



7450 ETHERNET SERVICE SWITCH
7750 SERVICE ROUTER
7950 EXTENSIBLE ROUTING SYSTEM
VIRTUALIZED SERVICE ROUTER

INTERFACE CONFIGURATION GUIDE

RELEASE 22.10.R1

3HE 18386 AAAD TQZZA 01

Issue 01

October 2022

© 2022 Nokia.

Use subject to Terms available at: www.nokia.com/terms/.

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.

This document includes Nokia proprietary and confidential information, which may not be distributed or disclosed to any third parties without the prior written consent of Nokia.

This document is intended for use by Nokia's customers ("You"/"Your") in connection with a product purchased or licensed from any company within Nokia Group of Companies. Use this document as agreed. You agree to notify Nokia of any errors you may find in this document; however, should you elect to use this document for any purpose(s) for which it is not intended, You understand and warrant that any determinations You may make or actions You may take will be based upon Your independent judgment and analysis of the content of this document.

Nokia reserves the right to make changes to this document without notice. At all times, the controlling version is the one available on Nokia's site.

No part of this document may be modified.

NO WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY OF AVAILABILITY, ACCURACY, RELIABILITY, TITLE, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, IS MADE IN RELATION TO THE CONTENT OF THIS DOCUMENT. IN NO EVENT WILL NOKIA BE LIABLE FOR ANY DAMAGES, INCLUDING BUT NOT LIMITED TO SPECIAL, DIRECT, INDIRECT, INCIDENTAL OR CONSEQUENTIAL OR ANY LOSSES, SUCH AS BUT NOT LIMITED TO LOSS OF PROFIT, REVENUE, BUSINESS INTERRUPTION, BUSINESS OPPORTUNITY OR DATA THAT MAY ARISE FROM THE USE OF THIS DOCUMENT OR THE INFORMATION IN IT, EVEN IN THE CASE OF ERRORS IN OR OMISSIONS FROM THIS DOCUMENT OR ITS CONTENT.

Copyright and trademark: Nokia is a registered trademark of Nokia Corporation. Other product names mentioned in this document may be trademarks of their respective owners.

© 2022 Nokia.

Table of contents

1	Getting started	8
1.1	About this guide	8
1.2	Interface configuration process	8
1.3	Conventions	9
1.3.1	Precautionary and information messages	9
1.3.2	Options or substeps in procedures and sequential workflows	10
2	Interfaces	11
2.1	Configuration overview	11
2.1.1	Chassis slots and card slots	11
2.1.2	XIOM modules	12
2.1.3	MDA-a, MDA-aXP, MDA, MDA-e, and MDA-s modules	13
2.1.4	XMAs/C-XMAs	15
2.1.5	Hardware licensing	16
2.1.6	Software license activation	18
2.1.7	Software license records	20
2.1.8	Oversubscribed Ethernet MDAs	21
2.1.8.1	Rate limiting	21
2.1.8.2	Packet classification and scheduling	21
2.2	Digital Diagnostics Monitoring	22
2.2.1	SFPs and XFPs	26
2.2.2	Statistics collection	26
2.3	Ports	27
2.3.1	Port types	27
2.3.2	Port features	29
2.3.2.1	Port State and Operational State	29
2.3.2.2	802.1x network access control	30
2.3.2.3	MACsec	37
2.3.2.4	SONET/SDH port attributes	53
2.3.2.5	SONET/SDH path attributes	53
2.3.2.6	APS	54
2.3.2.7	E-LMI	70
2.3.2.8	LLDP	70

2.3.2.9	Exponential Port Dampening.....	74
2.3.3	Per port aggregate egress queue statistics monitoring.....	76
2.3.4	Forward Error Correction.....	77
2.4	Datapath mapping.....	78
2.5	Port Cross-Connect.....	78
2.5.1	PXC terminology.....	78
2.5.2	Overview.....	79
2.5.3	Port-based PXC.....	80
2.5.4	Internal PXC.....	82
2.5.5	PXC sub-ports.....	84
2.5.6	Bandwidth considerations and QoS.....	86
2.5.6.1	Location selection for internal PXC.....	87
2.5.6.2	QoS.....	88
2.5.6.3	Queue allocation on PXC sub-ports.....	91
2.5.6.4	Pool allocations on PXC ports.....	91
2.5.7	Operational states.....	92
2.5.8	PXC statistics.....	92
2.5.8.1	Statistics on faceplate PXC ports.....	92
2.5.8.2	Statistics collection on internal (MAC-based) PXC.....	93
2.5.8.3	Statistics collection on PXC sub-ports.....	93
2.5.9	PXC LAG.....	93
2.5.10	Basic PXC provisioning.....	94
2.5.11	PXC mirroring and LI.....	98
2.5.12	Multi-chassis redundancy.....	98
2.5.13	Health monitoring on the PXC.....	98
2.5.14	Configuration example.....	99
2.6	FPE.....	101
2.7	LAG.....	105
2.7.1	LACP.....	105
2.7.1.1	LACP multiplexing.....	105
2.7.1.2	LACP tunneling.....	106
2.7.2	Active-standby LAG operation.....	106
2.7.3	LAG on access QoS consideration.....	107
2.7.3.1	Adapt QoS modes.....	108
2.7.3.2	Per-fp-ing-queuing.....	109
2.7.3.3	Per-fp-egr-queuing.....	110

2.7.3.4	Per-fp-sap-instance.....	110
2.7.4	Traffic load balancing options.....	111
2.7.4.1	Per-flow hashing.....	111
2.7.4.2	Adaptive load balancing.....	116
2.7.4.3	LAG port weight speed.....	118
2.7.4.4	LAG port hash-weight.....	119
2.7.4.5	Per-link hashing.....	123
2.7.4.6	Explicit per-link hash using LAG link mapping profiles.....	124
2.7.4.7	Consistent per-service hashing.....	125
2.7.4.8	ESM.....	126
2.7.4.9	IPv6 flow label load balancing.....	130
2.7.5	LAG hold-down timers.....	131
2.7.6	BFD over LAG links.....	132
2.7.7	Multi-chassis LAG.....	132
2.7.7.1	Overview.....	132
2.7.7.2	MC-LAG and SRRP.....	135
2.7.7.3	P2P redundant connection across Layer 2/3 VPN network.....	136
2.7.7.4	DSLAM dual homing in Layer 2/3 TPSDA model.....	137
2.7.8	LAG IGP cost.....	138
2.8	G.8031 protected Ethernet tunnels.....	139
2.9	G.8032 protected Ethernet rings.....	139
2.10	Ethernet port monitoring.....	139
2.11	802.3ah OAM.....	141
2.11.1	OAM events.....	144
2.11.1.1	Link monitoring.....	144
2.11.2	Remote loopback.....	149
2.11.3	802.3ah OAM PDU tunneling for Epipe service.....	149
2.11.3.1	802.3ah grace announcement.....	150
2.12	MTU configuration guidelines.....	154
2.12.1	Default MTU values.....	154
2.12.2	Modifying MTU defaults.....	155
2.12.3	Configuration example.....	155
2.13	Deploying preprovisioned components.....	156
2.14	Setting fabric speed.....	156
2.14.1	7750 SR-7/12/12e and 7450 ESS-7/12.....	156
2.14.2	7950 XRS-20/20e.....	157

2.15	Configuration process overview.....	158
2.16	Configuration notes.....	158
2.17	Configuring physical ports with CLI.....	159
2.17.1	Preprovisioning guidelines.....	159
2.17.1.1	Predefining entities.....	159
2.17.1.2	Preprovisioning a port.....	159
2.17.1.3	Maximizing bandwidth use.....	160
2.17.2	Basic configuration.....	160
2.17.3	Common configuration tasks.....	162
2.17.3.1	Configuring cards and MDAs.....	162
2.17.3.2	Configuring ports.....	163
2.18	Service management tasks.....	184
2.18.1	Modifying or deleting an MDA or XMA.....	184
2.18.2	Modifying a card type.....	185
2.18.3	Deleting a card.....	185
2.18.4	Deleting port parameters.....	185
2.18.5	Soft IOM reset.....	185
2.18.5.1	Soft reset.....	185
2.18.5.2	Deferred MDA reset.....	186
2.19	DWDM provisioning.....	187
2.19.1	Provisioning the DWDM coherent optic.....	187
2.19.2	Provisioning DWDM frequency.....	188
2.19.3	Provisioning DWDM coherent parameters.....	192
3	Standards and protocol support.....	197
3.1	Access Node Control Protocol (ANCP).....	197
3.2	Application Assurance (AA).....	197
3.3	Bidirectional Forwarding Detection (BFD).....	197
3.4	Border Gateway Protocol (BGP).....	197
3.5	Broadband Network Gateway (BNG) Control and User Plane Separation (CUPS).....	199
3.6	Certificate management.....	199
3.7	Circuit emulation.....	200
3.8	Ethernet.....	200
3.9	Ethernet VPN (EVPN).....	201
3.10	gRPC Remote Procedure Calls (gRPC).....	201
3.11	Intermediate System to Intermediate System (IS-IS).....	201

3.12	Internet Protocol (IP) Fast Reroute (FRR).....	203
3.13	Internet Protocol (IP) general.....	203
3.14	Internet Protocol (IP) multicast.....	204
3.15	Internet Protocol (IP) version 4.....	206
3.16	Internet Protocol (IP) version 6.....	206
3.17	Internet Protocol Security (IPsec).....	207
3.18	Label Distribution Protocol (LDP).....	209
3.19	Layer Two Tunneling Protocol (L2TP) Network Server (LNS).....	209
3.20	Multiprotocol Label Switching (MPLS).....	209
3.21	Multiprotocol Label Switching - Transport Profile (MPLS-TP).....	210
3.22	Network Address Translation (NAT).....	210
3.23	Network Configuration Protocol (NETCONF).....	211
3.24	Open Shortest Path First (OSPF).....	211
3.25	OpenFlow.....	212
3.26	Path Computation Element Protocol (PCEP).....	212
3.27	Point-to-Point Protocol (PPP).....	213
3.28	Policy management and credit control.....	213
3.29	Pseudowire (PW).....	213
3.30	Quality of Service (QoS).....	214
3.31	Remote Authentication Dial In User Service (RADIUS).....	214
3.32	Resource Reservation Protocol - Traffic Engineering (RSVP-TE).....	215
3.33	Routing Information Protocol (RIP).....	216
3.34	Segment Routing (SR).....	216
3.35	Simple Network Management Protocol (SNMP).....	217
3.36	Timing.....	219
3.37	Two-Way Active Measurement Protocol (TWAMP).....	220
3.38	Virtual Private LAN Service (VPLS).....	220
3.39	Voice and video.....	220
3.40	Wireless Local Area Network (WLAN) gateway.....	221
3.41	Yet Another Next Generation (YANG).....	221
3.42	Yet Another Next Generation (YANG) OpenConfig Modules.....	221

1 Getting started

1.1 About this guide

This guide describes system concepts and provides configuration examples to provision Input/Output modules (IOMs), XMA Control Modules (XCMs), also referred to as cards, Media Dependent Adapters (MDAs), XRS Media Adapters (XMAs), and ports.



Note: Unless otherwise indicated, this guide uses classic CLI command syntax and configuration examples.

This guide is organized into functional chapters and provides concepts and descriptions of the implementation flow, as well as Command Line Interface (CLI) syntax and command usage.

The topics and commands described in this document apply to the:

- 7450 ESS
- 7750 SR
- 7950 XRS
- Virtualized Service Router

For a list of unsupported features by platform and chassis, see the *SR OS R22.x.Rx Software Release Notes*, part number 3HE 18412 000 x TQZZA.

Command outputs shown in this guide are examples only; actual displays may differ depending on supported functionality and user configuration.



Note:

The SR OS CLI trees and command descriptions can be found in the following guides:

- *7450 ESS, 7750 SR, 7950 XRS, and VSR Classic CLI Command Reference Guide*
- *7450 ESS, 7750 SR, 7950 XRS, and VSR Clear, Monitor, Show, and Tools Command Reference Guide* (for both MD-CLI and Classic CLI)
- *7450 ESS, 7750 SR, 7950 XRS, and VSR MD-CLI Command Reference Guide*



Note:

This guide generically covers Release 22.x.Rx content and may contain some content that will be released in later maintenance loads. See the *SR OS R22.x.Rx Software Release Notes*, part number 3HE 18412 000 x TQZZA for information about features supported in each load of the Release 22.x.Rx software.

1.2 Interface configuration process

[Table 1: Configuration process](#) lists the tasks necessary to configure IOMs and XCMs (also referred to as cards), MDAs and XMAs, and ports.



