a setup which contains two CEs and four PEs. The CEs can be any routing/switching device that support the OUT_OF_SYNC signaling as described in IEEE Standard 802.3-2005 section 3 section 43.6.1. Figure 11: MC-LAG example topology shows the physical topology of the setup.

Figure 12: Network resiliency shows the use of both MC-LAG in the access network and pseudowire redundancy in the core network to provide a resilient end-to-end VLL service between CE-5 and CE-6.

Figure 12: Network resiliency



When an SDP is in standby, it sends the pseudowire status bit pwFwdingStandby to its peer.

Configuration

It is assumed that the following base configuration has been implemented on the PEs:

- Cards, MDAs, and ports
- Router interfaces
- IGP configured and converged
- MPLS
- SDPs configured between all PE routers

Either OSPF or IS-IS can be used as the IGP. Also, several protocols can be used for signaling the transport MPLS labels. Alternatively, GRE can be used for the transport tunnels. Likewise, several protocols can be used for signaling the SDPs. In this example, OSPF and LDP are used.

The following command is used to check if OSPF has converged (for example, on PE-1):

```
*A:PE-1# show router route-table

===============================================================================
Route Table (Router: Base)
===============================================================================
Dest Prefix[Flags]                            Type    Proto     Age        Pref
      Next Hop[Interface Name]                                    Metric
-------------------------------------------------------------------------------
192.0.2.1/32                                  Local   Local     00h16m53s  0
       system                                                         0
192.0.2.2/32                                  Remote  OSPF      00h14m26s  10
       192.168.12.2                                                   1000
192.0.2.3/32                                  Remote  OSPF      00h14m21s  10
       192.168.13.2                                                   1000
192.0.2.4/32                                  Remote  OSPF      00h14m21s  10
       192.168.12.2                                                   2000
192.168.12.0/30                               Local   Local     00h16m53s  0
       int-PE-1-PE-2                                                  0
192.168.13.0/30                               Local   Local     00h16m53s  0
       int-PE-1-PE-3                                                  0
192.168.24.0/30                               Remote  OSPF      00h14m26s  10
       192.168.12.2                                                   2000
192.168.34.0/30                               Remote  OSPF      00h14m21s  10
       192.168.13.2                                                   2000
-------------------------------------------------------------------------------
No. of Routes: 8
Flags: n = Number of times nexthop is repeated
       B = BGP backup route available
       L = LFA nexthop available
       S = Sticky ECMP requested
===============================================================================
```

The following command shows that the SDPs are up:

```
*A:PE-1# show service sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
12     0       1492    192.0.2.2        Up   Up          MPLS    L     TLDP
13     0       1492    192.0.2.3        Up   Up          MPLS    L     TLDP
14     0       1492    192.0.2.4        Up   Up          MPLS    L     TLDP
-------------------------------------------------------------------------------
Number of SDPs : 3
-------------------------------------------------------------------------------
Legend: R = RSVP, L = LDP, B = BGP, M = MPLS-TP, n/a = Not Applicable
        I = SR-ISIS, O = SR-OSPF, T = SR-TE, F = FPE
===============================================================================
```

MC-LAG for Epipe services

Step 1 - MC-LAG configuration on CEs.

The LAG configuration on the CEs is only included for completeness; any CE device could be used.

Auto-negotiation must be switched off or set to limited on all ports that will be included into the LAG in order to guarantee a specific port speed.

Note:
Disabling autonegotiation on Gigabit ports is not allowed because the IEEE 802.3 specification for Gigabit Ethernet requires autonegotiation to be enabled for far end fault detection.

Configure LACP on the LAG with at least one side of the LAG in **active** mode.

```
*A:CE-5# configure port 1/1/[1..4] ethernet autonegotiate limited
*A:CE-5# configure port 1/1/[1..4] no shutdown

# on CE-5:
configure
    lag 1
        port 1/1/1
        port 1/1/2
        port 1/1/3
        port 1/1/4
        lacp active administrative-key 32768
        no shutdown
    exit
exit
```

Step 2 - LAG configuration on PEs.

The PE ports connected to the CEs must be configured as access ports because they will be used in the redundant pseudowire service. The LAG must also be configured in access mode.

The LAG encapsulation type (null | dot1q | qinq) must match the port encapsulation type of the LAG members.

Auto-negotiation must be switched off or configured to limited.

Configure LACP on the LAG. At least 1 side of the LAG (PE or CE) must be configured in **active** mode.

```
# on PE-1:
configure
    port 1/1/3
        ethernet
            mode access
            autonegotiate limited
        exit
        no shutdown
    exit
    port 1/1/4
        ethernet
            mode access
            autonegotiate limited
        exit
        no shutdown
    exit
    lag 1
        mode access
        port 1/1/3
        port 1/1/4
        lacp active administrative-key 32768
        no shutdown
    exit
exit
```

Step 3 - MC-LAG configuration on PE-1 and PE-2

The redundant PEs must act as one virtual node toward the CE. They have to be able to communicate the same LACP parameters to the CE side.

The following parameters uniquely identify a LAG instance:

- lacp-key
- system-id
- system-priority

These three parameters must be configured with the same value on both redundant PEs.

Multi-chassis redundancy requires a peering session (which operates by an IP connection using UDP destination port 1025) that is configured toward the redundant PE system address to which MC-LAG redundancy is enabled, as follows. The peering session can be configured with MD5 authentication.

```
# on PE-1:
configure
    redundancy
        multi-chassis
            peer 192.0.2.2 create
                authentication-key "GBeGhtqKlY06VZKF2QK+xAzSig==" hash2
                mc-lag
                    lag 1 lacp-key 1 system-id 00:00:00:00:00:01 system-priority 100
                    no shutdown
                exit
                no shutdown
            exit
```

```
# on PE-2:
configure
    redundancy
        multi-chassis
            peer 192.0.2.1 create
                authentication-key "GBeGhtqKlY06VZKF2QK+xF3iEg==" hash2
                mc-lag
                    lag 1 lacp-key 1 system-id 00:00:00:00:00:01 system-priority 100
                    no shutdown
                exit
                no shutdown
            exit
```

Step 4 - MC-LAG verification.

Verify MC peers showing that the authentication and admin state are enabled.

```
*A:PE-1# show redundancy multi-chassis sync

===============================================================================
Multi-chassis Peer Table
===============================================================================
Peer
-------------------------------------------------------------------------------
Peer IP Address         : 192.0.2.2
Description             : (Not Specified)
Authentication          : Enabled
Source IP Address       : 192.0.2.1
Admin State             : Enabled
Warm standby            : No
Remote warm standby     : No
```

```
-------------------------------------------------------------------------------
Sync: Not-configured
-------------------------------------------------------------------------------
===============================================================================
===============================================================================
```

Step 5 - Verify MC-LAG peer status and LAG parameters.

```
*A:PE-1# show redundancy multi-chassis mc-lag peer 192.0.2.2

===============================================================================
Multi-Chassis MC-Lag Peer 192.0.2.2
===============================================================================
Last State chg  : 01/07/2020 12:33:32
Admin State     : Up                  Oper State           : Up
KeepAlive       : 10 deci-seconds     Hold On Ngbr Failure : 3
-------------------------------------------------------------------------------
Lag Id Lacp   Remote Source Oper   System Id          Sys   Last State Changed
       Key    Lag Id MacLSB MacLSB                    Prio
-------------------------------------------------------------------------------
1      1      1      Def    n/a    00:00:00:00:00:01  100   01/07/2020 12:33:33
-------------------------------------------------------------------------------
Number of LAGs : 1
===============================================================================
```

There is a fixed keepalive timer of 1 second. The **hold-on-neighbor-failure multiplier** command indicates the interval that the standby node will wait for packets from the active node before assuming a redundant-neighbor failure. The **hold-on-neighbor-failure** <*multiplier*> command is configurable in the **config>redundancy>multi-chassis>peer>mc-lag** context. The standby node will also assume a redundant-neighbor failure when there is no route available to the redundant-neighbor.

```
# on PE-1:
configure
    redundancy
        multi-chassis
            peer 192.0.2.2
                mc-lag
                    hold-on-neighbor-failure 10
                exit
            exit
        exit
    exit
```

In this example, the *lag-id* is 1 on both redundant PEs. This is not mandatory. If the *lag-id* on PE-2 is, for example 2, the following should be configured on PE-1:

```
# on PE-1:
configure
    redundancy
        multi-chassis
            peer 192.0.2.2
                mc-lag
                    lag 1 lacp-key 1 system-id 00:00:00:00:00:01 remote-lag 2
                exit
            exit
        exit
    exit
```

Step 6 - Verify MC-LAG

```
*A:PE-1# show lag 1
===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
-------------------------------------------------------------------------------
1              up      up      No       0         2        active
===============================================================================
```

```
*A:PE-2# show lag 1
===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
-------------------------------------------------------------------------------
1              up      down    No       0         0        standby
===============================================================================
```

In this case, the LAG on PE-1 is active (operationally up) whereas the LAG on PE-2 is standby (operationally down).

The default selection criteria is highest number of links and priority. In this example, the number of links and the priority of the links is the same on both redundant PEs. Whichever PE's LAG gets the operational status *up* first, will be the active LAG.

LAG ports of one PE could be preferred over the other PE by configuring port priority. For example, the following command lowers the priority of the LAG ports on PE-1, thus giving this LAG higher preference. The default priority is 32768, but it is modified to a value of 10, as follows:

```
# on PE-1:
configure
    lag 1
        port 1/1/3 priority 10
        port 1/1/4 priority 10
```

The selection criteria can be configured as highest-count, highest-weight or best-port (the default is highest count).

```
*A:PE-1# configure lag 1 selection-criteria
  - selection-criteria [best-port|highest-count|highest-weight] [slave-to-partner]
 [subgroup-hold-time <hold-time>]
  - no selection-criteria

 <best-port|highest*> : keywords
 <slave-to-partner>   : keyword
 <hold-time>          : [0..2000] tenths of a second | infinite
```

If highest-weight is configured, the sum of the weights of the LAG members is considered. The weight of an individual LAG member is calculated as priority 65535 (the default is 32768).

Step 7 - Verify detailed MC-LAG status on PE-1

```
*A:PE-1# show lag 1 detail

===============================================================================
LAG Details
===============================================================================
```

```
Description        : N/A
-------------------------------------------------------------------------------
Details
-------------------------------------------------------------------------------
Lag-id              : 1                     Mode                : access
Adm                 : up                    Opr                 : up
Thres. Last Cleared : 01/07/2020 12:42:16   Thres. Exceeded Cnt  : 0
Dynamic Cost        : false                 Encap Type          : null
Configured Address  : 04:0f:ff:00:01:41     Lag-IfIndex         : 1342177281
Hardware Address    : 04:0f:ff:00:01:41     Adapt Qos (access)  : distribute
Hold-time Down      : 0.0 sec               Port Type           : standard
Per-Link-Hash       : disabled
Include-Egr-Hash-Cfg: disabled              Forced              : -
Per FP Ing Queuing  : disabled              Per FP Egr Queuing  : disabled
Per FP SAP Instance : disabled
Access Bandwidth    : N/A                   Access Booking Factor: 100
Access Available BW : 0
Access Booked BW    : 0
LACP                : enabled               Mode                : active
LACP Transmit Intvl : fast                  LACP xmit stdby     : enabled
Selection Criteria  : highest-count         Slave-to-partner    : disabled
MUX control         : coupled
Subgrp hold time    : 0.0 sec               Remaining time      : 0.0 sec
Subgrp selected     : 1                     Subgrp candidate    : -
Subgrp count        : 1
System Id           : 04:0f:ff:00:00:00     System Priority     : 32768
Admin Key           : 32768                 Oper Key            : 1
Prtr System Id      : 04:1f:ff:00:00:00     Prtr System Priority : 32768
Prtr Oper Key       : 32768
Standby Signaling   : lacp
Port hashing        : port-speed            Port weight speed   : 0 gbps
Ports Up            : 2
Weights Up          : 2                     Hash-Weights Up     : 2
Monitor oper group  : N/A

MC Peer Address     : 192.0.2.2             MC Peer Lag-id      : 1
MC System Id        : 00:00:00:00:00:01     MC System Priority  : 100
MC Admin Key        : 1                     MC Active/Standby   : active
MC Lacp ID in use   : true                  MC extended timeout : false
MC Selection Logic  : local master decided
MC Config Mismatch  : no mismatch

-------------------------------------------------------------------------------
Port-id        Adm     Act/Stdby Opr     Primary   Sub-group     Forced  Prio
-------------------------------------------------------------------------------
1/1/3          up      active    up      yes       1             -       10
1/1/4          up      active    up                1             -       10

-------------------------------------------------------------------------------
Port-id        Role      Exp   Def   Dist  Col   Syn   Aggr  Timeout  Activity
-------------------------------------------------------------------------------
1/1/3          actor     No    No    Yes   Yes   Yes   Yes   Yes      Yes
1/1/3          partner   No    No    Yes   Yes   Yes   Yes   Yes      Yes
1/1/4          actor     No    No    Yes   Yes   Yes   Yes   Yes      Yes
1/1/4          partner   No    No    Yes   Yes   Yes   Yes   Yes      Yes
===============================================================================
```

After changing the LAG port priorities from default (32768) to 10, the LAG on PE-1 is in up/up state and the ports are in up/active/up status. This show command also displays actor and partner bits set in the LACP messages.

Step 8 - MC-LAG configuration on PE-3 and PE-4.

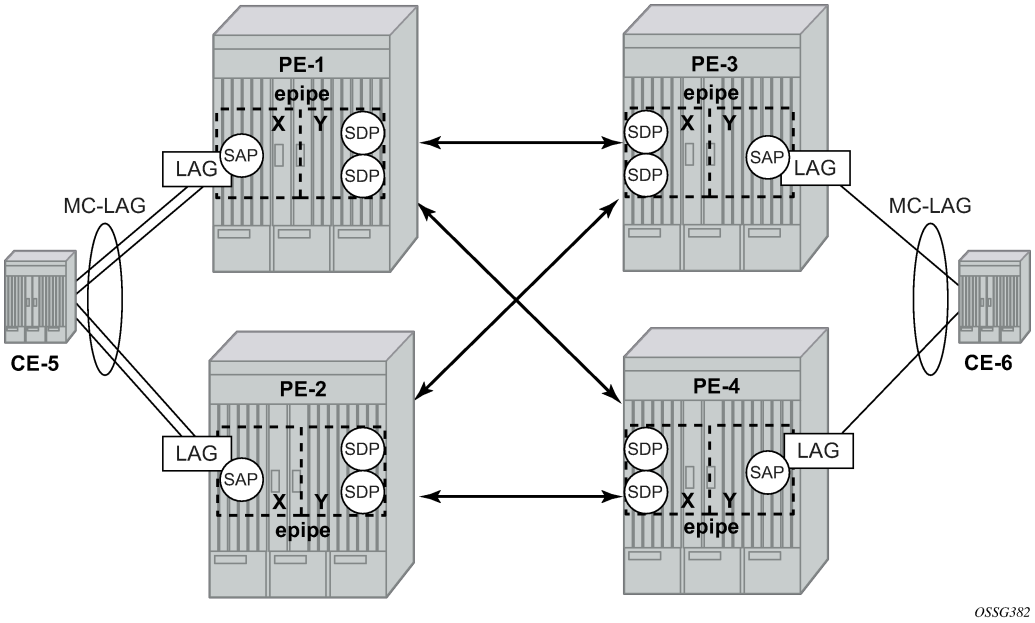
The MC-LAG configuration on PE-3 and PE-4 is similar to the configuration on PE-1 and PE-2. In this case, the priority of the LAG port on PE-4 is lowered to obtain the behavior in Figure 12: Network resiliency where the LAGs on PE-1 and PE-4 are active.

Step 9 - Pseudowire configuration.

Configure an Epipe service on every PE and create endpoints *X* and *Y* (the endpoint names can be any text string). Traffic can only be forwarded between two endpoints, for example, it is not possible for objects associated with the same endpoint to forward traffic to each other.

Associate the SAPs and spoke SDPs with the endpoints as shown in Figure 13: Association of SAPs/SDPs and endpoints.

Figure 13: Association of SAPs/SDPs and endpoints




```
# on PE-1:
configure
    service
        epipe 1 name "Epipe 1" customer 1 create
            endpoint "X" create
            exit
            endpoint "Y" create
            exit
            service-mtu 1400
            sap lag-1 endpoint "X" create
            exit
            spoke-sdp 13:1 endpoint "Y" create
            exit
            spoke-sdp 14:1 endpoint "Y" create
            exit
            no shutdown
        exit
    exit
exit
```

Likewise, an Epipe service, endpoints, SAPs and spoke SDPs must be configured on the other PE routers.

Step 10 - Pseudowire verification.

```
*A:PE-1# show service service-using

===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Up   1          Epipe 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-2# show service service-using

===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Down 1          Epipe 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-3# show service service-using

===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Down 1          Epipe 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
===============================================================================
```

```
*A:PE-4# show service service-using

===============================================================================
Services
===============================================================================
ServiceId    Type      Adm  Opr  CustomerId Service Name
-------------------------------------------------------------------------------
1            Epipe     Up   Up   1          Epipe 1
2147483648   IES       Up   Down 1          _tmnx_InternalIesService
2147483649   intVpls   Up   Down 1          _tmnx_InternalVplsService
-------------------------------------------------------------------------------
Matching Services : 3
-------------------------------------------------------------------------------
```

```
===============================================================================
```

The Epipe service on PE-2 and PE-3 is down and up on PE-1 and PE-4. This reflects the standby behavior shown in Figure 12: Network resiliency. However, after configuring ICB spoke SDPs (described later in this chapter), the Epipe will be in up/up status on all PE routers.

Step 11 - Verify SDP status

Local pseudowire bits indicate the status of the pseudowire on the PE node. These pseudowire bits will be sent to the peer. Peer pseudowire bits indicate the status of the pseudowire on the peer, as sent by the peer. The following example is taken on PE-2:

```
*A:PE-2# show service id 1 sdp 23:1 detail

===============================================================================
Service Destination Point (Sdp Id : 23:1) Details
===============================================================================
-------------------------------------------------------------------------------
 Sdp Id 23:1  -(192.0.2.3)
-------------------------------------------------------------------------------
Description     : (Not Specified)
SDP Id            : 23:1                    Type              : Spoke
Spoke Descr     : (Not Specified)
VC Type           : Ether                   VC Tag            : n/a
Admin Path MTU    : 0                       Oper Path MTU     : 1492
Delivery          : MPLS
Far End           : 192.0.2.3               Tunnel Far End    :
Oper Tunnel Far End: 192.0.2.3
LSP Types         : LDP
Hash Label        : Disabled                Hash Lbl Sig Cap  : Disabled
Oper Hash Label   : Disabled
Entropy Label     : Disabled

Admin State       : Up                      Oper State        : Up
MinReqd SdpOperMTU : 1400
Acct. Pol         : None                    Collect Stats     : Disabled
Ingress Label     : 524283                  Egress Label      : 524282
Ingr Mac Fltr-Id  : n/a                     Egr Mac Fltr-Id   : n/a
Ingr IP Fltr-Id   : n/a                     Egr IP Fltr-Id    : n/a
Ingr IPv6 Fltr-Id : n/a                     Egr IPv6 Fltr-Id  : n/a
Admin ControlWord : Not Preferred           Oper ControlWord  : False
Admin BW(Kbps)    : 0                       Oper BW(Kbps)     : 0
BFD Template      : None
BFD-Enabled       : no                      BFD-Encap         : ipv4
Last Status Change : 01/07/2020 12:51:04    Signaling         : TLDP
Last Mgmt Change  : 01/07/2020 12:50:53
Endpoint          : Y                       Precedence        : 4
PW Status Sig     : Enabled
Force Vlan-Vc     : Disabled                Force Qinq-Vc     : none
Class Fwding State : Down
Flags             : None
Local Pw Bits     : lacIngressFault lacEgressFault pwFwdingStandby
Peer Pw Bits      : lacIngressFault lacEgressFault pwFwdingStandby
Peer Fault Ip     : None
Peer Vccv CV Bits  : lspPing bfdFaultDet
Peer Vccv CC Bits  : mplsRouterAlertLabel

---snip---

-------------------------------------------------------------------------------
Segment Routing
-------------------------------------------------------------------------------
ISIS              : disabled
```

```
OSPF               : disabled
TE-LSP             : disabled
-------------------------------------------------------------------------------
Number of SDPs : 1
-------------------------------------------------------------------------------
===============================================================================
```

In this example, the remote side of the SDP is sending lacIngressFault lacEgressFault pwFwdingStandby flags. This is because the Epipe service on PE-3 is down because the MC-LAG is in standby/down status.

Link and node protection can be tested. The access links are protected by the MC-LAG, the PE routers are protected by the combination of MC-LAG/pseudowire redundancy. The SDPs can be protected by FRR, unless GRE is used.

Revertive behavior is expected when different MC-LAG port priorities are configured or if the number of MC-LAG ports is different on the MC-LAG peers: convergence takes place when the active PE fails and convergence takes place again when that PE is online again.

In case of revertive behavior, MC-LAG convergence might take less time than the setup of the spoke SDPs, thus creating a temporary black-hole. To avoid this situation, it is best to configure **hold-time up** on the LAG ports. In that case, the ports are kept in a down state for a configured period of time after the node has rebooted. This is done to ensure that the SDPs are operationally up when the MC-LAG convergence takes place. The **hold-time up** is expressed in seconds.

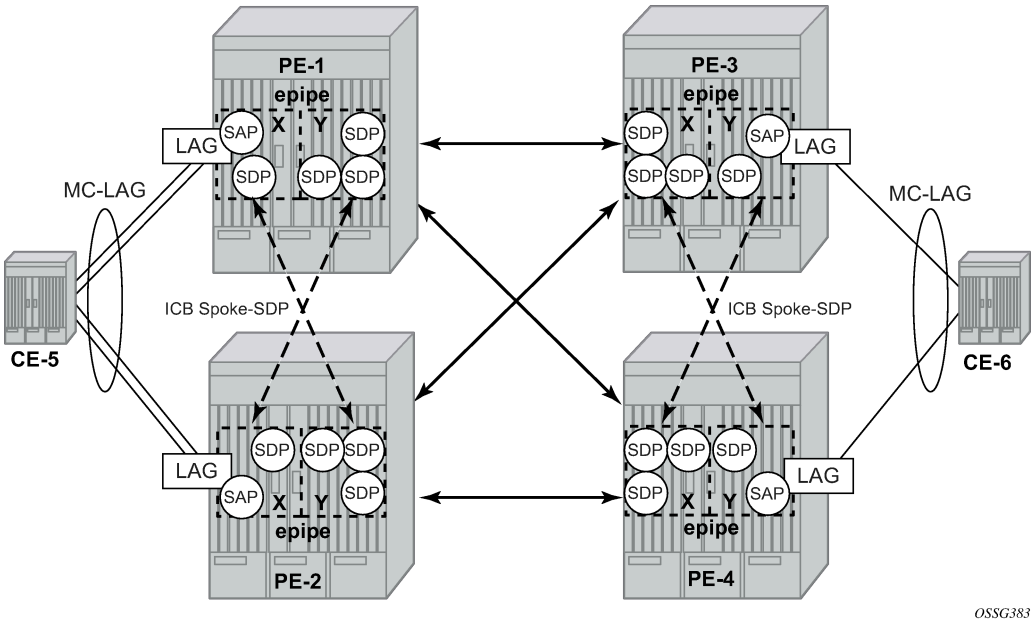
```
# on PE-1:
configure
    port 1/1/3
        ethernet
            hold-time up 50
        exit
    exit
    port 1/1/4
        ethernet
            hold-time up 50
        exit
    exit
```

Step 12 - Inter-Chassis Backup (ICB) pseudowire configuration.

In this setup, the configuration of ICBs is optional. It can be used to speed up convergence by forwarding in-flight packets during MC-LAG transition. Figure 15: Additional setup example 1 shows some setup examples where ICBs are required. ICBs cannot be configured at endpoints where the other object is a standard SAP, only MC-LAG SAPs and pseudowires are allowed with ICBs.

ICB SDPs and associated to endpoints as shown in Figure 14: ICB spoke SDPs and their association with the endpoints.

Figure 14: ICB spoke SDPs and their association with the endpoints



Two ICB spoke SDPs must be configured in the Epipe service on each PE router, one in each endpoint. Different SDP IDs can be used for the ICBs (as opposed to the regular pseudowires), but this is not necessary, because the far-end will be the same. The **vc-id** must be different however.

The ICB spoke SDPs must cross, one end should be associated with endpoint *X* and the other end (on the other PE) should be associated with endpoint *Y*. After configuring the ICB spoke SDPs, the Epipe service will be up/up on all four PE routers.

Only one spoke SDP will be forwarding. If there is an ICB and a MC-LAG SAP in an endpoint, the ICB will only forward if the SAP goes down. If an ICB resides in an endpoint together with other spoke SDPs, the ICB will only forward if there is no other active spoke SDP.