Note: For consistency across platforms, XMAs are modeled in the SR OS (CLI and SNMP) as MDAs.

Unless specified otherwise:

- the term “card” is used generically to refer to both IOMs and XCMs
- the term “MDA” is used generically to refer to both MDAs and XMA

Table 1: Configuration process

Area	Task	Section
Provisioning	Chassis slots and cards	Chassis slots and card slots
	MDAs	MDA-a, MDA-aXP, MDA, MDA-e, and MDA-s modules
	Ports	Ports
Interface Configuration	MTU Configuration	MTU configuration guidelines
	Configure fabric speed	Setting fabric speed
	Preprovisioning	Preprovisioning guidelines
	Configure cards and MDAs	Configuring cards and MDAs
	Configure ports	Configuring ports
	Service management	Service management tasks

1.3 Conventions

This section describes the general conventions used in this guide.

1.3.1 Precautionary and information messages

The following are information symbols used in the documentation.



DANGER: Danger warns that the described activity or situation may result in serious personal injury or death. An electric shock hazard could exist. Before you begin work on this equipment, be aware of hazards involving electrical circuitry, be familiar with networking environments, and implement accident prevention procedures.



WARNING: Warning indicates that the described activity or situation may, or will, cause equipment damage, serious performance problems, or loss of data.



Caution: Caution indicates that the described activity or situation may reduce your component or system performance.



Note: Note provides additional operational information.



Tip: Tip provides suggestions for use or best practices.

1.3.2 Options or substeps in procedures and sequential workflows

Options in a procedure or a sequential workflow are indicated by a bulleted list. In the following example, at step 1, the user must perform the described action. At step 2, the user must perform one of the listed options to complete the step.

Example: Options in a procedure

1. User must perform this step.
2. This step offers three options. User must perform one option to complete this step.
 - This is one option.
 - This is another option.
 - This is yet another option.

Substeps in a procedure or a sequential workflow are indicated by letters. In the following example, at step 1, the user must perform the described action. At step 2, the user must perform two substeps (a. and b.) to complete the step.

Example: Substeps in a procedure

1. User must perform this step.
2. User must perform all substeps to complete this action.
 - a. This is one substep.
 - b. This is another substep.

2 Interfaces

2.1 Configuration overview



Note:

- This document uses the term “preprovisioning” in the context of preparing or preconfiguring entities such as chassis slots, cards, Media Dependent Adapters (MDAs), ports, and interfaces, before initialization. These entities can be installed while remaining administratively disabled (shutdown). When the entity is in a no shutdown state (administratively enabled), then the entity is considered to be provisioned.
- For consistency across platforms, XRS Media Adapters (XMAs) and Compact XMAs (C-XMAs) are modeled as MDAs.
- Unless specified otherwise:
 - The term “card” is used generically to refer to both Input Output Modules (IOMs) and XCMs.
 - The term “MDA” is used generically to refer to both MDAs and XMAs.

Nokia routers provide the capability to configure chassis slots to accept specific card and MDA types and set the relevant configurations before the equipment is actually installed. The preprovisioning capability allows you to plan your configurations as well as monitor and manage your router hardware inventory. Ports and interfaces can also be preprovisioned. When the functionality is needed, the cards can be inserted into the appropriate chassis slots when required.

2.1.1 Chassis slots and card slots

Depending on the chassis type, the relationship between the chassis slot and card slot varies. Chassis slots represent the physical slots in the chassis where modules can be installed. Card slots represent the reference used in management interfaces when provisioning the modules and then using resources of those modules (for example, port references). See the appropriate platform Installation Guide for more information.

To preprovision a card slot, the card type must be specified. Users can enter card type information for each slot. When a card is installed in a slot and enabled, the system verifies that the installed card type matches the provisioned card type. If the parameters do not match, the card remains offline. A preprovisioned slot can remain empty without conflicting with populated slots.

The general syntax for the configuration of card slots is similar for all platforms, though the number of available slots varies by platform and chassis model. The supported card-types vary by chassis. See the appropriate platform Installation Guide for more information.

The 7950 XRS platforms accept XCMs in card slots. An XCM has two slots, each of which accept an XMA or C-XMA module. The C-XMA modules require a mechanical adapter to fit in an XMA slot.

In the **config** context, use the following CLI commands and syntax examples to provision the chassis slot and XCM:

```
A:XRS20>config# card 1
A:XRS20>config>card# card-type xcm-x20
```

The 7750 SR-2s/7s/14s platforms accept XCMs in card slots. The XCMs of the 7750 SR-2s/7s platforms have a single slot for an XMA or an XIOM module. The XCM of the 7750-14s have two slots for the XMA or XIOM modules.

The 7750 SR-1s platform supports a single XCM in a dedicated card slot. This XCM has a single XMA module. The type of XMA module is fixed based on the variant of 7750 SR-1s chassis. Both the XCM and the XMA must be provisioned.

The 7450 ESS-7/12, and 7750 SR-7/12, and 7750 SR-12e platforms accept either IMMs or IOMs in card slots. IOMs have two slots for pluggable MDAs. The IOM4-e and IOM4-e-B support MDA-e modules. The IOM5-e supports MDA-e-XP modules.

In the **config** context, use the following CLI commands and syntax examples to provision a card slot with an IOM:

```
A:SR12-1>config# card 1
A:SR12-1>config>card# card-type iom3
```

IMMs have integrated MDAs. The provisioning requirements depends on the generation of IMM that you use. See the IMM Installation Guide for more information.

The 7750 SR-a platforms support IOM-a cards in dedicated chassis slots. The 7750 SR-a4 supports one physical IOM-a in slot 3. This IOM-a is represented in the CLI as card 1. The 7750 SR-a8 supports two physical IOM-a cards, one in slot 3, the other in slot 6. These IOM-a cards are represented in the CLI as card 1 and card 2 respectively. The IOM-a does not have pluggable MDA slots. Each IOM-a can be configured to support up to four MDA-a or MDA-aXP modules. IOM-a cards are configured in the same manner as IOMs.

The 7750 SR-e platforms support the IOM-e modules in dedicated slots in the rear of each chassis. The 7750 SR-1e supports one physical IOM-e module. This IOM-e is represented in the CLI as card 1. The 7750 SR-2e supports two physical IOM-e cards. These IOM-e cards are represented in the CLI as card 1 and card 2 respectively. The 7750 SR-3e supports three physical IOM-e cards. These IOM-e cards are represented in the CLI as card 1, card 2, and card 3 respectively. The IOM-e does not have pluggable MDA slots. An IOM-e can be configured to support up to four MDA-e modules. IOM-e cards are configured in the same manner as IOMs.

2.1.2 XIOM modules

XIOM modules are modules that are used in 7750 SR-1s/2s/7s/14s platforms. These can be installed into an XCM instead of installing an XMA module. The XIOMs have two slots that support MDA-s modules (see [MDA-a, MDA-aXP, MDA, MDA-e, and MDA-s modules](#)).

The use of an XIOM introduces an additional index into the reference hierarchy. For example, a 7750 SR-14s with an XCM in card slot 1 can have an XMA in the first slot and an XIOM in the second slot. A possible configuration is shown in the following example:

```
#-----
echo "Card Configuration"
#-----
```

```

card 1
  card-type xcm-14s
  mda 1
    mda-type s36-100gb-qsfp28 level cr1600g
    no shutdown
  exit
  xiom x2
    xiom-type iom-s-3.0t level cr1600g+
    mda 1
      mda-type ms2-400g-qsfpdd+2-100g-qsfp28
      no shutdown
    exit
    mda 2
      mda-type ms16-100g-sfpdd+4-100g-qsfp28
      no shutdown
    exit
  no shutdown
exit

```

On the 7750 SR-1s/2s/7s/14s, the MDA-s modules are supported when an XIOM is installed into a slot within an XCM. Up to two MDA-s can be installed in an XIOM. MDA-s names in CLI start with the letters "ms" (for example, ms16-100g-sfpdd+4-100g-qsfp28).

2.1.3 MDA-a, MDA-aXP, MDA, MDA-e, and MDA-s modules

MDAs are pluggable adapter cards that provide physical interface connectivity. MDAs are available in a variety of interface and density configurations. MDA modules differ by chassis. See the individual chassis guide and the individual MDA installation guides for more information about specific MDAs.

On the 7450 ESS-7/12, 7750 SR-7/12, and 7750 SR-12e, MDAs plug into IOMs. (MDA-e modules plug into the IOM4-e and IOM4-e-B, MDA-e-XP modules plug into the IOM5-e). Up to two MDAs can be provisioned on an IOM.

IMMs are designed with fixed integrated media cards, which may require provisioning, depending on the generation of the IMM.

MDA-a and MDA-aXP modules are used in the 7750 SR-a and the MDA-e and ISA2 modules are used in the 7750 SR-e chassis. Up to four MDAs can be provisioned for each IOM.

In all cases, the card slot and IOM or IMM card-type must be provisioned before an MDA can be provisioned. A preprovisioned MDA slot can remain empty without interfering with services on populated equipment. When an MDA is installed and enabled, the system verifies that the MDA type matches the provisioned type. If the parameters do not match, the MDA remains offline.

On the 7450 ESS-7/12, 7750 SR-7/12, and 7750 SR-12e platforms, MDA names in the CLI start with the letter 'm' (for example, m10-1gb-xp-sfp).

The following example displays the **card**, **card-type**, **mda**, and **mda-type** command usage in the 7750 SR-12:

```

*A:Dut-B# configure card 6
*A:Dut-B>config>card# card-type "iom4-e"
*A:Dut-B>config>card# mda 1
*A:Dut-B>config>card>mda# mda-type me1-100gb-cfp2
*A:Dut-B>config>card>mda# no shutdown
*A:Dut-B>config>card>mda# exit
*A:Dut-B>config>card# mda 2

```

```
*A:Dut-B>config>card>mda# mda-type "me10-10gb-sfp+"
*A:Dut-B>config>card>mda# no shutdown
*A:Dut-B>config>card>mda# exit
*A:Dut-B>config>card# no shutdown
```

The following example displays the configuration:

```
*A:Dut-B# admin display-config
echo "Card Configuration"
#-----
card 6
  card-type iom4-e
  mda 1
    mda-type me1-100gb-cfp2
    no shutdown
  exit
  mda 2
    mda-type me10-10gb-sfp+
    no shutdown
  exit
  no shutdown
exit
#-----
```

The 7750 SR-a4 and 7750 SR-a8 support only MDA-a and MDA-aXP modules, which are identified in the CLI with an "ma" prefix (for example, ma4-10gb-sfp+), or "max" prefix (for example, maxp10-10gb-sfp+). Likewise, the 7750 SR-1e, 7750 SR-2e, and 7750 SR-3e support only MDA-e modules, which are identified in the CLI with an "me" prefix, such as me1-100gb-cfp2.

The following example shows the **card**, **card-type**, **mda**, and **mda-type** command usage in the 7750 SR-1e:

Classic CLI commands

```
A:SR1e>config# card 1
A:SR1e>config>card# card-type iom-e
A:SR1e>config>card# mda 1
A:SR1e>config>card>mda# mda-type me10-10gb-sfp+
A:SR1e>config>card>mda# exit
A:SR1e>config>card# mda 4
A:SR1e>config>card>mda# mda-type me1-100gb-cfp2
A:SR1e>config>card>mda# exit
```

The following example displays the configuration for the 7750 SR-1e:

Classic CLI configuration

```
A:SR1e# admin display-config
. . .
echo "Card Configuration"
#-----
card 1
  card-type iom-e
  mda 1
    mda-type me10-10gb-sfp+
  exit
  mda 4
    mda-type me1-100gb-cfp2
  exit
exit
```

```
-----  
A:SR1e#
```

The following example shows the **card**, **card-type**, **mda**, and **mda-type** command usage in the 7750 SR-3e:

MD-CLI commands

```
(gl)[/]  
A:admin@Dut-J# configure card 1  
(gl)[/configure card 1]  
A:admin@Dut-J# card-type iom-e  
*(gl)[/configure card 1]  
A:admin@Dut-J# mda 1 mda-type isa2-aa  
*(gl)[/configure card 1]  
A:admin@Dut-J# mda 2 mda-type me10-10gb-sfp+  
*(gl)[/configure card 1]  
A:admin@Dut-J# mda 3 mda-type me2-100gb-qsfp28  
*(gl)[/configure card 1]  
A:admin@Dut-J# mda 4 mda-type me40-1gb-csfp
```

The following example displays the configuration for the 7750 SR-3e:

MD-CLI configuration

```
*(gl)[/configure card 1]  
A:admin@Dut-J# info  
card-type iom-e  
mda 1 {  
    mda-type isa2-aa  
}  
mda 2 {  
    mda-type me10-10gb-sfp+  
}  
mda 3 {  
    mda-type me2-100gb-qsfp28  
}  
mda 4 {  
    mda-type me40-1gb-csfp  
}  
fp 1 {  
}
```

2.1.4 XMAs/C-XMAs



Note: For consistency across platforms, XMAs are modeled in the system as MDAs, and unless specified otherwise, the term MDA is used generically in this document to refer to both MDAs and C-XMA/XMAs. When the term XMA is used, it refers to both XMAs and C-XMAs unless specified otherwise.