The following output shows the additional Epipe service configuration on each PE:

```
# on PE-1:
configure
    service
        epipe 1
            spoke-sdp 12:1 endpoint "X" icb create
            exit
            spoke-sdp 12:2 endpoint "Y" icb create
            exit
```

```
# on PE-2:
configure
    service
        epipe 1
            spoke-sdp 21:1 endpoint "Y" icb create
            exit
            spoke-sdp 21:2 endpoint "X" icb create
            exit
```

```
# on PE-3:
```

```
configure
    service
        epipe 1
            spoke-sdp 34:1 endpoint "X" icb create
            exit
            spoke-sdp 34:2 endpoint "Y" icb create
            exit
```

```
# on PE-4:
configure
    service
        epipe 1
            spoke-sdp 43:1 endpoint "Y" icb create
            exit
            spoke-sdp 43:2 endpoint "X" icb create
            exit
```

Step 13 - Verification of active objects for each endpoint.

The following command shows which objects are configured for each endpoint and which is the active object at this moment:

```
*A:PE-1# show service id 1 endpoint

===============================================================================
Service 1 endpoints
===============================================================================
Endpoint name                : X
Description                  : (Not Specified)
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Standby Signaling Master     : false
Standby Signaling Slave      : false
Tx Active                    : lag-1
Tx Active Up Time            : 0d 00:09:47
Revert Time Count Down       : never
Tx Active Change Count       : 1
Last Tx Active Change        : 01/07/2020 12:49:52
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
SAP      : lag-1                                 Oper Status: Up
Spoke-sdp: 12:1 Prec:4 (icb)                     Oper Status: Up
===============================================================================
Endpoint name                : Y
Description                  : (Not Specified)
Creation Origin              : manual
Revert time                  : 0
Act Hold Delay               : 0
Standby Signaling Master     : false
Standby Signaling Slave      : false
Tx Active (SDP)              : 14:1
Tx Active Up Time            : 0d 00:06:02
Revert Time Count Down       : never
Tx Active Change Count       : 2
Last Tx Active Change        : 01/07/2020 12:53:37
-------------------------------------------------------------------------------
Members
-------------------------------------------------------------------------------
Spoke-sdp: 12:2 Prec:4 (icb)                     Oper Status: Up
Spoke-sdp: 13:1 Prec:4                           Oper Status: Up
Spoke-sdp: 14:1 Prec:4                           Oper Status: Up
```

```
===============================================================================
===============================================================================
```

On PE-1, the SAP and the spoke SDP 14:1 are active. The other objects do not forward traffic.

Step 14 - Other types of setups.

Figure 15: Additional setup example 1 and Figure 16: Additional setup example 2 show other setups that combine MC-LAG and pseudowire redundancy.

Figure 15: Additional setup example 1

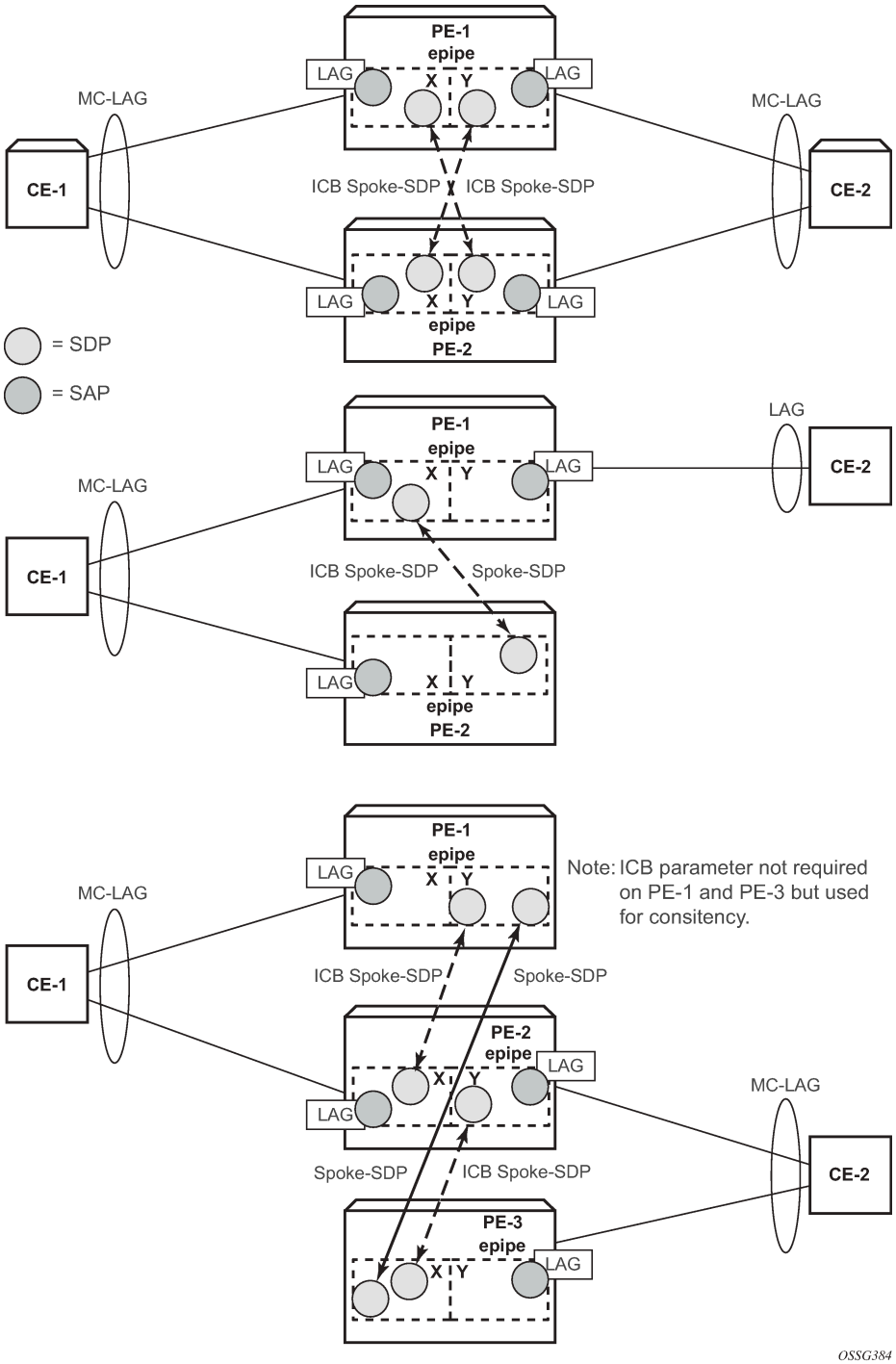
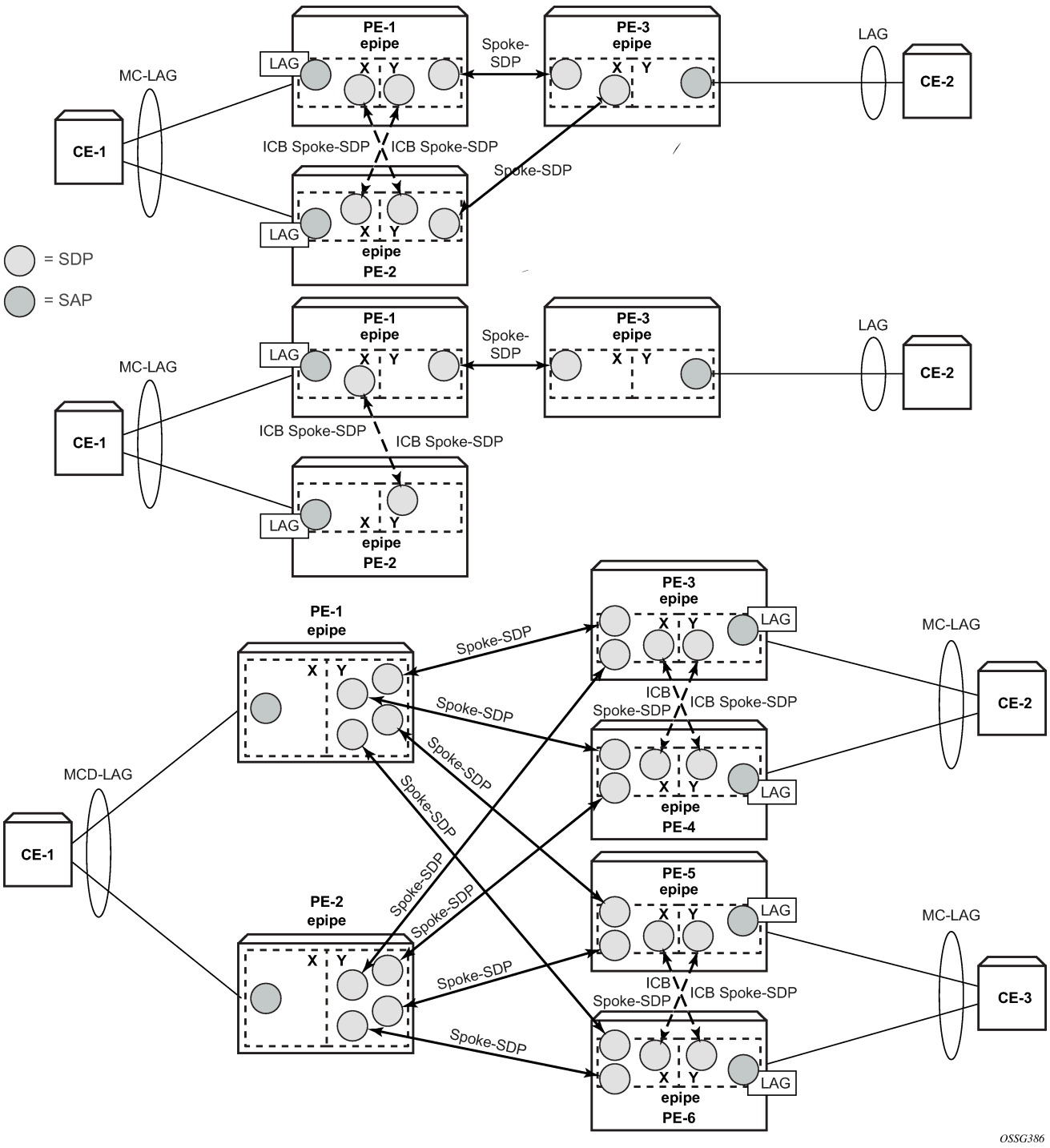


Figure 16: Additional setup example 2



MC-LAG for VPLS services

MC-LAG can also be configured for VPLS services. When the MC-LAG converges, the PE that transitions to standby state for the MC-LAG will send out an LDP address withdrawal message to all peers configured in the VPLS service. Both types of SDPs (spoke and mesh) support this feature. The PE peers will then flush all the MAC addresses learned via the PE that sent the LDP MAC address withdrawal message.

Because a VPLS service is a multipoint service, pseudowire redundancy is not required. The MC-LAG redundancy configuration is identical.

Forced switchover

MC-LAG convergence can be forced with the **tools perform lag** command:

```
*A:PE-1# tools perform lag force
  - force all-mc {active|standby}
  - force lag-id <lag-id> [sub-group <sub-group-id>] {active|standby}
  - force peer-mc <ip-address> {active|standby}

 <lag-id>             : [1..800]
 <sub-group-id>       : [1..16]
 <all-mc>             : keyword
 <ip-address>         : ipv4-address   - a.b.c.d
                        ipv6-address   - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
                                         d - [0..255]D
 <active|standby>     : keywords
```

```
*A:PE-1# tools perform lag force lag-id 1 standby
```

```
*A:PE-1# show lag 1

===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
-------------------------------------------------------------------------------
1              up      down    No       0         0        standby
===============================================================================
```

After the forced switchover, it is important to clear the forced switchover:

```
*A:PE-1# tools perform lag clear-force
  - clear-force all-mc
  - clear-force lag-id <lag-id> [sub-group <sub-group-id>]
  - clear-force peer-mc <ip-address>

 <lag-id>             : [1..800]
 <sub-group-id>       : [1..16]
 <all-mc>             : keyword
 <ip-address>         : ipv4-address   - a.b.c.d
                        ipv6-address   - x:x:x:x:x:x:x:x   (eight 16-bit pieces)
                                         x:x:x:x:x:x:d.d.d.d
                                         x - [0..FFFF]H
```

```
                                        d - [0..255]D
```

```
*A:PE-1# tools perform lag clear-force lag-id 1
```

```
*A:PE-1# show lag 1

===============================================================================
Lag Data
===============================================================================
Lag-id         Adm     Opr     Weighted Threshold Up-Count MC Act/Stdby
-------------------------------------------------------------------------------
1              up      up      No       0         2        active
===============================================================================
```

Conclusion

MC-LAG is a Nokia added value redundancy feature that offers fast access link convergence in Epipe and VPLS services for CE devices that support standard LACP. PE node convergence for VPLS services is enhanced by using LDP address withdrawal messages to flush the FDB on the PE peers. PE node convergence for Epipes is guaranteed by using pseudowire redundancy.

Port Cross-Connect (PXC)

This chapter provides information about Port Cross-Connect (PXC).

Topics in this chapter include:

- Applicability
- Overview
- Configuration
- Conclusion

Applicability

The chapter was initially written for SR OS Release 14.0.R5, but the CLI in the current edition is based on SR OS Release 21.2.R2.

Overview

The Port Cross-Connect (PXC) feature allows for a port, or number of ports, to be logically looped to themselves. The purpose of looping a port in this manner is to provide an "anchor point" function, such that traffic may ingress the node through any interface/port and be redirected to that anchor point.

When traffic is passed through the egress data path of the PXC, it can be used for additional packet processing that cannot be supported on the ingress data path, such as the removal of an encapsulation header. When traffic is looped back to the ingress data path of the PXC, it is processed as if it were the conventional service termination point. This essentially decouples the Input/Output (I/O) port through which packets ingress the node from the I/O port that implements the service termination. This decoupling removes the previous constraint for pseudowire-port (pw-port) whereby the I/O port through which packets ingress and egress the node was bound and could not be changed during, for example, a reconvergence event.

PXC provides two modes of operation: Distributed Versatile Service Module (DVSM) mode and Application Specific (AS) mode.

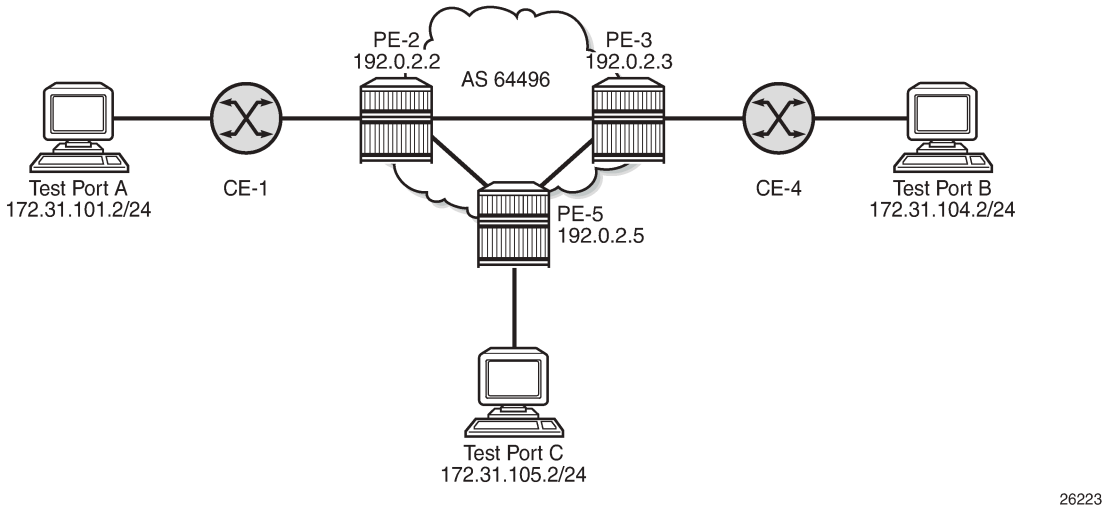
- The DVSM mode provides functionality like that of the VSM2 card, enabling the user to create an internal loopback through the card. This allows for back-to-back configurations similar to a VLAN cross-connect.
- The AS mode creates a Forwarding Path Extension (FPE) context through which the system can automatically create cross-connects to simplify user provisioning. Use-case examples for AS mode include PW port for business VPN services, VXLAN termination on a non-system interface, ESM over Pseudowire, and GRE tunnel termination.

This chapter describes the generic principles of PXC, combined with examples of both DVSM mode and AS mode.

Example topology

The topology shown in Figure 17: Example topology is used within this chapter to illustrate the use of PXC. PE-2, PE-3, and PE-5 form part of Autonomous System 64496 and run IS-IS level 2 together with LDP for the MPLS control plane. PE-2, PE-3, and PE-5 also peer in IBGP for the VPN-IPv4 address family. Test ports are connected to all PEs (in the case of PE-2 and PE-3, via CE routers) for the purpose of validating IP connectivity.

Figure 17: Example topology



PE-5 will host the PXC.

Configuration

PXC configuration

A PXC can consist of a single non-redundant port, or for redundancy and increased capacity, can consist of multiple ports that form member links of a Link Aggregation Group (LAG). Both options are described here.

Non-redundant PXC

The non-redundant PXC is created within the **port-xc** context and can be numbered from 1 to 64. A port must be assigned to the PXC before it is put into a **no shutdown** state, and that port must be in a **shutdown** state when it is assigned. There is no requirement for any kind of optical transceiver to be inserted in the port assigned to the PXC; it is only a logical loopback. When the port is assigned to the

PXC, it cannot be used for any other purpose besides a PXC-based service assignment (for example, a regular SAP could not be configured on this port).

```
# on PE-5:
configure
    port-xc
        pxc 1 create
            description "PXC non-redundant"
            port 1/2/1
            no shutdown
        exit
    exit
```

After the PXC has been put into a **no shutdown** state, two PXC sub-ports are automatically created by the system. The PXC sub-ports are identified by *.a* and *.b* suffixes of the parent PXC (in this example, pxc-1) and are created in hybrid mode with an MTU of 8700 bytes, both of which are non-configurable. The 8700-byte MTU represents the default port MTU (in this example, 8704 bytes) minus four bytes to allow for an internal VLAN tag that is used to identify each back-to-back sub-port. Finally, the encapsulation is set to dot1q, which is the default for hybrid ports. Q-in-Q encapsulation is also supported. It is also possible to configure dot1q encapsulation on one PXC sub-port and Q-in-Q encapsulation on the opposing PXC sub-port if, for example, there is a requirement to expose more VLAN tags on one side of the loop than the other side of the loop.

```
*A:PE-5# show port pxc 1
===============================================================================
Ports on Port Cross Connect 1
===============================================================================
Port          Admin Link Port    Cfg  Oper LAG/ Port Port Port   C/QS/S/XFP/
Id            State      State   MTU  MTU  Bndl Mode Encp Type   MDIMDX
-------------------------------------------------------------------------------
pxc-1.a       Down  Yes  Link Up 8700 8700    - hybr dotq xgige
pxc-1.b       Down  Yes  Link Up 8700 8700    - hybr dotq xgige
===============================================================================
```

After the PXC creation, the PXC sub-port CLI configuration is automatically generated and can be accessed in the same way as a conventional physical port, using the syntax "port pxc-n.l" where "n" represents the assigned PXC number and "l" represents the sub-port letter (a or b). As shown in the previous output, the sub-ports are in an admin down state following automatic creation and need to be manually put into a **no shutdown** state, as follows:

```
# on PE-5:
configure
    port pxc-1.a
        no shutdown
    exit
    port pxc-1.b
        no shutdown
    exit
```

The physical port assigned to the PXC must also now be put into a **no shutdown** state in order for the PXC to become operational:

```
# on PE-5:
configure
    port 1/2/1
        no shutdown
    exit
```

The command in the following output can then be used to verify the operational state of the PXC:

```
*A:PE-5# show port-xc pxc 1

===============================================================================
Port Cross-Connect Information
===============================================================================
PXC    Admin     Oper      Port        Description
Id     State     State     Id
-------------------------------------------------------------------------------
1      Up        Up        1/2/1       PXC non-redundant
===============================================================================
```

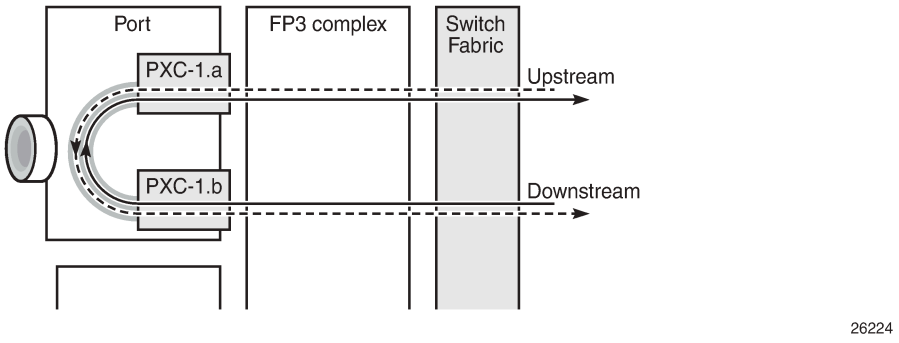
Similarly, the operational state of each of the sub-ports can be verified as follows. The physical link is indicated as being present even though there is no transceiver installed in this port.

```
*A:PE-5# show port pxc-1.a

===============================================================================
Ethernet Interface
===============================================================================
Description        : Port cross-connect
Interface          : pxc-1.a                    Oper Speed       : 10 Gbps
Link-level         : Ethernet                   Config Speed     : N/A
Admin State        : up                         Oper Duplex      : full
Oper State         : up
Config Duplex      : N/A
Physical Link      : Yes                        MTU              : 8700
Single Fiber Mode  : No                         Min Frame Length : 64 Bytes
IfIndex            : 1090523137                 Hold time up     : 0 seconds
Last State Change  : 05/18/2021 09:15:26        Hold time down   : 0 seconds
Last Cleared Time  : N/A
Phys State Chng Cnt: 0
---snip---
```

Figure 18: Non-redundant PXC shows a representation of the non-redundant PXC configuration. Both upstream and downstream traffic will pass twice through the FP data-path and port. For example, downstream traffic passes through the FP complex and PXC-1.b. The traffic is then looped back to PXC-1.a, and back into the FP complex. Similarly, upstream traffic passes through the FP complex to PXC-1.a. It is then looped back to PXC-1.b and back into the FP complex.

Figure 18: Non-redundant PXC



When using a PXC, the physical port effectively simulates two (sub-)ports, which creates two egress traffic paths: one upstream and one downstream. When the receive side of the PXC port receives those paths, it needs to distinguish between them, and this is where the internal additional VLAN tag is used.

The difference between this PXC configuration and a conventional port not looped or configured as PXC is as follows. With a conventional port, ingress traffic passes through the port and ingress data-path of the FP complex only once, and egress traffic passes through the egress data-path of the FP complex and port only once.

Redundant PXC

For a redundant PXC, the fundamental building blocks are identical to those of the non-redundant PXC, but there are a few additional configuration steps required to construct the LAGs to which the redundant PXC ports belong.

The redundant PXC example consists of two ports: 1/2/2 and 1/2/3 in the following output. In this case, the redundant PXC ports belong to the same IMM, but different IMMs can be used for increased redundancy. Two PXCs are created and each one is assigned one of the redundant PXC ports. Both PXCs are put into a **no shutdown** state.

```
# on PE-5:
configure
    port-xc
        pxc 2 create
            description "PXC redundant"
            port 1/2/2
            no shutdown
        exit
        pxc 3 create
            description "PXC redundant"
            port 1/2/3
            no shutdown
        exit
    exit
```

As with the non-redundant PXC, when the PXC has been put into a **no shutdown** state, two PXC sub-ports with .a and .b suffixes are automatically created by the system for each PXC port:

```
*A:PE-5# show port pxc [2..3]

===============================================================================
Ports on Port Cross Connect 2
===============================================================================
Port          Admin Link Port    Cfg  Oper LAG/ Port Port Port   C/QS/S/XFP/
Id            State      State   MTU  MTU  Bndl Mode Encp Type   MDIMDX
-------------------------------------------------------------------------------
pxc-2.a       Down  Yes  Link Up 8700 8700    - hybr dotq xgige
pxc-2.b       Down  Yes  Link Up 8700 8700    - hybr dotq xgige
===============================================================================

===============================================================================
Ports on Port Cross Connect 3
===============================================================================
Port          Admin Link Port    Cfg  Oper LAG/ Port Port Port   C/QS/S/XFP/
Id            State      State   MTU  MTU  Bndl Mode Encp Type   MDIMDX
-------------------------------------------------------------------------------
pxc-3.a       Down  Yes  Link Up 8700 8700    - hybr dotq xgige
pxc-3.b       Down  Yes  Link Up 8700 8700    - hybr dotq xgige
```

```
===============================================================================
```

The PXC sub-ports, together with the physical port, must then all be put into a **no shutdown** state:

```
# on PE-5:
configure
    port pxc-2.a
        no shutdown
    exit
    port pxc-2.b
        no shutdown
    exit
    port pxc-3.a
        no shutdown
    exit
    port pxc-3.b
        no shutdown
    exit
    port 1/2/2
        no shutdown
    exit
    port 1/2/3
        no shutdown
    exit
```