XMAs are supported on the 7750 SR-1s/2s/7s/14s and 7950 XRS platforms. XMAs plug into XCMs. XCMs must be provisioned before an XMA can be provisioned with a type.

The XMA information must be configured before ports can be configured. After you configure the XCM, use the following CLI commands to provision XMAs.

A maximum of two XMA's can be configured on an XCM. The following example displays the card slot, card type, MDA slot, and MDA type command usage:

```
A:XRS20>config# card 1
A:XRS20>config>card# card-type xcm-x20
A:XRS20>config>card# mda 1
A:XRS20>config>card>mda# mda-type cx2-100g-cfp
A:XRS20>config>card>mda# power-priority-level 130
A:XRS20>config>card>mda# exit
A:XRS20>config>card# mda 2
A:XRS20>config>card>mda# mda-type cx20-10g-sfp
A:XRS20>config>card>mda# power-priority-level 135
A:XRS20>config>card>mda# exit
```

The following example displays the configuration:

```
A:XRS20# admin display-config
. . .
-----
echo "Card Configuration "
#-----
card 1
  card-type xcm-x20
  mda 1
    mda-type cx2-100g-cfp
    power-priority-level 130
  exit
  mda 2
    mda-type cx20-10g-sfp
    power-priority-level 135
  exit
exit
-----
A:XRS20#
```

On the 7950 XRS, the **show card state** output displays an "x" in the name of the XMA and "cx" in the name of a C-XMA:

```
A:Dut-A# show card state
=====
Card State
=====
Slot/  Provisioned Type      Admin Operational  Num  Num  Comments
Id     Equipped Type (if different)  State State  Ports MDA
-----
1      xcm-x20                up    up          2
1/1    cx20-10g-sfp           up    up          20
1/2    cx20-10g-sfp           up    up          20
2      xcm-x20                up    up          2
2/1    cx20-10g-sfp           up    up          20
A      cpm-x20                up    up                    Active
B      cpm-x20                up    up                    Standby
=====
```

2.1.5 Hardware licensing

With the introduction of pay-as-you-grow licensing, level-based hardware assemblies (for example, FP4 and FP5 IOMs and XMA's) now include variants with license levels. These levels define the capacity and

functionality of the assembly. The capacity controls aspects such as the number and types of connectors and breakout options that can be configured, as well as the total connector bandwidth. Licensing also controls the number of user (based on configuration) hardware egress queues and egress policers that are available per forwarding plane. For a more complete description of the levels available for a particular assembly, see the associated installation guide.

The license level must be provisioned for the assembly at the same time as the card type or MDA type is provisioned. Each assembly has a set of levels applicable to that particular IOM or XMA that are defined using mnemonic strings. For example, an assembly may have a level 'cr1200g' which refers to a functional level of 'core routing' and a capacity maximum bandwidth of 1.2 Tb/s. A second example of a license level is 'he2400g+', which refers to a functional level of 'high scale edge routing' and a capacity level of a bandwidth of 2.4 Tb/s but with Intelligent Fan In/Out to a higher bandwidth.

When an assembly is installed in the chassis, the license level encoded into the equipped assembly must match the value provisioned for the assembly. If they do not match, the assembly cannot become active in the chassis. The only exception is that a variant of the assembly with the maximum functional and capacity level is allowed to come up in a slot provisioned as any level; the restrictions in effect are at the provisioned level, but this allows this specific assembly to be used to replace any other level of that assembly if necessary.

The following example shows the provisioning of an XCM with two XMAs. The first XMA is a two complex, 2.4T 24-connector QSFP28 XMA with a license level of er2400g (edge routing, 2.4 Tb/s) and the second XMA is a two complex, 2.4T 6-connector CFP8 XMA with a license level of he1600g (high scale edge routing, 4 connector 1.2Tbps):

```
*A:bkvm20# configure card 4 card-type "xcm2-x20"
*A:bkvm20# configure card 4 mda 1 mda-type "x24-100g-qsfp28" level "er2400g"
*A:bkvm20# configure card 4 mda 2 mda-type "x6-400g-cfp8" level "he1600g"
*A:bkvm20# show mda
=====
MDA Summary
=====
```

Slot	Mda	Provisioned Type Equipped Type (if different)	Admin State	Operational State
4	1	x24-100g-qsfp28:er2400g	up	up
	2	x6-400g-cfp8:he1600g	up	up

```
=====
```

The **show card** and **show mda** output display both the type and level of the assembly and indicates when there is a difference between the provisioned and installed levels. In the following example, the first XMA has a provisioned value matching the installed assembly and the second XMA has a difference in the provisioned and installed assembly.

```
*A:bkvm20# show mda 4/1 detail
=====
MDA 4/1 detail
=====
```

Slot	Mda	Provisioned Type Equipped Type (if different)	Admin State	Operational State
4	1	x24-100g-qsfp28:er2400g	up	up

```
MDA Licensing Data
  Licensed Level      : er2400g
  Description         : 2.4T, 24c, Edge Routing
*A:bkvm20# show mda 4/2 detail
=====
MDA 4/2 detail
=====
```

Slot	Mda	Provisioned Type Equipped Type (if different)	Admin State	Operational State
4	2	x6-400g-cfp8:he2400g x6-400g-cfp8:he1600g	up	provisioned
MDA Licensing Data				
Licensed Level			:	he1600g
Description			:	1.6T, 4c, High Scale Edge Routing

The connector and bandwidth constraints of a card or MDA are viewed using the **show licensing** command.

```
*A:bkvm18>config>card>mda# show licensing 1/1
=====
Connector          MAC  Licensed  Restrictions
-----
1/1/c1             1    Yes       None
1/1/c2             1    Yes       None
1/1/c3             1    Yes       None
1/1/c4             1    Yes       None
1/1/c5             1    No        No Breakout Allowed
1/1/c6             1    No        No Breakout Allowed
1/1/c7             2    Yes       None
1/1/c8             2    Yes       None
1/1/c9             2    Yes       None
1/1/c10            2    Yes       None
1/1/c11            2    No        No Breakout Allowed
1/1/c12            2    No        No Breakout Allowed
1/1/c13            3    Yes       None
...
```

The number of hardware egress user queues and egress user policers (total, allocated, and free), that are dependent on the operational license level of the card, XIOM or MDA containing the FP, are displayed using the **tools dump resource-usage card fp** command.

```
*A:PE# tools dump resource-usage card 1 fp 1
=====
Resource Usage Information for Card Slot #1 FP #1
=====
Total  Allocated  Free
-----
...
Egress User Queues | 131072  384  130688
Egress User Policers | 393215  0  393215
...
=====
```

The **tools dump resource-usage card fp** command also displays the number of hardware egress queues available on the related FP that is dependent on the configured allocation of the percentage of ingress queues.

2.1.6 Software license activation

The pay-as-you-grow licensing of the 7750 SR hardware platforms includes the ability to distribute software based licenses to operational systems in a live network. This is done by the creation of a license key file containing various individual licenses, making this file available to a system, and then activating that license key file on the system. Nokia provides the Configuration License Manager (CLM) tool to assist

in this process. Normally, the CLM would perform all the steps in license distribution to a target system and only the assignment of individual licenses within a system itself need to be managed using the management mechanism for the system itself.

The license key file is a secure distribution mechanism for the licenses. Within the license key file, are one or more license keys. Each license key is assigned for a particular target system identified using a UUID. This UUID is tied to the chassis of the system and therefore the license key can only be validated on the system with that chassis. Each license key is also tied to a specific major software release of SR OS.

When a license key file is available to a target system and is activated it is first validated to ensure it is applicable for the specified target. For the license key of the SR OS hardware platform to be valid, the operator must ensure that the:

- license is for a 7xxx platform
- UUID of the system matches the one encoded in the "UUID-locked" license key
- SR OS software version (the major release number) matches the one encoded in the license key
- license file is not expired

Validation or activation of the license key file results in zero or one license key that is valid for the running software on the target. If there is a valid license key, the records contained in that key are read and made available to the system (see [Software license records](#)).

There may be additional license keys in the file that are for the target system but for a different software release. This is the case when an upgrade of the system is planned. Those license keys shall be considered available however, the feature licenses contained within them are not available for use. These can be seen using the **show system license available-licenses** command.

```
A:bkvm20# show system license available-licenses
=====
Available Licenses
License name   : sr-regress@list.nokia.com
License uuid   : ab516e50-2413-44aa-9f7c-34b4e5b64d19
Machine uuid   : ab516e50-2413-44aa-9f7c-34b4e5b64d19
License desc   : 7xxx Platform
License prod   : 7xxx Platform
License sros   : TiMOS-[BC]-16.0.*
Current date   : FRI NOV 03 15:53:54 UTC 2017
Issue date     : FRI SEP 22 20:55:14 UTC 2017
Start date     : FRI SEP 15 00:00:00 UTC 2017
End date       : THU MAR 15 00:00:00 UTC 2018
-----
License name   : sr-regress@list.nokia.com
License uuid   : ab516e50-2413-44aa-9f7c-34b4e5b64d19
Machine uuid   : ab516e50-2413-44aa-9f7c-34b4e5b64d19
License desc   : 7xxx Platform
License prod   : 7xxx Platform
License sros   : TiMOS-[BC]-17.0.*
Current date   : FRI NOV 03 15:53:54 UTC 2017
Issue date     : FRI SEP 22 20:55:14 UTC 2017
Start date     : FRI SEP 15 00:00:00 UTC 2017
End date       : THU MAR 15 00:00:00 UTC 2018
-----
2 license(s) available.
=====
```

On a system boot, the license key file pointed to by the BOF is activated. If new license key files are activated on a system, the BOF should be updated to point to the new license key file. If there are active licenses in system that are in use and the node reboots and the license key file pointed to by the BOF

either fails to validate or does not contain the in-use license records, the system is considered to be in unlicensed state. In this state, the node reboots every 60 minutes until the records are no longer in-use or a valid license key is provided that includes all the in use license records.

2.1.7 Software license records

The activate license key shall unlock a set of license records for use within the system. In Release 16.0, only Hardware Upgrade license records are distributed in the license keys. These can be used to upgrade the hardware capacity or the hardware functional level of a card or an XMA. These upgrades define a starting level and an upgraded level for the target assembly. Multiple instances of the same upgrade can be assigned to a system in the license key. The list of these records and the number in-use and available can be checked with the **show licensing-entitlements** command.

```
*A:bkvm20# show licensing entitlements

=====
License                               Available      In-Use        State
-----
MDA Upgrades
  cr1200g-cr1600g                      1              1            VALID
  cr1200g-er1200g                      1              0            VALID
  cr1600g-cr2400g                      1              0            VALID
  cr1600g-er1600g                      2              0            VALID
  cr2400g-er2400g                      1              0            VALID
  er1200g-er1600g                      4              2            VALID
  er1200g-he1200g                      1              0            VALID
  er1600g-er2400g                      1              0            VALID
  er1600g-he1600g                      1              0            VALID
  er2400g-he2400g                      1              0            VALID
  he1200g-he1600g                      1              0            VALID
  he1600g-he2400g                      1              0            VALID
=====
*
```

An upgrade can be assigned to a particular card or XMA using the **upgrade** command within the **card** or **mda** context. Multiple upgrades can be assigned to the same card or XMA if needed to gradually augment the base card. In the following example, four upgrades have been applied in sequence to the base level of cr1600g to bring the XMA from CR to ER then ER to HE then to increase capacity first from 1600 Gb/s to 2400 Gb/s and then from 2400 Gb/s to 2400 Gb/s with aggregation to 3600 Gb/s.

```
=====
MDA 1/1 detail
=====
Slot  Mda  Provisioned Type                               Admin  Operational
      Mda  Equipped Type (if different)                   State  State
-----
1     1     s36-100gb-qsfp28:cr1600g                       up     provisioned
      (not equipped)

MDA Licensing Data
Licensed Level      : he2400g+
Description         : 2.4T /w agg to 3.6T, 36p, High Scale Edge
                    Routing
Upgrade 1           : cr1600g-er1600g
Upgrade 2           : er1600g-he1600g
Upgrade 3           : he1600g-he2400g
Upgrade 4           : any2400g-2400g+
```

2.1.8 Oversubscribed Ethernet MDAs

The 7750 SR and 7450 ESS support oversubscribed Ethernet MDAs. These have more bandwidth toward the user than the capacity between the MDA and IOM.