After the associated components have been put into a **no shutdown** state, the operational state of the PXCs can be verified:

```
*A:PE-5# show port-xc pxc [2..3]

===============================================================================
Port Cross-Connect Information
===============================================================================
PXC    Admin     Oper      Port        Description
Id     State     State     Id
-------------------------------------------------------------------------------
2      Up        Up        1/2/2       PXC redundant
===============================================================================

===============================================================================
Port Cross-Connect Information
===============================================================================
PXC    Admin     Oper      Port        Description
Id     State     State     Id
-------------------------------------------------------------------------------
3      Up        Up        1/2/3       PXC redundant
===============================================================================
```

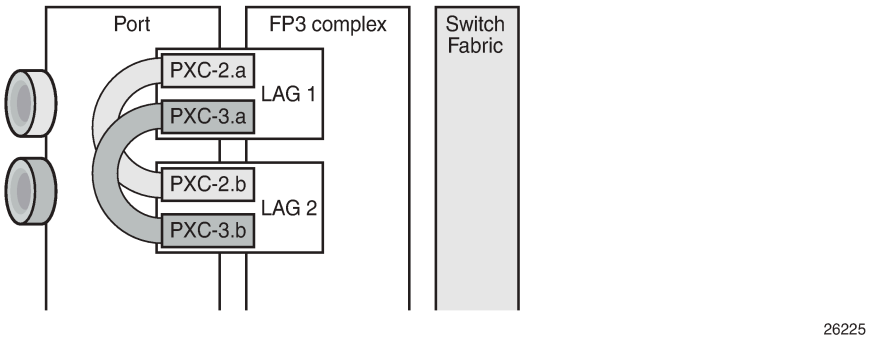
The PXC sub-ports are then associated with two LAGs to essentially form an internal back-to-back LAG. To do this, both sub-ports with the .a suffix belong to one LAG instance, and both sub-ports with the .b suffix belong to the other LAG instance. Like any other LAG member links, PXC sub-ports in a LAG must be configured with the same physical attributes, such as speed and duplex. Both LAG instances are configured with **mode hybrid** to match the mode of the physical ports. Setting the mode to **hybrid** automatically sets the **encap-type** to **dot1q**.

```
# on PE-5:
configure
    lag 1 name "lag-1"
        mode hybrid
        encap-type dot1q
```

```
        port pxc-2.a
        port pxc-3.a
        no shutdown
    exit
    lag 2 name "lag-2"
        mode hybrid
        encap-type dot1q
        port pxc-2.b
        port pxc-3.b
        no shutdown
    exit
```

Figure 19: PXC redundant mode with LAG shows a representation of the redundant PXC with LAG. Both upstream and downstream traffic will pass twice through the FP data-path and port.

Figure 19: PXC redundant mode with LAG



When the LAGs are configured and the associated PXC sub-ports assigned as member links, the operational status can be verified. Note that at the LAG level, each of the configured LAG instances is not aware that it is internally connected to another LAG instance, even though the member sub-ports are logically looped. It would be possible, for example, to put LAG 1 into an admin shutdown state and not affect the operational state of LAG 2. LACP is not supported for PXC LAG; however, it is possible to run the 802.3ah Ethernet in the First Mile (EFM) at PXC sub-port level, if required.

```
*A:PE-5# show lag 1 detail

===============================================================================
LAG Details
===============================================================================
Description        : N/A
-------------------------------------------------------------------------------
Details
-------------------------------------------------------------------------------
Lag-id              : 1                     Mode                : hybrid
Lag-name            : lag-1
Adm                 : up                    Opr                 : up
Thres. Last Cleared : 05/18/2021 08:35:32   Thres. Exceeded Cnt  : 0
Dynamic Cost        : false                 Encap Type          : dot1q
Configured Address  : 02:1f:ff:00:01:41     Lag-IfIndex         : 1342177281
Hardware Address    : 02:1f:ff:00:01:41     Adapt Qos (access)  : distribute
Hold-time Down      : 0.0 sec               Port Type           : standard
Per-Link-Hash       : disabled
Include-Egr-Hash-Cfg: disabled
Per FP Ing Queuing  : disabled              Per FP Egr Queuing  : disabled
Per FP SAP Instance : disabled
Access Bandwidth    : N/A                   Access Booking Factor: 100
```

```
Access Available BW : 0
Access Booked BW    : 0
LACP                : disabled
Standby Signaling   : lacp
Port hashing        : port-speed           Port weight speed    : 0 gbps
Ports Up            : 2
Weights Up          : 2                    Hash-Weights Up      : 20
Monitor oper group  : N/A
Adaptive loadbal.   : disabled             Tolerance            : N/A

-------------------------------------------------------------------------------
Port-id        Adm     Act/Stdby Opr     Primary   Sub-group     Forced  Prio
-------------------------------------------------------------------------------
pxc-2.a        up      active    up      yes       1            -       32768
pxc-3.a        up      active    up                1            -       32768
===============================================================================
```

DVSM mode

DVSM mode enables the creation of a back-to-back cross-connect. This back-to-back connection can be network-to-network, access-to-access, or a combination such as network-to-access. To provide an example of using DVSM mode, PE-3 in Figure 17: Example topology functions as a Layer 2 backhaul device, and PE-5 housing the PXC functions as the Layer 3 service edge. A pseudowire is extended from PE-3 to PE-5, where it is terminated in a VPRN, providing point-to-point connectivity between CE-4 and PE-5.

VLAN 100 is extended from CE-4 to PE-3, where it is indexed into an Epipe service. The SAP is service-delimiting; therefore, the VLAN is removed before frames are encapsulated into the pseudowire. The Epipe then has a single non-redundant spoke-SDP to PE-5 with VC-ID 11. The service configuration on PE-3 is as follows:

```
# on PE-3:
configure
    service
        sdp 35 mpls create
            far-end 192.0.2.5
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 11 name "Epipe 11" customer 1 create
            sap 1/1/3:100 create
                no shutdown
            exit
            spoke-sdp 35:11 create
                no shutdown
            exit
            no shutdown
        exit
```

At PE-5, the configuration of the corresponding end of the Epipe service is shown in the following output. This service consists of a single spoke-SDP toward PE-3 with VC-ID 11 to match the VC-ID advertised by PE-3, and a single SAP toward the PXC port. The syntax takes the form "pxc-n.l:vlan" where "n" is the PXC identifier, "l" is the sub-port letter (in this case .a), and "vlan" represents the VLAN identifier of the SAP.

As shown in the following output, the Epipe service uses PXC 1, which is the non-redundant PXC port. This is only an example; it could similarly use the redundant PXC port, in which case the SAP syntax would be the conventional LAG syntax (for example, lag-1:100, lag-2:100). Also note that although VLAN 100 is used both at PE-3's Epipe SAP and PE-5's Epipe PXC SAP, there is no correlation or dependence between the two. Both VLAN tags are service-delimiting and are subsequently stripped before the Ethernet frame is encapsulated into the pseudowire payload, so any valid VLAN value could be used at either point. The service configuration on PE-5 is as follows:

```
# on PE-5:
configure
    service
        sdp 53 mpls create
            far-end 192.0.2.3
            ldp
            keep-alive
                shutdown
            exit
            no shutdown
        exit
        epipe 11 name "Epipe 11" customer 1 create
            sap pxc-1.a:100 create
                no shutdown
            exit
            spoke-sdp 53:11 create
                no shutdown
            exit
            no shutdown
        exit
```

The VPRN configuration at the corresponding side of the PXC port is shown in the following output. The VPRN has two interfaces: the first is toward a directly connected test port used to verify IP connectivity, and the second ("to-CE-4") is toward CE-4 and has a SAP with a PXC syntax. The PXC syntax represents the same PXC and VLAN identifiers as the preceding Epipe configuration, but the PXC sub-port is .b, to represent the "other side" of the PXC logical loopback. Therefore, the VLAN values must match to create the back-to-back connection. Although not shown in the output (for brevity), a BGP session is configured between PE-5 and CE-4 for route exchange. The remainder of the VPRN parameters are generic and are not explained here.

```
# on PE-5:
configure
    service
        vprn 10 name "VPRN 10 using PXC DVSM" customer 1 create
            autonomous-system 64496
            interface "Test-Port-C" create
                address 172.31.105.1/24
                sap 1/1/3:100 create
                exit
            exit
            interface "to-CE-4" create
                address 192.168.45.2/30
                sap pxc-1.b:100 create
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 64496:10
                    vrf-import "vrf10-import"
```

```
                    vrf-export "vrf10-export"
                    no shutdown
                exit
            exit
            bgp
                group "EBGP"
            ---snip---
            no shutdown
        exit
```

PXC port dimensioning

When the VPRN service at PE-5 is put into a **no shutdown** state, the EBGP session to CE-4 is established. The relevant routes are exchanged between CE-4 and PE-5 and traffic can be exchanged between test ports B (connected to CE-4) and C (connected to PE-5). Initially, traffic is sent from test port B toward port C at a rate of 100 packets/s. Traffic is intentionally sent in only one direction (in this example) to emphasize a point regarding PXC port dimensioning and capacity planning, as follows.

The PXC in use by the Epipe/VPRN service is PXC 1, which uses physical port 1/2/1. The following output shows a snapshot of a monitor command against the physical port. Although traffic is only being sent in a single direction (test port B behind CE-4 toward test port C connected to PE-5), the input/output rate of packets per second is the same at 100 packets/s. This is because the physical port consists of two PXC sub-ports that are looped. In this example, traffic is output from pxc-1.a when traffic is sent from the Epipe SAP into the PXC port, and traffic is input at pxc-1.b when traffic is received by the VPRN SAP from the PXC port. Because both upstream/ingress traffic and downstream/egress traffic will be seen as output packets using the available capacity of the physical port, this needs to be considered when capacity is being planned.

```
*A:PE-5# monitor port 1/2/1 rate interval 3

===============================================================================
Monitor statistics for Port 1/2/1
===============================================================================
                                                    Input                 Output
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 3 sec (Mode: Rate)
-------------------------------------------------------------------------------
Octets                                             51600                  51600
Packets                                              100                    100
Errors                                                 0                      0
Bits                                              412800                 412800
Utilization (% of port capacity)                   ~0.00                  ~0.00
```

QoS continuity

The application of ingress/egress SAP QoS policies is fundamentally the same for a PXC-based SAP as it is for a conventional SAP. However, there is a difference with regard to how ingress Forwarding Class (FC) mappings are maintained throughout the PXC in DVSM mode. On a conventional SAP, ingress packets are classified and mapped to an FC. That FC mapping is maintained (as part of the fabric header) when the packet transits through the system and is ultimately used to define the egress queue and egress marking, such as MPLS EXP bits or dot1p bits.

However, the PXC sub-ports are subtly different. Consider SAP ingress traffic entering the VPRN at PE-5 from the locally connected test port C destined toward test port B at CE-4. At the ingress to PE-5, this traffic is mapped to FC Expedited Forwarding (EF) and forwarded into the PXC port through SAP pxc-1.b:100. When the traffic is forwarded out of the (PXC) SAP, the fabric header is removed as if it were a conventional SAP, and therefore, the information conveying the FC mapping is lost. When the traffic arrives at the opposing PXC sub-port SAP (in this case, pxc-1.a:100), a further FC classification is undertaken, and without some non-default configuration, traffic will be classified as FC Best Effort (BE). Therefore, it is a requirement to use non-default ingress/egress QoS policies through the PXC port in order to maintain FC continuity. A relatively simple way to do to this is through the use of dot1p markings.

To illustrate how this FC continuity is achieved, and in general how QoS is applied to PXC ports, an example of the relevant policies applied to PE-5's egress traffic toward CE-4 is used.

The first of the following outputs provides an example of the SAP-egress QoS policy applied at the VPRN PXC SAP (pxc-1.b:100). There are three classes in use: BE, Assured-Forwarding (AF), and EF. These FCs are remapped to queues 1, 2, and 3, respectively, and each queue is mapped to a parent H-QoS scheduler. Because the FCs must be maintained through the PXC loop, dot1p markings are used to distinguish between them. FC EF uses dot1p 5, FC AF uses dot1p 3, and FC BE uses dot1p 1. The SAP egress QoS policy is configured on PE-5 as follows:

```
# on PE-5:
configure
    qos
        sap-egress 2 name "SAP egress 2" create
            queue 1 create
                parent "aggregate-rate" level 2 weight 10
            exit
            queue 2 best-effort create
                parent "aggregate-rate" level 2 weight 40 cir-level 2
                rate 5000 cir max
            exit
            queue 3 expedite create
                parent "aggregate-rate" cir-level 3
                rate 2000 cir 2000
            exit
            fc af create
                queue 2
                dot1p 3
            exit
            fc be create
                queue 1
                dot1p 1
            exit
            fc ef create
                queue 3
                dot1p 5
            exit
        exit
```

The configuration of the Tier 1 scheduler "aggregate-rate" referenced by the child queues in the preceding SAP-egress QoS policy is shown in the following output. The scheduler in turn references a **port-scheduler-policy** using the command **port-parent**. Parenting to a port-scheduler is optional, but allows for inclusion of Preamble and Inter-Frame Gap (IFG) in the QoS scheduling algorithm, which is otherwise not included by a conventional H-QoS scheduler. The **port-scheduler-policy** "port-scheduler" is not referenced directly by the Tier 1 scheduler, but rather the port-scheduler is inherited by any child queues on

the port to which the port-scheduler is applied. In this case, the **port-scheduler-policy** "port-scheduler" is applied to the PXC sub-port pxc-1.b as follows:

```
# on PE-5:
configure
    qos
        port-scheduler-policy "port-scheduler" create
        exit
        scheduler-policy "egress-hqos-scheduler" create
            tier 1
                scheduler "aggregate-rate" create
                    port-parent
                    rate 1
                exit
            exit
        exit
    exit
    port pxc-1.b
        ethernet
            egress-scheduler-policy "port-scheduler"
        exit
        no shutdown
    exit
```

Finally, the SAP-egress QoS policy is applied to the PXC sub-port SAP within the **vprn interface** context. The H-QoS scheduler is also attached and an override of the rate configured. In summary, the SAP-egress QoS policy configuration looks exactly like that used on a conventional SAP, other than the dot1p markings used for FC continuity, which may not always be used or required.

```
# on PE-5:
configure
    service
        vprn 10 name "VPRN 10 using PXC DVSM" customer 1 create
            interface "to-CE-4" create
                address 192.168.45.2/30
                sap pxc-1.b:100 create
                    egress
                        scheduler-policy "egress-hqos-scheduler"
                        scheduler-override
                            scheduler "aggregate-rate" create
                                rate 20000
                            exit
                        exit
                        qos 2
                    exit
                exit
            exit
```

On the opposing side of the PXC loop, the dot1p markings imposed by the VPRN SAP egress are used to reclassify traffic back to its original FC mapping. The following output shows the SAP-ingress QoS policy applied at the Epipe PXC sub-port SAP (pxc-1.a:100). As shown in this output, dot1p 5 is mapped to FC EF, dot1p 3 is mapped to FC AF, and dot1p 1 is mapped to FC BE, thereby retaining the FC mappings through the PXC port.

```
# on PE-5:
configure
    qos
        sap-ingress 11 name "SAP ingress 11" create
            queue 1 create
            exit
```

```
            queue 2 best-effort create
                rate max cir max
            exit
            queue 3 expedite create
                rate max cir max
            exit
            fc "af" create
                queue 2
            exit
            fc "be" create
                queue 1
            exit
            fc "ef" create
                queue 3
            exit
            dot1p 1 fc "be"
            dot1p 3 fc "af"
            dot1p 5 fc "ef"
        exit
```

```
# on PE-5:
configure
    service
        epipe 11 name "Epipe 11" customer 1 create
            sap pxc-1.a:100 create
                ingress
                    qos 11
                exit
                no shutdown
            exit
        exit
```

The preceding configuration shows the required QoS policies for downstream traffic (VPRN egress to Epipe ingress). Corresponding QoS policies must also be configured for upstream traffic (Epipe egress to VPRN ingress). For brevity, they are not shown here.

AS mode

AS mode creates an FPE context that is used to provide information to the system about which PXC ports or LAGs are paired, so that the configuration process can be simplified by automatic provisioning of cross-connects. To illustrate the use of AS mode, the redundant PXC (formed of LAG 1 and 2) configured earlier in this chapter is used. However, redundancy is not a requirement. Non-redundant PXC ports can also be used with AS mode.

For AS mode, a similar setup to the DVSM example is used, with Epipe termination into a VPRN. This provides a generic view of the applicability of AS mode, but also allows a direct comparison between the DVSM and AS mode approaches. Again, PE-3 in Figure 17: Example topology functions as a Layer 2 backhaul device and PE-5 hosts the PXC functions as the Layer 3 service edge. A pseudowire is extended from PE-3 to PE-5 where it will be terminated in a VPRN, providing point-to-point connectivity between CE-4 and PE-5.

The following output illustrates the configuration of the Epipe service at PE-3. CE-4 uses Q-in-Q encapsulation on the PE-CE link to PE-3 with SVLAN tag 100 and CVLAN tag 1024. At PE-3, it is indexed into an Epipe service using a q.* SAP to make the CVLAN tag transparent (part of the payload). As the

spoke-SDP toward PE-5 is also configured with **force-vlan-vc-forwarding**, both SVLAN and CVLAN tags will be encapsulated in the pseudowire payload.

```
# on PE-3:
configure
    service
        epipe 13 name "Epipe 13" customer 1 create
            sap 1/1/3:100.* create
                no shutdown
            exit
            spoke-sdp 35:13 create
                force-vlan-vc-forwarding
                no shutdown
            exit
            no shutdown
        exit
```

As in the previous configuration example, LAG 1 and LAG 2 are used for PXC redundancy. LAG 1 has the PXC sub-ports pxc-2.a and pxc-3.a as member links, while LAG 2 has the PXC sub-ports pxc-2.b and pxc-3.b as member links. For AS mode, the next requirement is to configure the FPE construct and assign the paired LAG instances to that FPE. When entering the **fwd-path-ext** context, the **sdp-id-range** must be configured before any **fpe** instances can be created. The **sdp-id-range** allocates a block of SDP identifiers to be used for the automatic cross-connects between service applications and the FPE. Up to 128 SDP identifiers can be allocated in the range 1 to 17407.

After the **sdp-id-range** is configured, the **fpe** instance is created and the user enters the **fpe** context. The **path** command is used to assign redundant or non-redundant PXC objects to the FPE. In the case of a non-redundant FPE, the **path** command would refer to a **pxc** instance. In the case of a redundant FPE, the **path** syntax requires that each of the paired LAG instances is assigned to cross-connect "a" or cross-connect "b". Each FPE has two fundamental components, known as the transit side and the terminating side. The transit side is the side where additional traffic preprocessing is carried out, such as header removal or manipulation. It can be considered as the side closest to the network. The terminating side is the side where the preprocessed traffic is terminated in a service. When an FPE is used, the system automatically assigns cross-connect "a" to the transit side, and cross-connect "b" to the terminating side. In the following example, the command **path xc-a lag-1 xc-b lag-2** assigns LAG 1 to cross-connect "a" and LAG 2 to cross-connect "b". This means that LAG 1 is the transit side while LAG 2 is the terminating side.

The application of the FPE also needs to be configured. In this example, **pw-port** is selected to allow for support of pseudowire-SAP (including Enhanced Subscriber Management (ESM) over pseudowire). The other available options (for example, vxlan-termination) are beyond the scope of this chapter.

```
# on PE-5:
configure
    fwd-path-ext
        sdp-id-range from 17280 to 17407
        fpe 1 create
            path xc-a lag-1 xc-b lag-2
            pw-port
        exit
    exit
```

After the LAG instance is assigned to the FPE, it can no longer be used for other general purposes, such as IP interfaces and/or SAPs. Any attempt to do so is blocked in CLI. The operational state of the FPE can be verified as shown in the following output. It is also useful to be able to identify the services and

pw-ports that are mapped to an FPE. This can be obtained using the **show fwd-path-ext fpe <number> associations** command.

```
*A:PE-5# show fwd-path-ext fpe 1

===============================================================================
FPE Id: 1
===============================================================================
Description        : (Not Specified)
Path               : lag 1, lag 2
Pw Port            : Enabled                          Oper     : up
Sub Mgmt Extension : Disabled                         Oper     : N/A
Vxlan Termination  : Disabled                         Oper     : down
Segment-Routing V6 : Disabled
===============================================================================
```

The next step is to configure a pseudowire-port (pw-port) that will be used for terminating services. The creation of the **pw-port** creates a new context in which the only required configuration is to define the encapsulation type as dot1q or qinq. In this instance, the **pw-port** will support **encap-type qinq**.

```
# on PE-5:
configure
    pw-port 1 create
        encap-type qinq
    exit
```

The operational state of the pw-port is captured as a reference at this point, so that a comparison can be made later in the configuration process.

```
*A:PE-5# show pw-port 1

===============================================================
PW Port Information
===============================================================
PW Port   Encap          SDP:VC-Id               IfIndex
---------------------------------------------------------------
1         qinq           N/A                     1526726657
===============================================================
```

At PE-5, the requirement now is to link the spoke-SDP from PE-3 to the configured pw-port (pw-port 1) via the FPE. To do this, an Epipe service must be used that is configured for multi-segment pseudowire working, using the creation-time attribute **vc-switching**. The Epipe service consists of a single spoke-SDP toward PE-3 with a VC-ID matching that signaled by PE-3 (VC-ID 13). The second endpoint within the Epipe service uses the command **pw-port 1 fpe 1** to reference the previously configured pw-port and FPE objects. This command essentially creates an internal cross-connect between the Epipe service and the pw-port via the configured FPE object.

```
# on PE-5:
configure
    service
        epipe 13 name "Epipe 13" customer 1 vc-switching create
            pw-port 1 fpe 1 create
                no shutdown
            exit
            spoke-sdp 53:13 create
                no shutdown
            exit
            no shutdown
```

```
        exit
```

The following output shows the SDPs belonging to the preceding vc-switched Epipe service configured. The first SDP with identifier 53:13 is the pseudowire toward PE-3 with VC-ID 13. The second SDP has identifier 17280:1 allocated from the preconfigured **sdp-id-range**, and has a type of Fpe. In the configuration of **fpe 1**, the **path** command assigned LAG 1 to cross-connect "a" (**xc-a**) and LAG 2 to cross-connect "b" (**xc-b**). Also, cross-connect "b" is always automatically assigned to the terminate side of the FPE. Therefore, the Far End address is shown as fpe_1.b, in order to terminate the service.

```
*A:PE-5# show service id 13 sdp

===============================================================================
Services: Service Destination Points
===============================================================================
SdpId            Type     Far End addr    Adm     Opr       I.Lbl     E.Lbl
-------------------------------------------------------------------------------
53:13            Spok     192.0.2.3       Up      Up        524280    524283
17280:1          Fpe      fpe_1.b         Up      Up        524282    524281
-------------------------------------------------------------------------------
Number of SDPs : 2
-------------------------------------------------------------------------------
===============================================================================
```

With the vc-switching Epipe service configured and operational, the state of the pw-port can again be shown in the following output. Before the configuration of the vc-switching Epipe, the pw-port had no SDP identifier or VC-ID. Now both entries exist; automatically created by the system when **pw-port 1 fpe 1** was configured as an endpoint within the vc-switching Epipe. The SDP identifier of 17281 is allocated from the preconfigured **sdp-id-range**.

```
*A:PE-5# show pw-port 1

===============================================================
PW Port Information
===============================================================
PW Port   Encap          SDP:VC-Id               IfIndex
---------------------------------------------------------------
1         qinq           17281:100001            1526726657
===============================================================
```

The output for SDP 17281 shows that the Far End is fpe_1.a (transit), the Delivery (Del) is MPLS, the LSP type is FPE (F), and that no signaling (Sig) is used for this internal SDP, as follows:

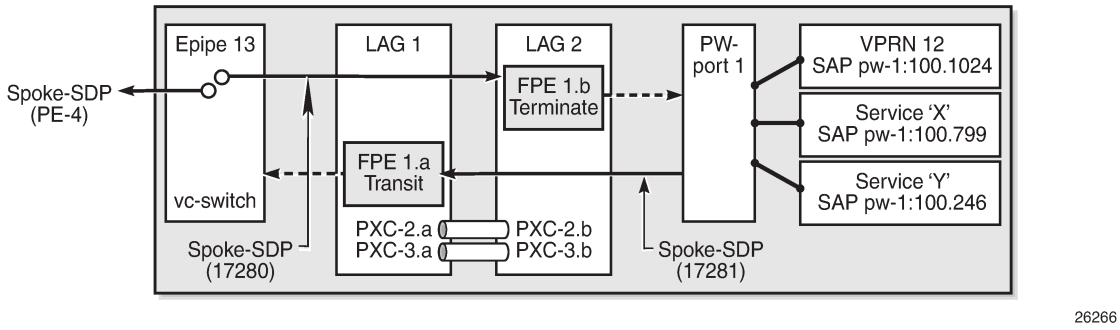
```
*A:PE-5# show service sdp 17281

===============================================================================
Service Destination Point (Sdp Id : 17281)
===============================================================================
SdpId  AdmMTU  OprMTU  Far End          Adm  Opr         Del     LSP   Sig
-------------------------------------------------------------------------------
17281  0       8678    fpe_1.a          Up   Up          MPLS    F     None
===============================================================================
```

In SR OS, the combination of SDP ID and VC-ID is always associated with a service. When using AS mode, the system automatically creates an internal VPLS service with ID 2147383649 and a name of _tmns_InternalVplsService. This VPLS includes all internal SDPs dynamically created for binding pw-ports to the transit side of the corresponding FPE. The VPLS is an internal construct that does not affect forwarding.