A traffic management function is implemented on the MDA to control the data entering the IOM. This function consists of two parts:

- rate limiting
- packet classification and scheduling

2.1.8.1 Rate limiting

The oversubscribed MDA limits the rate at which traffic can enter the MDA on a per-port basis. If a port exceeds its configured limits then the excess traffic is discarded, and 802.3x flow control frames (pause frames) are generated.

2.1.8.2 Packet classification and scheduling

The classification and scheduling function implemented on the oversubscribed MDA ensures that traffic is correctly prioritized when the bus from the MDA to the IOM is overcommitted. This could occur if the policing parameters configured are such that the sum of the traffic being admitted into the MDA is greater than the capacity between the MDA and the IOM.

The classification function uses the bits set in the DSCP or Dot1p fields of the customer packets to perform classification. It can also identify locally addressed traffic arriving on network ports as Network Control packets. This classification on the oversubscribed MDA uses the following rules:

- If the service QoS policy for the SAP (port or VLAN) uses the default classification policy, all traffic is classified as Best Effort (be).
- If the service QoS policy for the SAP contains a Dot1p classification, the Dot1p field in the customer packets is used for classification on the MDA.
- If the service QoS policy for the SAP contains a DSCP classification, the DSCP field in the customer packets is used for classification on the MDA.
- If a mix of Dot1p and DSCP classification definitions are present in the service QoS policy, then the field used to perform classification is the type used for the highest priority definition. For example, if High Priority 1 is the highest priority definition and it specifies that the DSCP field should be used, then the DSCP field is used for classification on the MDA and the Dot1p field is ignored.
- If the service QoS policy for the SAP specifies IP or MAC filters for forwarding class identification, then traffic is treated as Best Effort. Full MAC or IP classification is not possible on the MDA (but is possible on the IOM).
- The packet is classified into 16 classes. Typically, these are the eight forwarding classes and each packet is assigned one priority per forwarding class. After classification, the packet is offered to the queuing model. This queuing model is limited to three queues each having four thresholds. These thresholds define whether an incoming packet, after classification, is accepted in the queue or not. [Table 2: Typical mapping of classes onto queues/threshold](#) shows typical mapping of classes onto queues and thresholds.

Table 2: Typical mapping of classes onto queues/threshold

Counter	{Queue	Threshold	Traffic class}
0	{2	3	"fc-nc / in-profile"}
1	{2	2	"fc-nc / out-profile"}
2	{2	1	"fc-h1 / in-profile"}
3	{2	0	"fc-h1 / out-profile"}
4	{1	3	"fc-ef / in-profile"}
5	{1	2	"fc-ef / out-profile"}
6	{1	1	"fc-h2 / in-profile"}
7	{1	0	"fc-h2 / out-profile"}
8	{0	3	"fc-l1 / in-profile"}
9	{0	3	"fc-l1 / out-profile"}
10	{0	2	"fc-af / in-profile"}
11	{0	2	"fc-af / out-profile"}
12	{0	1	"fc-l2 / in-profile"}
13	{0	1	"fc-l2 / out-profile"}
14	{0	0	"fc-be / in-profile"}
15	{0	0	"fc-be / out-profile"}

A counter is associated with each mapping. The above is an example and is dependent on the type of classification (such as dscp-exp, dot1p, and so on). When the threshold of a particular class is reached, packets belonging to that class are not accepted in the queue. The packets are dropped and the associated counter is incremented.

The scheduling of the three queues is done in a strict priority, highest priority basis is associated with queue 2. This means that scheduling is done at queue level, not on the class that resulted from the classification. As soon as a packet has been accepted by the queue there is no way to differentiate it from other packets in the same queue (for example, another classification result not exceeding its threshold). All packets queued in the same queue, have the same priority from a scheduling point of view.

2.2 Digital Diagnostics Monitoring

Some Nokia SFPs, XFPs, QSFPs, CFPs and the MSA DWDM transponder have the Digital Diagnostics Monitoring (DDM) capability where the transceiver module maintains information about its working status in device registers including:

- temperature
- supply voltage
- transmit (TX) bias current
- TX output power
- received (RX) optical power

For QSFPs and CFPs, DDM Temperature and Supply voltage is available only at the Module level as shown in [Table 4: DDM alarms and warnings](#).

See the [Statistics collection](#) section for details about the QSFP and CFP example DDM and DDM Lane information.

For the QSFPs and CFPs, the number of lanes is indicated by DDM attribute "Number of Lanes: 4".

Subsequently, each lane threshold and measured values are shown per lane.

If a lane entry is not supported by the specific QSFP or CFP specific model, then it is shown as "-" in the entry.

An example QSFP and CFP lane information is provided below:

```

Transceiver Data
Transceiver Type   : QSFP+
Model Number      : 3HE06485AAAA01 ALU  IPUIBMY3AA
TX Laser Wavelength: 1310 nm                Diag Capable   : yes
Number of Lanes   : 4
Connector Code    : LC                      Vendor OUI      : e4:25:e9
Manufacture date  : 2012/02/02              Media           : Ethernet
Serial Number     : 12050188
Part Number       : DF40GELR411102A
Optical Compliance : 40GBASE-LR4
Link Length support: 10km for SMF
=====
Transceiver Digital Diagnostic Monitoring (DDM)
=====
Value High Alarm  High Warn   Low Warn   Low Alarm
-----
Temperature (C)   +35.6      +75.0      +70.0      +0.0      -5.0
Supply Voltage (V) 3.23       3.60       3.50       3.10      3.00
=====
Transceiver Lane Digital Diagnostic Monitoring (DDM)
=====
High Alarm  High Warn   Low Warn   Low Alarm
-----
Lane Tx Bias Current (mA)      78.0       75.0       25.0       20.0
Lane Rx Optical Pwr (avg dBm)  2.30       2.00       -11.02     -13.01
=====
Lane ID Temp(C)/Alm   Tx Bias(mA)/Alm   Tx Pwr(dBm)/Alm   Rx Pwr(dBm)/Alm
-----
1                      -                 43.5              -              0.42
2                      -                 46.7              -              -0.38
3                      -                 37.3              -              0.55
4                      -                 42.0              -              -0.52
=====

Transceiver Type   : CFP
Model Number      : 3HE04821ABAA01 ALU  IPUIBHJDAA
TX Laser Wavelength: 1294 nm                Diag Capable   : yes
Number of Lanes   : 4
Connector Code    : LC                      Vendor OUI      : 00:90:65
    
```

```

Manufacture date   : 2011/02/11           Media           : Ethernet
Serial Number     : C22CQYR
Part Number       : FTLC1181RDNL-A5
Optical Compliance : 100GBASE-LR4
Link Length support: 10km for SMF
=====
Transceiver Digital Diagnostic Monitoring (DDM)
=====
Value High Alarm  High Warn   Low Warn   Low Alarm
-----
Temperature (C)   +48.2     +70.0     +68.0     +2.0     +0.0
Supply Voltage (V) 3.24      3.46      3.43      3.17     3.13
=====
Transceiver Lane Digital Diagnostic Monitoring (DDM)
=====
High Alarm   High Warn   Low Warn   Low Alarm
-----
Lane Temperature (C) +55.0     +53.0     +27.0     +25.0
Lane Tx Bias Current (mA) 120.0     115.0     35.0      30.0
Lane Tx Output Power (dBm) 4.50      4.00      -3.80     -4.30
Lane Rx Optical Pwr (avg dBm) 4.50      4.00      -13.00    -16.00
=====
Lane ID Temp(C)/Alm   Tx Bias(mA)/Alm   Tx Pwr(dBm)/Alm   Rx Pwr(dBm)/Alm
-----
1      +47.6           59.2              0.30              -10.67
2      +43.1           64.2              0.27              -10.31
3      +47.7           56.2              0.38              -10.58
4      +51.1           60.1              0.46              -10.37
=====

```

The transceiver is programmed with warning and alarm thresholds for low and high conditions that can generate system events. These thresholds are programmed by the transceiver manufacturer.

There are no CLI commands required for DDM operations, however, the **show>port port-id detail** command displays DDM information in the Transceiver Digital Diagnostics Monitoring output section.

DDM information is populated into the router's MIBs, so the DDM data can be retrieved by Network Management using SNMP. Also, RMON threshold monitoring can be configured for the DDM MIB variables to set custom event thresholds if the factory-programmed thresholds are not at the wanted levels.

The following are potential uses of the DDM data:

- **optics degradation monitoring**

With the information returned by the DDM-capable optics module, degradation in optical performance can be monitored and trigger events based on custom or the factory-programmed warning and alarm thresholds.

- **link or router fault isolation**

With the information returned by the DDM-capable optics module, any optical problem affecting a port can be quickly identified or eliminated as the potential problem source.

Supported real-time DDM features are summarized in [Table 3: Real-time DDM information](#).

Table 3: Real-time DDM information

Parameter	User units	SFP/XFP units	SFP	XFP	MSA DWDM
Temperature	Celsius	C	✓	✓	✓

Parameter	User units	SFP/XFP units	SFP	XFP	MSA DWDM
Supply Voltage	Volts	μV	✓	✓	
TX Bias Current	mA	μA	✓	✓	✓
TX Output Power	dBm (converted from mW)	mW	✓	✓	✓
RX Received Optical Power ⁴	dBm (converted from dBm) (Avg Rx Power or OMA)	mW	✓	✓	✓
AUX1	parameter dependent (embedded in transceiver)			✓	
AUX2	parameter dependent (embedded in transceiver)			✓	

The factory-programmed DDM alarms and warnings that are supported are summarized in [Table 4: DDM alarms and warnings](#).

Table 4: DDM alarms and warnings

Parameter	SFP/XFP units	SFP	XFP	Required?	MSA DWDM
Temperature - High Alarm - Low Alarm - High Warning - Low Warning	C	Yes	Yes	Yes	Yes
Supply Voltage - High Alarm - Low Alarm - High Warning - Low Warning	μV	Yes	Yes	Yes	No
TX Bias Current - High Alarm - Low Alarm - High Warning - Low Warning	μA	Yes	Yes	Yes	Yes
TX Output Power	mW	Yes	Yes	Yes	Yes

Parameter	SFP/XFP units	SFP	XFP	Required?	MSA DWDM
- High Alarm - Low Alarm - High Warning - Low Warning					
RX Optical Power - High Alarm - Low Alarm - High Warning - Low Warning	mW	Yes	Yes	Yes	Yes
AUX1 - High Alarm - Low Alarm - High Warning - Low Warning	parameter dependent (embedded in transceiver)	No	Yes	Yes	No
AUX2 - High Alarm - Low Alarm - High Warning - Low Warning	parameter dependent (embedded in transceiver)	No	Yes	Yes	No

2.2.1 SFPs and XFPs

The availability of the DDM real-time information and the warning and alarm status is based on the transceiver. It may or may not indicate that DDM is supported. Although some Nokia SFPs support DDM, Nokia has not required DDM support in releases before Release 6.0. Non-DDM and DDM-supported SFPs are distinguished by a specific value in their EEPROM.

For SFPs that do not indicate DDM support in their EEPROM, DDM data is available although the accuracy of the information has not been validated or verified.

For non-Nokia transceivers, DDM information may be displayed, but Nokia is not responsible for formatting, accuracy, and so on.

2.2.2 Statistics collection

The DDM information and warnings and alarms are collected at one minute intervals, so the minimum resolution for any DDM events when correlating with other system events is one minute.

In the Transceiver Digital Diagnostic Monitoring section of the **show port port-id detail** command output:

- if the present measured value is higher than either or both of the High Alarm and High Warn thresholds, an exclamation mark "!" displays along with the threshold value
- if the present measured value is lower than either or both of the Low Alarm and Low Warn thresholds, an exclamation mark "!" displays along with the threshold value

```

B:SR7-101# show port 2/1/6 detail
.....
=====
Transceiver Digital Diagnostic Monitoring (DDM), Internally Calibrated
=====
      Value High Alarm  High Warn   Low Warn  Low Alarm
-----
Temperature (C)      +33.0+98.0  +88.0     -43.0-45.0
Supply Voltage (V)   3.31 4.12   3.60      3.00 2.80
Tx Bias Current (mA)5.7 60.0   50.00.1  0.0
Tx Output Power (dBm) -5.45 0.00  -2.00     -10.50  -12.50
Rx Optical Power (avg dBm) -0.65-3.00! -4.00!   -19.51  -20.51
=====
    
```

2.3 Ports

2.3.1 Port types

Before a port can be configured, the slot must be provisioned with a card type and MDA type.