Figure 20: AS mode with redundant FPE shows the components of the FPE from vc-switching Epipe to pw-port.

Figure 20: AS mode with redundant FPE



Next, bind the VPRN service to the pw-port with the relevant VLAN delimiters. CE-4 is using SVLAN tag 100 and CVLAN tag 1024 and both VLANs are encapsulated inside the pseudowire as payload. The following VPRN configuration has two interfaces: the first is toward a directly connected test port used to verify IP connectivity, and the second is toward CE-4 and has a SAP with a pw-port syntax. The SAP pw-1:100.1024 represents pw-port 1 with Q-in-Q encapsulation using SVLAN tag 100 and CVLAN tag 1024 as service delimiters. Although not shown (for brevity), a BGP session is configured between PE-5 and CE-4 for route exchange. The remainder of the VPRN parameters are generic and are not explained here.

```
# on PE-5:
configure
    service
        vprn 12 name "VPRN 12 using PXC AS" customer 1 create
            autonomous-system 64496
            interface "Test-Port-C" create
                address 172.31.105.1/24
                sap 1/1/3:100 create
                exit
            exit
            interface "to-CE-4" create
                address 192.168.45.2/30
                sap pw-1:100.1024 create
                exit
            exit
            bgp-ipvpn
                mpls
                    auto-bind-tunnel
                        resolution any
                    exit
                    route-distinguisher 64496:12
                    vrf-target target:64496:12
                    no shutdown
                exit
            exit
            bgp
                group "EBGP"
            ---snip---
            no shutdown
            exit
        exit
```

FPE port dimensioning

After the VPRN service at PE-5 is put into a **no shutdown** state, the EBGP session to CE-4 is established. The relevant routes are exchanged between CE-4 and PE-5 and traffic can be exchanged between test ports B (behind CE-4) and C (connected to PE-5). Initially, traffic is sent unidirectionally from test port C (connected to PE-5) toward port B (connected to CE-4) at a rate of 100 packets/s. To provide a level of entropy for the generated traffic, 100 destination IP addresses are used in the range 172.31.104.2 through 172.31.104.101, and 100 source IP addresses are used in the range 172.31.105.2 through 172.31.105.101.

The following output shows a snapshot of a monitor command against LAG 2 (xc-b, or terminating side) incorporating both physical ports. First, note that the input and output rate of packets per second are equal at 100 packets/s, which is not intuitive for a unidirectional traffic flow. This is because the LAG statistics are essentially a copy of the physical port statistics and the physical port consists of two PXC sub-ports that are looped. Logically, this unidirectional traffic flow is forwarded in a single upstream direction from pxc-2.a/pxc-3.a to pxc-2.b/pxc-3.b. Physically, the unidirectional traffic is transmitted by ports 1/2/2 and 1/2/3, then received by the same ports through the loop. Second, note that traffic is load-balanced over both member links (PXC sub-ports) of the LAG. This is because conventional LAG load-balancing mechanisms are used for the FPE LAG, which in the case of a VPRN SAP-to-network relies on source/destination IP address (with optional Layer 4, which is not currently configured).

```
*A:PE-5# monitor lag 2 rate interval 3

===============================================================================
Monitor statistics for LAG ID 2
===============================================================================
Port-id      Input packets                    Output packets
             Input bytes                      Output bytes
             Input errors [Input util %]      Output errors [Output util %]
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 9 sec (Mode: Rate)
-------------------------------------------------------------------------------
1/2/2!       41                               41
             22041                            22041
             0                          0.00  0                          0.00
1/2/3!       59                               59
             32159                            32159
             0                          0.00  0                          0.00
-------------------------------------------------------------------------------
Totals       100                              100
             54200                            54200
             0                          0.00  0                          0.00
! indicates that the port is assigned to a port-xc.
```

Traffic is then generated unidirectionally upstream from test port B (connected to CE-4) toward port C (connected to PE-5) at a rate of 100 packets/s. Again, to provide a level of entropy for the generated traffic, 100 destination IP addresses are used in the range 172.31.105.2 through 172.31.105.101, and 100 source IP addresses are used in the range 172.31.104.2 through 172.31.104.101. The input/output rates of packets per second are the same, as previously explained. Again, traffic is load-balanced over both member links (PXC sub-ports). This is because hashing of traffic through a vc-switched Epipe service uses source/destination IP information (and optional Layer 4 information, which is not currently configured).

```
*A:PE-5# monitor lag 2 rate interval 3
```

```
===============================================================================
Monitor statistics for LAG ID 2
===============================================================================
Port-id      Input packets                    Output packets
             Input bytes                      Output bytes
             Input errors [Input util %]      Output errors [Output util %]
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
At time t = 9 sec (Mode: Rate)
-------------------------------------------------------------------------------
1/2/2!       44                               44
             23848                            23848
             0                          0.00  0                          0.00
1/2/3!       56                               56
             30352                            30352
             0                          0.00  0                          0.00
-------------------------------------------------------------------------------
Totals       100                              100
             54200                            54200
             0                          0.00  0                          0.00
! indicates that the port is assigned to a port-xc.
```

QoS continuity

When using AS mode, the FPE construct creates internal cross-connects between the vc-switching Epipe and the pw-port. These internal cross-connects function as MPLS tunnels that transit through internal network interfaces on the PXC sub-ports. The internal network interfaces use the default network policy 1 for egress marking and ingress classification/FC mapping. Like all default QoS policies, this network policy cannot be modified (or deleted). Also, it is not possible to use a non-default network policy, because there is no router interface to which the non-default policy can be attached.

The internal cross-connects also use the default network-queue policy named "default". While this policy also cannot be modified, it is possible to configure and apply a non-default network-queue policy (including a port-scheduler-policy, if required) at PXC sub-port level. An example of how this would be applied is shown in the following output. Where redundant PXC ports are used in an LAG instance, the queue-policy must be applied to the primary link of the LAG, which is then automatically applied to all other member links. (The primary link of the LAG can be identified using the command "show lag n port".)

```
# on PE-5:
configure
    port pxc-2.a
        ethernet
            network
                queue-policy "non-default"
            exit
        exit
        no shutdown
    exit
```

To demonstrate QoS continuity through the FPE, the following is established:

- **Downstream**: Traffic is generated from test port C (connected to PE-5) toward test port B (connected to CE-4) with DiffServ marking EF at a rate of 100 packets/s. At PE-5 SAP ingress, this traffic is mapped into FC EF.

- **Upstream**: Traffic is generated from test port B (connected to CE-4) toward test port C (connected to PE-5) with DiffServ marking EF at a rate of 100 packets/s. At PE-3, a SAP-ingress QoS policy is used to map the traffic into FC EF.
- The default network QoS policy 1 is used on all network interfaces at PE-3 and PE-5. On egress, this policy marks FC EF as MPLS EXP 5. On ingress, MPLS EXP 5 is mapped to FC EF.
- The default network queue-policy "default" is used on all network interfaces at PE-3 and PE-5. This maps FC EF traffic to queue 6 at ingress and egress.

First, QoS continuity for downstream traffic is validated. The following output shows the relatively simple SAP-egress QoS policy that is applied to the egress of the VPRN interface (pw-port) toward CE-4. No classification of traffic and mapping to FCs are present in the policy, because the classification and mapping have already taken place on the SAP ingress at PE-5 (the SAP facing the test port C).

```
# on PE-5:
configure
    qos
        sap-egress 12 create
            queue 1 create
                parent "aggregate-rate" level 2 weight 10
            exit
            queue 2 best-effort create
                parent "aggregate-rate" level 2 weight 40 cir-level 2
                rate 5000 cir max
            exit
            queue 3 expedite create
                parent "aggregate-rate" cir-level 3
                rate 2000 cir 2000
            exit
            fc af create
                queue 2
            exit
            fc be create
                queue 1
            exit
            fc ef create
                queue 3
            exit
        exit
```

The configuration of the Tier 1 scheduler "aggregate-rate" referenced by the child queues in the preceding SAP-egress QoS policy is as follows. The Tier 1 scheduler references a **port-scheduler-policy** using the command **port-parent**. Parenting to a port-scheduler is optional, but allows for inclusion of Preamble and IFG in the QoS scheduling algorithm, which otherwise are not included. The Tier 1 scheduler does not directly reference the **port-scheduler-policy** by name, but rather inherits any port-scheduler configured on the port to which the child queues are mapped. In this example, the port-scheduler-policy "port-scheduler" is applied to PXC sub-port pxc-2.b (terminating side). This is the primary link of LAG 2 and ensures that the same port-scheduler-policy is automatically applied to other member ports.

```
# on PE-5:
configure
    qos
        port-scheduler-policy "port-scheduler" create
        exit
        scheduler-policy "egress-hqos-scheduler" create
            tier 1
                scheduler "aggregate-rate" create
                    port-parent
                    rate 1
```

```
            exit
        exit
    exit
exit
port pxc-2.b
    ethernet
        egress-scheduler-policy "port-scheduler"
    exit
    no shutdown
exit
```

Finally, the SAP-egress QoS policy is applied to the pw-port SAP within the VPRN. The egress H-QoS scheduler is also attached and an override of the rate is configured.

```
# on PE-5:
configure
    service
        vprn 12 name "VPRN 12 using PXC AS" customer 1 create
            interface "to-CE-4" create
                sap pw-1:100.1024 create
                    egress
                        scheduler-policy "egress-hqos-scheduler"
                        scheduler-override
                            scheduler "aggregate-rate" create
                                rate 25000
                            exit
                        exit
                        qos 12
                    exit
                exit
            exit
```

When traffic is generated downstream toward CE-4 in FC EF at a rate of 100 packets/s, the first point of verification is the VPRN pw-port SAP egress. The following output is a **monitor** of the SAP showing that traffic is correctly mapped to queue 3.

```
*A:PE-5# monitor service id 12 sap pw-1:100.1024 rate

===============================================================================
Monitor statistics for Service 12 SAP pw-1:100.1024
===============================================================================
---snip---
-------------------------------------------------------------------------------
At time t = 11 sec (Mode: Rate)
-------------------------------------------------------------------------------
---snip---
-------------------------------------------------------------------------------
Sap per Queue Stats
-------------------------------------------------------------------------------
                        Packets                 Octets                  % Port
                                                                        Util.
---snip---
Egress Queue 3
For. In/InplusProf    : 0                       0                       0.00
For. Out/ExcProf      : 100                     51600                   0.04
Dro. In/InplusProf    : 0                       0                       0.00
Dro. Out/ExcProf      : 0                       0                       0.00
```

Monitoring of network interfaces does not show queue statistics (and is not supported on PXC sub-ports), but a verification of the sub-port statistics on the transit side (LAG 1) shows that packets are incrementing in ingress queue 6 on both sub-ports, as follows:

```
*A:PE-5# show port pxc-2.a detail | match "Ingress Queue  6" post-lines 4
Ingress Queue  6             Packets                 Octets
     In Profile forwarded  :    711                    382518
     In Profile dropped    :    0                      0
     Out Profile forwarded :    0                      0
     Out Profile dropped   :    0                      0
```

```
*A:PE-5# show port pxc-3.a detail | match "Ingress Queue  6" post-lines 4
Ingress Queue  6             Packets                 Octets
     In Profile forwarded  :    404                    217352
     In Profile dropped    :    0                      0
     Out Profile forwarded :    0                      0
     Out Profile dropped   :    0                      0
```

The last point of verification is the network egress interface toward PE-3. Again, a check at the physical port level shows that packets are incrementing in egress queue 6. Therefore, we can conclude that QoS/FC continuity is maintained in the downstream direction.

```
*A:PE-5# show port 1/1/2 detail | match "Egress Queue  6" post-lines 4
Egress Queue  6              Packets                 Octets
     In/Inplus Prof fwded  :    2394                   1297548
     In/Inplus Prof dropped:    0                      0
     Out/Exc Prof fwded    :    0                      0
     Out/Exc Prof dropped  :    0                      0
```