Nokia routers support the following port types:

- **Ethernet**

Supported Ethernet port types include:

- Fast Ethernet (10/100BASE-T)
- Gb Ethernet (1GbE, 1000BASE-T)
- 10 Gb Ethernet (10GbE, 10GBASE-X)
- 40 Gb Ethernet (40GbE)
- 100 Gb Ethernet (100GbE)

Router ports must be configured as either access, hybrid, or network. The default is network.

- **access ports**

Access ports are configured for customer facing traffic on which services are configured. If a Service Access Port (SAP) is to be configured on the port or channel, it must be configured as an access port or channel. When a port is configured for access mode, the appropriate encapsulation type must be configured to distinguish the services on the port or channel. After a port has been configured for access mode, one or more services can be configured on the port or channel depending on the encapsulation value.

- **network ports**

Network ports are configured for network-facing traffic. These ports participate in the service provider transport or infrastructure network. Dot1q is supported on network ports.

- **hybrid ports**

Hybrid ports are configured for access and network-facing traffic. While the default mode of an Ethernet port remains network, the mode of a port cannot be changed between the access, network, and hybrid values unless the port is shut down and the configured SAPs or interfaces are deleted. Hybrid ports allow a single port to operate in both access and network modes. The MTU of a port in hybrid mode is the same as in network mode, except for the 10/100 MDA. The default encapsulation for hybrid port mode is dot1q; it also supports QinQ encapsulation on the port level. Null hybrid port mode is not supported. After the port is changed to hybrid, the default MTU of the port is changed to match the value of 9212 bytes currently used in network mode (higher than an access port). This is to ensure that both SAP and network VLANs can be accommodated. The only exception is when the port is a 10/100 Fast Ethernet. In those cases, the MTU in hybrid mode is set to 1522 bytes, which corresponds to the default access MTU with QinQ, which is larger than the network dot1q MTU or access dot1q MTU for this type of Ethernet port. The configuration of all parameters in access and network contexts continues to be done within the port using the same CLI hierarchy as in existing implementation. The difference is that a port configured in mode hybrid allows both ingress and egress contexts to be configured concurrently. An Ethernet port configured in hybrid mode can have two values of encapsulation type: dot1q and QinQ. The NULL value is not supported because a single SAP is allowed, and can be achieved by configuring the port in the access mode, or a single network IP interface is allowed, which can be achieved by configuring the port in network mode. Hybrid mode can be enabled on a LAG port when the port is part of a single chassis LAG configuration. When the port is part of a multi-chassis LAG configuration, it can only be configured to access mode because MC-LAG is not supported on a network port and consequently is not supported on a hybrid port. The same restriction applies to a port that is part of an MC-Ring configuration.

For a hybrid port, the amount of the allocated port buffers in each of ingress and egress is split equally between network and access contexts using the following **config>port>hybrid-buffer-allocation>ing-weight access access-weight network network-weight** [0 to 100] and **config>port>hybrid-buffer-allocation>egr-weight access access-weight network network-weight** commands.

Adapting the terminology in buffer-pools, the port's access active bandwidth and network active bandwidth in each ingress and egress are derived as follows (egress formulas shown only):

- $\text{total-hybrid-port-egress-weights} = \text{access-weight} + \text{network-weight}$
- $\text{hybrid-port-access-egress-factor} = \text{access-weight} / \text{total-hybrid-port-egress-weights}$
- $\text{hybrid-port-network-egress-factor} = \text{network-weight} / \text{total-hybrid-port-egress-weights}$
- $\text{port-access-active-egress-bandwidth} = \text{port-active-egress-bandwidth} \times \text{hybrid-port-access-egress-factor}$
- $\text{port-network-active-egress-bandwidth} = \text{port-active-egress-bandwidth} \times \text{hybrid-port-network-egress-factor}$

- **WAN PHY**

10 G Ethernet ports can be configured in WAN PHY mode (using the **ethernet xgig** config). When configuring the port to be in WAN mode, you can change specific SONET/SDH parameters to reflect the SONET/SDH requirements for this port.

- **SONET-SDH and TDM**

Supported SONET-SDH and TDM port types include:

- DS-1/E-1 channel
- OC3/STM-1

- OC12/STM-4

- **Link Aggregation (LAG)**

LAG can be used to group multiple ports into one logical link. The aggregation of multiple physical links allows for load sharing and offers seamless redundancy. If one of the links fails, traffic is redistributed over the remaining links.

- **APS**

Automatic Protection Switching (APS) is a means to provide redundancy on SONET equipment to guard against linear unidirectional or bidirectional failures. The network elements (NEs) in a SONET/SDH network constantly monitor the health of the network. When a failure is detected, the network proceeds through a coordinated pre-defined sequence of steps to transfer (or switchover) live traffic to the backup facility (called protection facility.) This is done very quickly to minimize lost traffic. Traffic remains on the protection facility until the primary facility (called working facility) fault is cleared, at which time the traffic may optionally be reverted to the working facility.

- **Optical Transport Network (OTN)**

Including OTU2, OTU2e, OTU3, and OTU4. OTU2 encapsulates 10-Gigabit Ethernet WAN and adds FEC (Forward Error Correction). OTU2e encapsulates 10-Gigabit Ethernet LAN and adds FEC (Forward Error Correction). OTU3 encapsulated OC768 and adds FEC. OTU4 encapsulates 100-Gigabit Ethernet and adds FEC.

- **connector**

A QSFP28 (or QSFP-DD) connector that can accept transceiver modules including breakout connectors to multiple physical ports. For example, a QSFP28 connector can support ten 10 Gb Ethernet ports. The connectors themselves cannot be used as ports in other commands, however, the breakout ports can be used as any Ethernet port.

2.3.2 Port features

2.3.2.1 Port State and Operational State

There are two port attributes that are related and similar but have slightly different meanings: Port State and Operational State (or Operational Status).

The following descriptions are based on normal individual ports. Many of the same concepts apply to other objects that are modeled as ports in the router such as APS groups but the show output descriptions for these objects should be consulted for the details.

- **Port State**

- Displayed in port summaries such as **show port** or **show port 1/1**
- `tmnxPortState` in the TIMETRA-PORT-MIB
- Values: None, Ghost, Down (linkDown), Link Up, Up

- **Operational State**

- Displayed in the show output of a specific port such as **show port 2/1/3**
- `tmnxPortOperStatus` in the TIMETRA-PORT-MIB
- Values: Up (inService), Down (outOfService)

The behavior of Port State and Operational State are different for a port with link protocols configured (Eth OAM, Eth CFM or LACP for Ethernet ports, LCP for PPP/POS ports). A port with link protocols configured only transitions to the **Up** Port State when the physical link is up and all the configured protocols are up. A port with no link protocols configured transitions from Down to Link Up and then to Up immediately after the physical link layer is up.

The linkDown and linkUp log events (events 2004 and 2005 in the SNMP application group) are associated with transitions of the port Operational State. Note that these events map to the RFC 2863, *The Interfaces Group MIB*, (which obsoletes RFC 2233, *The Interfaces Group MIB using SMIv2*) linkDown and linkUp traps as mentioned in the SNMPv2-MIB.

An Operational State of **Up** indicates that the port is ready to transmit service traffic (the port is physically up and any configured link protocols are up). The relationship between port Operational State and Port State is shown in [Table 5: Relationship of Port State and Oper State](#).

Table 5: Relationship of Port State and Oper State

	Operational state (Oper State or Oper Status) (as displayed in "show port x/y/z")	
Port State (as displayed in the show port summary)	For ports that have no link layer protocols configured	For ports that have link layer protocols configured (PPP, LACP, 802.3ah EFM, 802.1ag Eth-CFM)
Up	Up	Up
Link Up (indicates the physical link is ready)	Up	Down
Down	Down	Down

2.3.2.2 802.1x network access control

Nokia routers support network access control of client devices (PCs, STBs, and so on) on an Ethernet network using the IEEE. 802.1x standard. 802.1x is known as Extensible Authentication Protocol (EAP) over a LAN network or EAPOL.

2.3.2.2.1 802.1x modes

Nokia routers support port-based network access control for Ethernet ports only. Every Ethernet port can be configured to operate in one of three different operation modes, controlled by the port-control parameter:

- **force-auth**

This mode disables 802.1x authentication and causes the port to transition to the authorized state without requiring any authentication exchange. The port transmits and receives normal traffic without requiring 802.1x-based host authentication. This is the default setting.

- **force-unauth**

This mode causes the port to remain in the unauthorized state, ignoring all attempts by the hosts to authenticate. The switch cannot provide authentication services to the host through the interface.

- **auto**

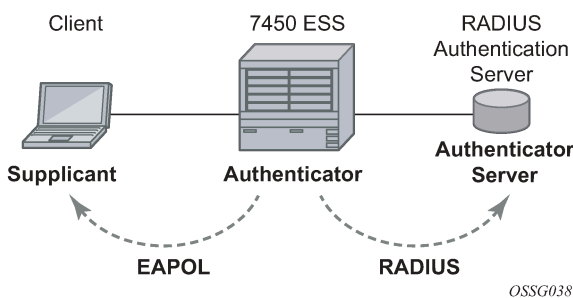
This mode enables 802.1x authentication. The port starts in the unauthorized state, allowing only EAPOL frames to be sent and received through the port. Both the router and the host can initiate an authentication procedure as described below. The port remains in unauthorized state (no traffic except EAPOL frames is allowed) until the first client is authenticated successfully. After this, traffic is allowed on the port for all connected hosts.

2.3.2.2.2 802.1x basics

The IEEE 802.1x standard defines three participants in an authentication conversation (see [Figure 1: 802.1x architecture](#) that shows an example with the 7450 ESS).

the supplicant	the end-user device that requests access to the network
the authenticator	controls access to the network. Both the supplicant and the authenticator are referred to as Port Authentication Entities (PAEs).
the authentication server	performs the actual processing of the user information

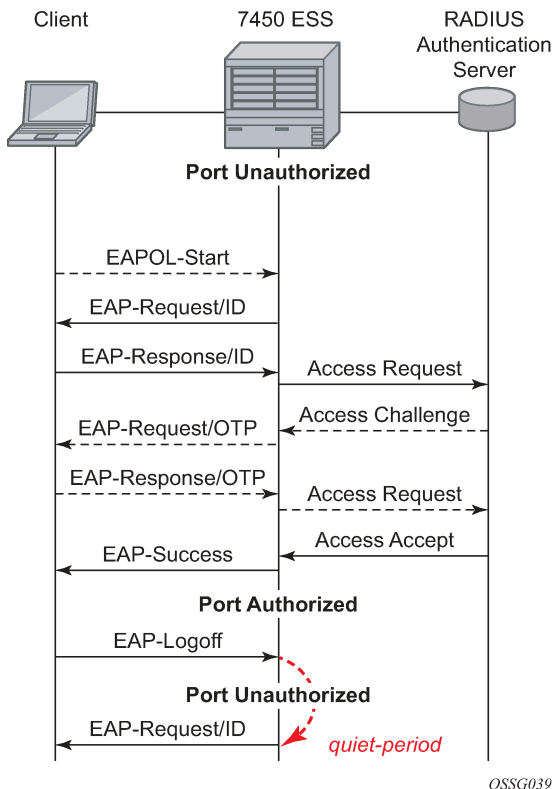
Figure 1: 802.1x architecture



The authentication exchange is carried out between the supplicant and the authentication server, the authenticator acts only as a bridge. The communication between the supplicant and the authenticator is done through the Extended Authentication Protocol (EAP) over LANs (EAPOL). On the back end, the communication between the authenticator and the authentication server is done with the RADIUS protocol. The authenticator is therefore a RADIUS client, and the authentication server a RADIUS server.

The messages involved in the authentication procedure are shown in [Figure 2: 802.1x authentication scenario](#). The router initiates the procedure when the Ethernet port becomes operationally up, by sending a special PDU called EAP-Request/ID to the client. The client can also initiate the exchange by sending an EAPOL-start PDU, if it does not receive the EAP-Request/ID frame during bootup. The client responds on the EAP-Request/ID with a EAP-Response/ID frame, containing its identity (typically username + password).

Figure 2: 802.1x authentication scenario



OSSG039

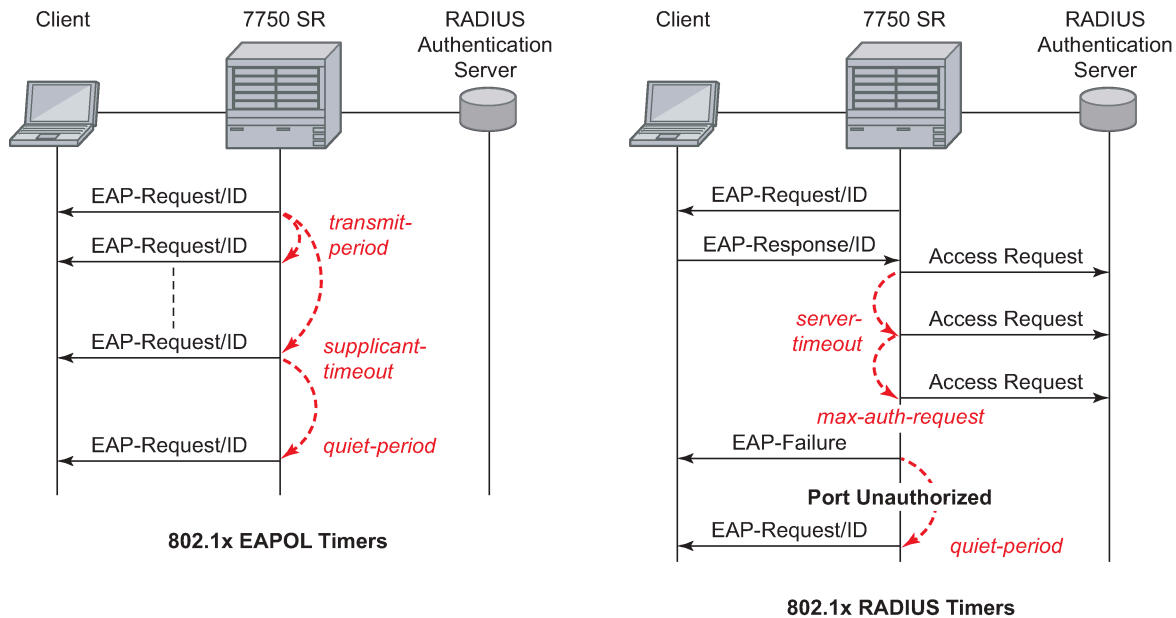
After receiving the EAP-Response/ID frame, the router encapsulates the identity information into a RADIUS AccessRequest packet, and sends it off to the configured RADIUS server.

The RADIUS server checks the supplied credentials, and if approved returns an Access Accept message to the router. The router notifies the client with an EAP-Success PDU and puts the port in authorized state.

2.3.2.2.3 802.1x timers

The 802.1x authentication procedure is controlled by a number of configurable timers and scalars. There are two separate sets, one for the EAPOL message exchange and one for the RADIUS message exchange. See [Figure 3: 802.1x EAPOL timers \(left\) and RADIUS timers \(right\)](#) for an example of the timers on the 7750 SR.

Figure 3: 802.1x EAPOL timers (left) and RADIUS timers (right)



OSSG040-7750

EAPOL timers:

- **transmit-period**

This timer indicates how many seconds the Authenticator listens for an EAP-Response/ID frame. If the timer expires, a new EAP-Request/ID frame is sent and the timer restarted. The default value is 60. The range is 1 to 3600 seconds.

- **supplicant-timeout**

This timer is started at the beginning of a new authentication procedure (transmission of first EAP-Request/ID frame). If the timer expires before an EAP-Response/ID frame is received, the 802.1x authentication session is considered as having failed. The default value is 30. The range is 1 to 300.

- **quiet-period**

This timer indicates number of seconds between authentication sessions. It is started after logout, after sending an EAP-Failure message or after expiry of the *supplicant-timeout* timer. The default value is 60. The range is 1 to 3600.

RADIUS timer and scalar:

- **max-auth-req**

This scalar indicates the maximum number of times that the router sends an authentication request to the RADIUS server before the procedure is considered as having failed. The default value is value 2. The range is 1 to 10.

- **server-timeout**

This timer indicates how many seconds the authenticator waits for a RADIUS response message. If the timer expires, the access request message is sent again, up to *max-auth-req* times. The default value is 60. The range is 1 to 3600 seconds.

The router can also be configured to periodically trigger the authentication procedure automatically. This is controlled by the enable **re-authentication** and **re-auth-period** parameters. Reauth-period indicates the period in seconds (since the last time that the authorization state was confirmed) before a new authentication procedure is started. The range of reauth-period is 1 to 9000 seconds (the default is 3600 seconds, one hour). Note that the port stays in an authorized state during the re-authentication procedure.

2.3.2.2.4 802.1x tunneling

Tunneling of untagged 802.1x frames received on a port is supported for both Epipe and VPLS service using either null or default SAPs (for example 1/1/1:*) when the **config>port>ethernet>dot1x port-control** is set to **force-auth**.

When tunneling is enabled on a port (using the command **configure port port-id ethernet dot1x tunneling**), untagged 802.1x frames are treated like user frames and are switched into Epipe or VPLS services which have a corresponding null SAP or default SAP on that port. In the case of a default SAP, it is possible that other non-default SAPs are also present on the port. Untagged 802.1x frames received on other service types, or on network ports, are dropped.

When tunneling is required, it is expected that it is enabled on all ports into which 802.1x frames are to be received. The configuration of dot1x must be configured consistently across all ports in LAG as this is not enforced by the system.

Note that 802.1x frames are treated like user frames, that is, tunneled, by default when received on a spoke or mesh SDP.

2.3.2.2.5 Per-host authentication

Per-host authentication enables SR OS to authenticate each host individually and allows or disallows the PDUs from this host through the port. Per-host authentication is configurable using the CLI.

When dot1x tunneling is disabled, the port does not allow any PDUs to pass through, with the exception of dot1x packets, which are extracted.

When **per-host-authentication** is configured on the port for dot1x, each host is authenticated individually according to the RADIUS policy and host traffic is allowed or disallowed through the port. After the first successful host authentication, the behavior is the following:

- On downstream (that is, traffic from the network to the host), the port is authorized and allows all traffic to go through.
- On upstream (that is, traffic from the host to the network), the port is authorized, but allows only traffic from the authenticated hosts. When a host is allowed through the port, all of the PDUs for that host are allowed to pass through the port, including untagged or tagged packets. The traffic from any unauthenticated host is disallowed.

For per-host authentication, EAPOL packets are sent to the RADIUS server using the RADIUS protocol. The calling station identifier is the source MAC address of the host and is usually present in the packet. The identifier is used to allow or disallow the host source MAC address based on the RADIUS success or failure answer.

The hosts are authenticated periodically. If a host is authenticated and placed on the allow list and a subsequent authentication fails, that host is removed from the allow list.

If a host authenticates unsuccessfully multiple times, that host is put on a disallow list for a specific amount of time. That is, enabling per-host authentication provides per-host (source MAC) DoS mitigation.

Duplicate MAC addresses are not allowed on the port.

All logs display per-host authentication.

2.3.2.2.5.1 Per-host authentication interaction with dot1x

When per host authentication is first enabled, all MAC addresses on the port are denied. The user can allow MAC addresses using the static source MAC or dot1x host authentication. The following considerations apply when dot1x authentication is used.

- If the 802.1x authentication mode is configured as **force-auth** (using the command **config>port>ethernet>dot1x port-control**), any host that sends EAPOL frames is authenticated without requiring any exchange with the RADIUS server.
- If **config>system>security dot1x** is configured as **shutdown**, the port behavior is the same as in the **force-auth** case.
- If the 802.1x authentication mode is configured as **auto**, the hosts are authenticated using RADIUS. However, if **config>system>security dot1x** is configured as **shutdown**, the **force-auth** behavior takes effect.

2.3.2.2.5.2 Static allow source MAC

A host can be added to the Allow MAC list statically, without being authenticated using dot1x. In this case, the host source MAC address must be added manually using the CLI.

If the same host is added to the list using dot1x and the CLI, the static configuration takes precedence. If the host is added using the CLI, the host is placed on the Allow list. If the same host tries to authenticate using RADIUS and the authentication fails, the host is still allowed through the port because it was statically added using the CLI command **config>port>ethernet>dot1x>per-host-authentication>allowed-source-macs mac-address mac-address**.

2.3.2.2.5.3 Tagged dot1x authentication

dot1x packets can arrive tagged or untagged on the authenticator port from the host. SR OS can be configured to tunnel or extract tagged dot1x packets. SR OS forwards tagged dot1x packets only.

The tunneling or extracting of tagged dot1x packets can be enabled for dot1q (**tunnel-dot1q**) and QinQ (**tunnel-qinq**) encapsulation types.

Each of the encapsulation types configured on the port can be configured to tunnel dot1x packets or extract dot1x packets to be authenticated using a configured RADIUS policy.

The extraction or tunneling of tagged packets applies to any tag value.

2.3.2.2.5.4 Dot1x and LAG

For dot1x authentication support, when the primary port member of the LAG is configured with dot1x, all members inherit the dot1x functionality. Dot1x packets can be extracted on any LAG member and sent to the RADIUS server for processing and authentication. After a successful authentication, the host is allowed on all LAG members. The host dot1x packets can be extracted on one LAG member, while the actual traffic traverses another LAG member. The following is the behavior of dot1x in a LAG bundle:

- When port members are added to the LAG and dot1x is enabled, all ports inherit the same dot1x configuration as the primary port member of the LAG.
- If a host source address (SA) is authenticated through one of the LAG member ports, all ports on the LAG bundle are authorized and pass traffic.
- When a new port member is added to the LAG, if the LAG bundle has been authenticated and is authorized, the new port member is authorized as well.
- Dot1x configuration changes are allowed on the primary LAG member only. A port can be added to a LAG only if its dot1x configuration aligns with that of the primary LAG member. If at least one LAG member is authorized, all LAG members are authorized.

In an upgrade scenario, when an older configuration file (**admin save**) is executed on a new release, a warning is displayed instead of an error for a command that violates the dot1x configuration change behavior; the violating command is ignored.

- If a port is removed from the LAG bundle, the port becomes unauthorized and the EAP negotiation should authorize the port again. This is true for all ports in the LAG bundle, primary or not.
- When Random Early Discard (RED) updates are received during an ISSU on a LAG member in standby, the following updates are ignored:
 - enable dot1x on a LAG member
 - authorize a LAG member

When a port is added to a LAG during ISSU, its dot1x configuration is reset to the default values.

2.3.2.2.5.5 SR host authentication behaviour

SR allows the same MAC source address (MAC SA) on different ports if the MAC address is authenticated. Multiple hosts with the same MAC address can reside and get authenticated on different ports.

2.3.2.2.5.6 Authentication lists

The following authentication lists are supported:

- **authenticated host list**

This list contains up to 1000 hosts. Only hosts that have been authenticated through RADIUS and are allowed through the port are included in this list.

- **unauthenticated host list**

This list contains up to 2000 hosts. Only hosts that have failed authentication or are in the process of being authenticated are included in this list.

If this list reaches the 2000-host limit and a new host is being authenticated, the new host bumps off the list the first host that has failed authentication. The following sequence shows an example:

Unauthenticated list

Host 1 authenticating

Host 2 failed authentication

....

Host 2000 authenticating

Host 2001 just arrived, this host should bump Host 2 off in the list, not Host 1.

If all hosts are in authenticating state, the new Host 2001 is not allowed on the list.

2.3.2.2.6 802.1x configuration and limitations

Configuration of 802.1x network access control on the router consists of two parts:

- generic parameters, which are configured under **config>system>security>dot1x**
- port-specific parameters, which are configured under **config>port>ethernet>dot1x**

The following considerations apply:

- If per-host authentication is not configured, the authentication of any host on the port provides access to the port for any device, even if only a single client has been authenticated.
- 802.1x authentication can only be used to gain access to a pre-defined Service Access Point (SAP). It is not possible to dynamically select a service (such as a VPLS service) depending on the 802.1x authentication information.
- If 802.1x access control is enabled and a high rate of 802.1x frames are received on a port, that port is blocked for a period of 5 minutes as a DoS protection mechanism.

2.3.2.2.6.1 Disabling the 802.1x functionality on a port

By default, the 802.1x functionality consisting of packet extraction and processing on the CPM is enabled on each port.