Next, the upstream QoS continuity is verified. PE-3 is marking traffic generated by test port B to FC EF, which in turn is marked as MPLS EXP 5 by PE-3's default network QoS policy. The following output taken at PE-5 shows that packets are incrementing in ingress queue 6 of the network interface toward PE-3 and confirms that traffic is correctly marked as FC EF at ingress.

```
*A:PE-5# show port 1/1/2 detail | match "Ingress Queue  6" post-lines 4
Ingress Queue  6             Packets                 Octets
     In Profile forwarded  :    3458                   1874236
     In Profile dropped    :    0                      0
     Out Profile forwarded :    0                      0
     Out Profile dropped   :    0                      0
```

The next point of verification is the egress side of the PXC sub-ports (pxc-2.a and pxc-3.a) forming the transit side (LAG 1). The sub-port statistics verify that packets are incrementing in egress queue 6 of both sub-ports (as traffic is being load-balanced).

```
*A:PE-5# show port pxc-2.a detail | match "Egress Queue  6" post-lines 4
Egress Queue  6              Packets                 Octets
     In/Inplus Prof fwded  :    12441                  6693258
     In/Inplus Prof dropped:    0                      0
     Out/Exc Prof fwded    :    0                      0
     Out/Exc Prof dropped  :    0                      0
```

```
*A:PE-5# show port pxc-3.a detail | match "Egress Queue  6" post-lines 4
Egress Queue  6              Packets                 Octets
     In/Inplus Prof fwded  :    12893                  6936434
     In/Inplus Prof dropped:    0                      0
     Out/Exc Prof fwded    :    0                      0
```

```
     Out/Exc Prof dropped  :    0                       0
```

PXC sub-ports operate in hybrid mode. When the upstream traffic arrives on the PXC sub-ports that form the terminating side of the FPE (pxc-2.b and pxc-3.b), it is mapped to the pw-port SAP-ingress queues, bypassing the ingress network QoS policy and associated ingress network queues. As a result, the MPLS EXP-to-FC mapping cannot be fulfilled and traffic requires reclassification and remapping to the correct FC by the SAP-ingress QoS policy. The following output shows the SAP-ingress QoS policy applied to the pw-port SAP within the VPRN. Because the EXP-to-FC mapping could not be completed, FC reclassification is required in order to map traffic to its original FC before transiting the FPE. In this example, DSCP is used. Also, FC EF is mapped to queue 3.

```
#on PE-5:
configure
    qos
      sap-ingress 12 create
            queue 1 create
                parent "aggregate-rate" level 2 weight 10
            exit
            queue 2 best-effort create
                parent "aggregate-rate" level 2 weight 40 cir-level 2
                rate 5000 cir max
            exit
            queue 3 expedite create
                parent "aggregate-rate" cir-level 3
                rate 2000 cir 2000
            exit
            queue 11 multipoint create
                rate max cir max
            exit
            fc "af" create
                queue 2
            exit
            fc "be" create
                queue 1
            exit
            fc "ef" create
                queue 3
            exit
            dscp af31 fc "af"
            dscp be fc "be"
            dscp ef fc "ef"
        exit
```

For completeness, the configuration of the Tier 1 scheduler "aggregate-rate" referenced by the child queues in the preceding SAP-ingress QoS policy is as follows. Unlike the egress counterpart, there is no parenting to a port-scheduler because this is an egress function only.

```
# on PE-5:
configure
    qos
        scheduler-policy "ingress-hqos-scheduler" create
            tier 1
                scheduler "aggregate-rate" create
                    rate 1
                exit
            exit
        exit
    exit
```

The SAP-ingress QoS policy is applied to the pw-port SAP within the VPRN, together with the ingress H-QoS scheduler. An override of the scheduler rate is also applied.

```
# on PE-5:
configure
    service
        vprn 12 name "VPRN 12 using PXC AS" customer 1 create
            interface "to-CE-4" create
                sap pw-1:100.1024 create
                    ingress
                        scheduler-policy "ingress-hqos-scheduler"
                        scheduler-override
                            scheduler "aggregate-rate" create
                                rate 25000
                            exit
                        exit
                        qos 12
                    exit
                exit
            exit
```

With the SAP-ingress policy applied, a monitor output of the SAP in the following output verifies that the packets are being received in queue 3 at a rate of 100 packets/s. This verifies the FC continuity in the upstream direction, noting that reclassification and remapping of FC is required at SAP ingress.

```
*A:PE-5# monitor service id 12 sap pw-1:100.1024 rate

===============================================================================
Monitor statistics for Service 12 SAP pw-1:100.1024
===============================================================================
---snip---
-------------------------------------------------------------------------------
Sap Statistics
-------------------------------------------------------------------------------
---snip---
                        Packets                 Octets
---snip---

Ingress Queue 3 (Unicast) (Priority)
Off. HiPrio           : 0                       0                       0.00
Off. LowPrio          : 100                     51647                   0.04
Dro. HiPrio           : 0                       0                       0.00
Dro. LowPrio          : 0                       0                       0.00
For. InProf           : 0                       0                       0.00
For. OutProf          : 100                     51647                   0.04
```

OAM continuity

The FPE pw-port functionality may be used by redundant routers to provide resilient service termination for a Layer 2 backhaul node implementing a mechanism such as active/standby pseudowire. In SR OS, an active/standby pseudowire is modeled as an Epipe or VPLS service with an endpoint object containing two spoke-SDPs. This form of redundancy relies on the propagation of the Pseudowire Status TLV within an LDP Notification message to convey the operational status of the pseudowires and thereby indicate which one of the pseudowires is active and which one is standby.

The FPE construct uses the concept of a multi-segment pseudowire, implementing Switching-PE (S-PE) functionality to instantiate dynamic cross-connects through the FPE. To verify that LDP status signaling is maintained through this S-PE function, the following is established:

- The Epipe service at PE-3 used for Layer 2 backhaul to the FPE is modified to include an **endpoint** object referenced by two spoke-SDPs.
- The first spoke-SDP has a far end of PE-2 and is configured as **precedence primary**, so becomes the active pseudowire.
- The second spoke-SDP has a far end of PE-5 and is configured with the default precedence 4, so becomes the standby pseudowire.
- Because the endpoint object is configured for **standby-signaling-master**, PE-3 will signal a status of standby toward PE-5.

For completeness, the configuration of the Epipe service at PE-3 is as follows:

```
# on PE-3:
configure
    service
       epipe 13 name "Epipe 13" customer 1 create
            endpoint "redundant-Layer3" create
                standby-signaling-master
            exit
            sap 1/1/3:100.* create
                no shutdown
            exit
            spoke-sdp 32:13 endpoint "redundant-Layer3" create
                precedence primary
                no shutdown
            exit
            spoke-sdp 35:13 endpoint "redundant-Layer3" create
                no shutdown
            exit
            no shutdown
        exit
```

As shown in the following output, PE-3 has the spoke-SDP to PE-5 (sdp 35:13) as administratively and operationally up, but is signaling a status of standby (pwFwdingStandby).

```
*A:PE-3# show service id 13 sdp 35:13 detail | match expression
                                          "Local Pw Bits|Peer Pw Bits|Admin State"
Admin State        : Up                      Oper State        : Up
Local Pw Bits      : pwFwdingStandby
Peer Pw Bits       : None
Admin State        : Disabled                Oper State        : Disabled
```

At PE-5, the signaled status is acknowledged at the far end of the pseudowire in the Peer Pw Bits field.

```
*A:PE-5# show service id 13 sdp 53:13 detail | match expression
                                           "Local Pw Bits|Peer Pw Bits|Admin State"
Admin State        : Up                      Oper State        : Up
Local Pw Bits      : None
Peer Pw Bits       : pwFwdingStandby
Admin State        : Disabled                Oper State        : Disabled
```

Typically, an S-PE would propagate the status TLV received from one pseudowire segment into the opposing pseudowire segment in order to provide end-to-end status signaling. However, when using FPE, the SR OS Service Manager process correlates between a pseudowire and its corresponding pw-port SAPs, so can take the necessary actions based upon the operational state of each. Therefore, it is not necessary for the S-PE to propagate the status TLV from one segment to another. This is illustrated in the following output at PE-5, which shows the second segment of the multi-segment pseudowire toward the

terminating side fpe_1.b. As described, the status bits are not copied between single segments and all local/peer pseudowire bits remain unset.

```
*A:PE-5# show service id 13 sdp 17280:1 detail | match expression
                                          "Admin State|Peer Pw Bits|Local Pw Bits"
Admin State        : Up                    Oper State        : Up
Local Pw Bits      : None
Peer Pw Bits       : None
Admin State        : Disabled              Oper State        : Disabled
```

The pw-port 1 used throughout in this example is internally bound to SDP 17281, as shown in the first of the following outputs. The second output shows that this SDP is operationally down with the flag "stitchingSvcTxDown".

```
*A:PE-5# show pw-port 1

===============================================================
PW Port Information
===============================================================
PW Port   Encap          SDP:VC-Id               IfIndex
---------------------------------------------------------------
1         qinq           17281:100001            1526726657
===============================================================
```

```
*A:PE-5# show service sdp 17281 detail | match "SDP: 17281 Pw-port: 1" post-lines 10
SDP: 17281 Pw-port: 1
-------------------------------------------------------------------------------
VC-Id               : 100001               Admin Status       : up
Encap               : qinq                 Oper Status        : down
VC Type             : ether
Dot1Q Ethertype     : 0x8100               QinQ Ethertype     : 0x8100
Control Word        : Not Preferred
Entropy Label       : Disabled

Admin Ingress label : 524281               Admin Egress label : 524282
Oper Flags          : stitchingSvcTxDown
```

At service level, the first of the following two outputs shows the state of the SAP bound to pw-port 1. As shown, the operational state is down with an indication that this is due to the port being operationally down. The second output shows that this SAP status is propagated to IP interface level because the interface "to-CE-4" is also shown as operationally down.

```
*A:PE-5# show service id 12 sap pw-1:100.1024 detail | match expression
                                                          "Admin State|Flags"
Admin State        : Up                    Oper State        : Down
Flags              : PortOperDown
```

```
*A:PE-5# show router 12 interface "to-CE-4"

===============================================================================
Interface Table (Service: 12)
===============================================================================
Interface-Name                   Adm       Opr(v4/v6)  Mode    Port/SapId
   IP-Address                                                  PfxState
-------------------------------------------------------------------------------
to-CE-4                          Up        Down/Down   VPRN    pw-1:100.1024
   192.168.45.2/30                                             n/a
-------------------------------------------------------------------------------
Interfaces : 1
```

```
===============================================================================
```

To verify a failover, the state of the active/standby pseudowire is transitioned by failing the active pseudowire between PE-3 and PE-2. This causes PE-3 to declare the pseudowire to PE-5 active, which clears the standby status bits. This action causes the SDP (17281) bound to pw-port 1 to become operationally up, followed by pw-port 1 and its associated SAPs, followed by the VPRN IP interface "to-CE-4".

```
164 2021/05/18 09:43:51.723 UTC MINOR: SVCMGR #2103 Base
"Status of service 2147483649 (customer 1) changed to administrative state: up, operational
 state: up"

165 2021/05/18 09:43:51.723 UTC MINOR: SVCMGR #2313 Base
"Status of SDP Bind 53:13 in service 13 (customer 1) peer PW status bits changed to none"

166 2021/05/18 09:43:51.724 UTC MAJOR: SVCMGR #2210 Base
"Processing of an access port state change event is finished and the status of all affected
 SAPs on port pw-1 has been updated."

167 2021/05/18 09:43:51.724 UTC WARNING: SNMP #2005 vprn12 to-CE-4
"Interface to-CE-4 is operational"
```

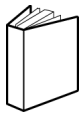
This example in the AS mode section illustrated how notification of a downstream failure is propagated through the components of the PXC in AS mode and reflected in the status of the pw-port (and its associated services). Also, if a pw-port fails due to a PXC failure (for example, the physical port fails), it is just as important that the operational state is propagated externally. In the case of pseudowire backhaul (as in the example), this would be achieved by setting the LDP pseudowire status bits to psnIngressFault and psnEgressFault toward the far end.

Conclusion

This chapter demonstrates the principles of PXC configuration. The PXC can be used to provide a relatively simple back-to-back cross-connect operation in DVSM mode, or it can be used in AS mode to provide an integrated path through the FPE with automated cross-connects used to simplify the provisioning process. In both DVSM mode and AS mode, the PXC can be configured as redundant or non-redundant. A relatively simple use-case of terminating an Epipe into a VPRN has been demonstrated for both modes.

There are a large number of use-cases where frame/packet preprocessing is required before service termination. The workaround for these use-cases has previously been a physical external loop, but can now be resolved logically and internally through use of the PXC.

Customer document and product support



Customer documentation

Customer documentation welcome page

Technical support

Product support portal

Documentation feedback

Customer documentation feedback