Use the following commands to administratively disable the 802.1x functionality on a port by not extracting the dot1x packets to the CPM:

- **configure>port>ethernet>dot1x shutdown** (classic CLI)
- **configure port ethernet dot1x admin-state** (MD-CLI)

2.3.2.3 MACsec

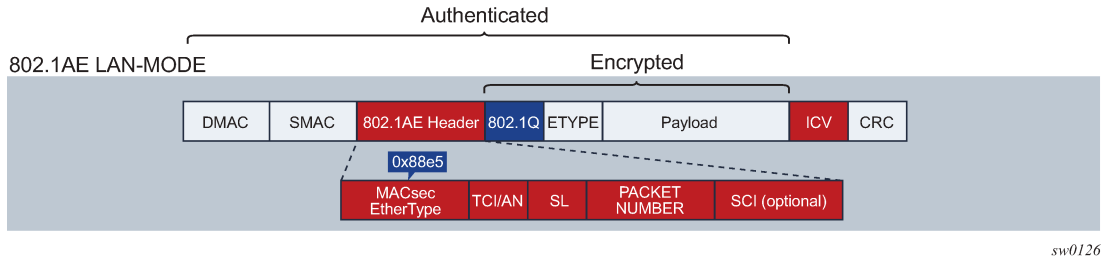
Media Access Control Security (MACsec) is an industry-standard security technology that provides secure communication for almost all types of traffic on Ethernet links. MACsec provides point-to-point and point-to-multipoint security on Ethernet links between directly-connected nodes or nodes connected via a Layer 2 cloud. MACsec can identify and prevent most security threats, including:

- denial of service
- intrusion
- man-in-the-middle
- masquerading
- passive wiretapping
- playback attacks

MACsec Layer 2 encryption is standardized in IEEE 802.1AE. MACsec encrypts anything from the 802.1AE header to the end of the payload including 802.1Q. MACsec leaves the DMAC and SMAC in clear text.

Figure 4: 802.1 AE LAN-MODE shows the 802.1AE LAN-Mode structure.

Figure 4: 802.1 AE LAN-MODE



The forwarding on a MACsec packet is performed using the destination MAC address, which is in clear text.

2.3.2.3.1 MACsec 802.1AE header (SecTAG)

The MACsec 802.1AE header includes a security TAG (SecTAG) field that contains the following:

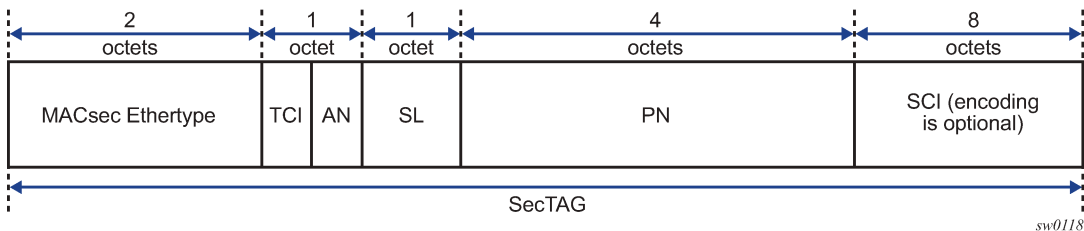
- association number within the channel
- packet number to provide a unique initialization vector for encryption and authentication algorithms as well as protection against replay attack
- optional LAN-Wide secure channel identifier

The security field, which is identified by the MACsec EtherType, conveys the following information:

- TAG Control Information (TCI)
- Association Number (AN)
- Short Length (SL)
- Packet Number (PN)
- Optionally-encoded Secure Channel Identifier (SCI)

Figure 5: SecTAG format shows the format of the SecTAG.

Figure 5: SecTAG format



2.3.2.3.2 MACsec encryption mode

There are two main modes of encryption in MACsec:

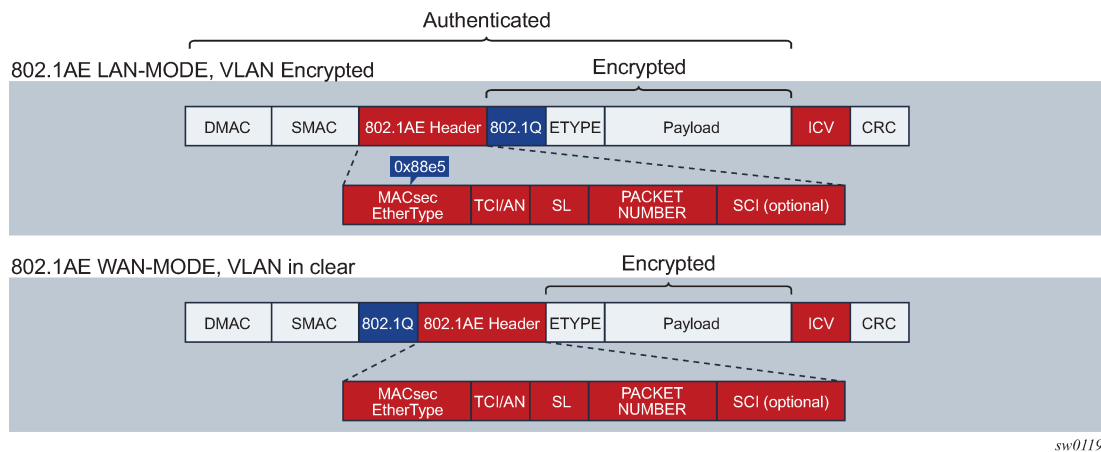
- VLAN in clear text (WAN Mode)
- VLAN encrypted

802.1AE dictates that the 802.1Q VLAN needs to be encrypted. Some vendors give the option of configuring the MACsec on a port with VLAN in clear text.

SR OS supports both modes. On the 7750 SR and 7450 ESS, 1/10 Gig cards support both mode of operation.

[Figure 6: 802.1 AE LAN and WAN modes and VLAN encrypted/clear](#) shows the VLAN encrypted and VLAN in clear.

Figure 6: 802.1 AE LAN and WAN modes and VLAN encrypted/clear



sw0119

2.3.2.3.2.1 MACsec encryption per traffic flow encapsulation matching

In Release 16.0 and later, MACsec can be applied to a selected sub-set of the port traffic, based on the type and value of the packet encapsulation. The SR OS can be configured to match and encrypt the following traffic encapsulation types:

- all encap traffic arriving on port including untagged, single-tag, and double-tag. This is the default behavior of MACsec and the only option supported in releases before 16.0.
- untagged only traffic.
- single-tag or dot1q traffic. In this mode, MACsec can apply to a specific tag or wild card tag where all single-tag traffic is matched.
- double-tag or QinQ traffic. In this mode, MACsec can apply to a specific service tag, a specific service and customer tag, or a wild card for any QinQ traffic.

MKA PDUs are generated specifically for the traffic encapsulation type that is being matched.

2.3.2.3.3 MACsec key management modes

There are four main, key management modes in MACsec. [Table 6: MACsec key management modes](#) describes these management modes.

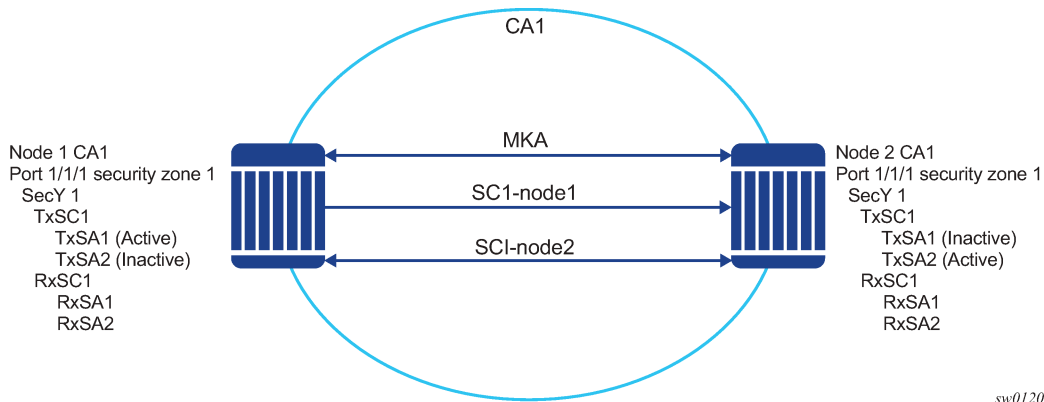
Table 6: MACsec key management modes

Keying	Explanation	SR OS support	Where used
Static SAK	Manually configures each node with a static SAK, SAM, or CLI		Switch to switch
Static CAK PRE SHARED KEY	Uses a dynamic MACsec Key Management (MKA) and uses a configured pre shared key to drive the CAK. The CAK encrypts the SAK between two peers and authenticates the peers	✓	Switch to switch
Dynamic CAK EAP Authentication	Uses a dynamic MKA and an EAP Master System Key (MSK) to drive the CAK. The CAK encrypts the SAK between two peers and authenticates the peers		Switch to switch
Dynamic CAK MSK distribution via RADIUS and EAP-TLS	Stores the MSKs in the Radius server and distributes to the hosts via EAP-TLS. This is typically used in the access networks where a large number of hosts use MACsec and connect to an access switch. MKA uses MSK to drive the CAK. The CAK encrypts the SAK between 2 peers and authenticates the peers		Host to switch

2.3.2.3.4 MACsec terminology

Figure 7: MACsec concepts for static-CAK illustrates some of the main concepts used in MACsec for the static-CAK scenario.

Figure 7: MACsec concepts for static-CAK



sw0120

Table 7: MACsec terms describes the MACsec terminology.

Table 7: MACsec terms

MACsec term	Description
CA: Connectivity Association	A security relationship, established and maintained by key agreement protocols (MKA), that comprises a fully-connected subset of the service access points in stations attached to a single LAN that are to be supported by MACsec.
MKA: MACsec Key Agreement Protocol	Control protocol between MACsec peers, which is used for peer aliveness and encryption key distribution. MACsec Key Agreement is responsible for discovering, authenticating, and authorizing the potential participants in a CA.
SecY: MAC Security Entity	Operates the MAC Security protocol within a system. Manages and identifies the SC and the corresponding active SA.
SC: Security Channel	SC provides a unidirectional point-to-point or point-to-multipoint communication. Each SC contains a succession of SAs and each SC has a different SAK.
SA: Security Association	In the cases of SR OS 2 SA per SC, each with a different SAK, each SC comprises a succession of SAs. Each SA is identified by the SC identifier, concatenated with a two-bit association number. The Secure Association Identifier (SAI), therefore created, allows the receiving SecY to identify the SA, and therefore the SAK used to decrypt and authenticate the received frame. The AN, and therefore the SAI, is only unique for the SAs that can be used or recorded by participating SecYs at any instant. MACsec Key Agreement is responsible for creating and distributing SAKs to each of the SecYs in a CA. This key creation and distribution is independent of the cryptographic operation of each of the SecYs. The decision to replace one SA with its successor is made by the SecY that transmits using the SC, after MACsec Key Agreement has informed it that all the other SecYs are prepared to receive using

MACsec term	Description
	that SA. No notification, other than receipt of a secured frame with a different SAI is sent to the receiver. A SecY must always be capable of storing SAKs for two SAs for each inbound SC, and of swapping from one SA to another without notice. Certain LAN technologies can reorder frames of different priority, so reception of frames on a single SC can use interleaved SA.
SAK: Security Association Key	SAK is the encryption key used to encrypt the datapath of MACsec.

2.3.2.3.5 MACsec static CAK

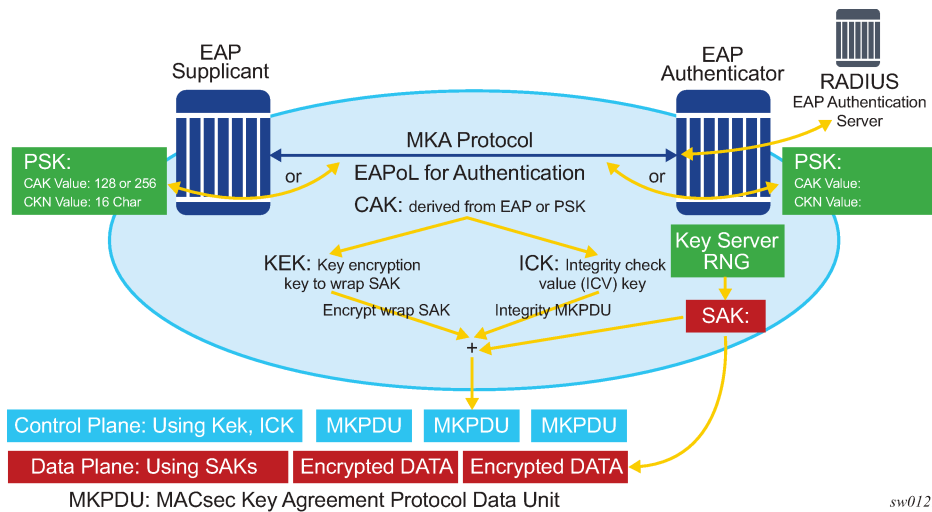
MACsec uses SAs for encryption of packets. SA is a security relationship that provides security guarantees for frames transmitted from one member of a CA to the others. Each SA contains a single secret key (SAK) where the cryptographic operations used to encrypt the datapath PDUs.

SAK is the secret key used by an SA to encrypt the channel.

When enabled, MACsec uses a static CAK security mode. Two security keys, a connectivity association key (CAK) that secures control plane traffic and a randomly-generated secure association key (SAK) that secures data plane traffic are used to secure the point-to-point or point-to-multipoint Ethernet link. Both keys are regularly exchanged between both devices on each end of the Ethernet link to ensure link security.

Figure 8: MACsec generating the CAK illustrates MACsec generating the CAK.

Figure 8: MACsec generating the CAK



The node initially needs to secure the control plane communication to distribute the SAKs between two or more members of a CA domain.

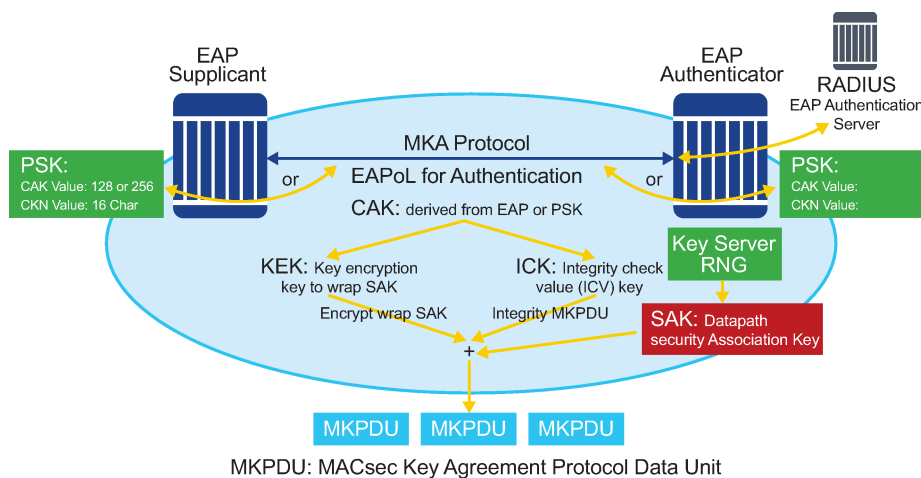
The securing of control plane is done via CAK. To generate the CAK, there are two main methods:

- EAPoL (SR OS does not support EAPoL)
- pre shared key (CAK and CKN values are configured manually via CLI). The following CAK and CKN rules apply.

- CAK is a 32 hexadecimal characters for 128-bit key and 64 hexadecimal characters for 256-bit key depending on which algorithm is used for control plane encryption (for example, **aes-128-cmac** or **aes-256-cmac**).
- CKN is a 32 octets char (64 hex) and it is the connectivity association key name which identifies the CAK. This allows each of the MKA participants to select which CAK to use to process a received MKPDU. MKA places no restriction on the format of the CKN, except that it must comprise an integral number of octets, between 1 and 32 (inclusive), and that all potential members of the CA use the same CKN.
- CKN and CAK must match on peers to create a MACsec Secure CA.

Figure 9: MACsec control plane and encryption illustrates the MACsec control plane authentication and encryption.

Figure 9: MACsec control plane and encryption



sw0122

After the CAK is generated, it can obtain two other keys. These keys are:

- Key Encryption Key (KEK)** used to wrap and encrypt the SAKs
- Integrity Connection Value (ICV) Key (ICK)** used for an integrity check of each MKPDU sent between two CAs

The key server then creates a SAK, that is shared with the CAs of the security domain, and that SAK secures all data traffic traversing the link. The key server continues to periodically create and share a randomly-created SAK over the point-to-point link for as long as MACsec is enabled.

The SAK is encrypted via the AES-CMAC, the KEK as encryption key, and ICK as integration key.

2.3.2.3.6 SAK rollover

SR OS regenerates the SAK after the following events:

- when a new host has joined the CA domain and MKA hellos are received from this host
- when the sliding window is reaching the end of its 32-bit or 64-bit length

- when a new PSK is configured and a rollover of PSK has been executed

2.3.2.3.7 MKA

Each MACsec peer operates the MACsec Key Agreement Protocol (MKA). Each node can operate multiple MKAs based on the number of CA to which the node belongs. Each MKA instance is protected by a distinct secure connectivity Association key (CAK), that allows each PAE to ensure that information for a specific MKA instance is only accepted from other peer that also possess that CAK, and therefore identifying the peers as members or potential members of the same CA. See [MACsec static CAK](#) for information about the CAK identification process done via CKN.

2.3.2.3.7.1 MKA PDU generation

[Table 8: MKA PDU generation](#) describes the MKA PDUs generated for different traffic encapsulation matches.

Table 8: MKA PDU generation

Configuration	Configuration example (<s-tag>.<c-tag>)	MKA packet generation	Traffic pattern match/behavior
All-encap	config>port>ethernet>dot1x>macsec sub-port 10 encap-match all-encap ca-name 10	untagged MKA packet	Matches all traffic on port, including untagged, single-tag, and double-tag. Default behavior; only available behavior in releases before 16.0.
UN-TAG	config>port>ethernet>dot1x>macsec sub-port 1 encap-match untagged ca-name 2	untagged MKA packet	Matches only untagged traffic on port
802.1Q single S-TAG (specific S-TAG)	config>port>ethernet>dot1x>macsec sub-port 2 encap-match dot1q 1 ca-name 3	MKA packet generated with S-TAG=1	Matches only single-tag traffic on port with tagID of 1
802.1Q single S-TAG (any S-TAG)	config>port>ethernet>dot1x>macsec sub-port 3 encap-match dot1q * ca-name 4	untagged MKA packet	Matches any dot1q single-tag traffic on port

Configuration	Configuration example (<s-tag>.<c-tag>)	MKA packet generation	Traffic pattern match/behavior
802.1ad double tag (both tag have specific TAGs)	<code>config>port>ethernet>dot1x>macsec</code> <code>sub-port 4</code> <code>encap-match qinq 1.1</code> <code>ca-name 5</code>	MKA packet generated with S-tag=1 and C-TAG=1	Matches only double-tag traffic on port with service tag of 1 and customer tag of 1
802.1ad double tag (specific S-TAG, any C-TAG)	<code>config>port>ethernet>dot1x>macsec</code> <code>sub-port 6</code> <code>encap-match qinq 1.*</code> <code>ca-name 7</code>	MKA packet generated with S-TAG=1	Matches only double-tag traffic on port with service tag of 1 and customer tag of any
802.1ad double tag (any S-TAG, any C-TAG)	<code>config>port>ethernet>dot1x>macsec</code> <code>sub-port 7</code> <code>encap-match qinq *.*</code> <code>ca-name 8</code>	untagged MKA packet	Matches any double-tag traffic on port

2.3.2.3.7.2 Tags in clear behavior by traffic encapsulation types

[Table 9: Tags in clear behavior](#) describes how single or double tags in clear configuration under a connectivity association affects different traffic flow encryptions.

By default all tags are encrypted in CA. An MKA can be generated without any tags (un-tag) but the data being matched can be based on dot1q or QinQ.

Table 9: Tags in clear behavior

Configuration	Traffic pattern match/behavior	Sub-port's CA configuration: no tag in clear text	Sub-port's CA configuration: single-tag in clear text	Sub-port's CA configuration: double tag in clear text
PORT All-encap	Matches all traffic on port, including untagged, single-tag, double-tag (Release 15.0 default behavior)	MKA PDU: untagged Untagged traffic: encrypted Single-tag traffic: encrypted, no tag in clear Double-tag traffic: encrypted, no tag in clear	MKA PDU: untagged Untagged traffic: in clear Single-tag traffic: encrypted, single-tag in clear Double-tag traffic: encrypted, single-tag in clear	MKA PDU: untagged Untagged traffic: in clear Single-tag traffic: in clear Double-tag traffic: encrypted, double-tag in clear

Configuration	Traffic pattern match/behavior	Sub-port's CA configuration: no tag in clear text	Sub-port's CA configuration: single-tag in clear text	Sub-port's CA configuration: double tag in clear text
untagged	Matches only untagged traffic on port	MKA PDU: untagged Untagged traffic: encrypted Single-tag traffic: not matched by this MACsec policy Double-tag traffic: not matched by this MACsec policy	N/A	N/A
802.1Q single tag (specific tag)	Matches only single-tag traffic on port with the configured tag value	MKA PDU: untagged Untagged traffic: not matched by this MACsec policy Single-tag traffic: tag is encrypted Double-tag traffic: not matched by this MACsec policy	MKA PDU: same tag as the one configured under encap-match Untagged traffic: not matched by this MACsec policy Single-tag traffic: tag is in clear Double-tag traffic: not matched by this MACsec policy	N/A
802.1Q single tag (any tag)	Matches all single-tag traffic on port	MKA PDU: untagged Untagged traffic: not matched by this MACsec policy Single-tag traffic: encrypted Double-tag traffic: not matched by this MACsec policy	MKA PDU: untagged Untagged traffic: not matched by this MACsec policy Single-tag traffic: encrypted with single tag in clear Double-tag traffic: not matched by this MACsec policy	N/A
802.1ad double tag (both tag have specific values)	Matches only double-tag traffic on port with both configured tag values	MKA PDU: untagged Untagged traffic: not matched by this MACsec policy Single-tag traffic: not matched by this MACsec policy	MKA PDU: single tag, equal to S-TAG Untagged traffic: not matched by this MACsec policy Single-tag traffic: not matched by this MACsec policy	MKA PDU: double tag, equal to the values configured under the encap-match Untagged traffic: not matched by this MACsec policy

Configuration	Traffic pattern match/behavior	Sub-port's CA configuration: no tag in clear text	Sub-port's CA configuration: single-tag in clear text	Sub-port's CA configuration: double tag in clear text
		Double-tag traffic matching both configured tags: encrypted, no tag in clear	Double-tag traffic matching both configured tags: single S-TAG in clear	Single-tag traffic: not matched by this MACsec policy Double-tag traffic matching both configured tags: encrypted, both tags in clear
802.1ad double tag (specific S-TAG, any C-TAG)	Matches only double-tag traffic on port with the configured S-TAG	MKA PDU: untagged Untagged traffic: not matched by this MACsec policy Single-tag traffic: not matched by this MACsec policy Double-tag traffic matching the configured S-TAG: encrypted, no tag in clear	MKA PDU: single tag, equal to S-TAG Untagged traffic: not matched by this MACsec policy Single-tag traffic: not matched by this MACsec policy Double-tag traffic matching the configured S-TAG: S-TAG tag in clear	MKA PDU: single tag, equal to S-TAG Untagged traffic: not matched by this MACsec policy Single-tag traffic: not matched by this MACsec policy Double-tag traffic matching the configured S-TAG: both tags in clear
802.1ad double tag (any S-TAG, any C-TAG)	Matches all double-tag traffic on port	MKA PDU: untagged Untagged traffic: not matched by this MACsec policy Single-tag traffic: not matched by this MACsec policy Double-tag traffic: encrypted, no tag in clear	MKA PDU: untagged Untagged traffic: not matched by this MACsec policy Single-tag traffic: not matched by this MACsec policy Double-tag traffic: S-TAG tag in clear	MKA PDU: untagged Untagged traffic: not matched by this MACsec policy Single-tag traffic: not matched by this MACsec policy Double-tag traffic: both tags in clear

2.3.2.3.8 Pre-shared key

A peer may support the use of one or more pre-shared keys (PSKs). An instance of MKA operates for each PSK that is administratively configured as active.

A pre-shared key may be created by NSP, or entered in CLI manually.

Each PSK is configured with two fields. The two fields are:

- CKN (connectivity association name)
- CAK value

The CAK name (CKN) is required to be unique per port among the configured sub-ports, and can be used to identify the key in subsequent management operations.

Each static CAK configuration can have two pre-shared key entries for rollover. The active PSK index dictates the CAK that is used for encrypting the MKA PDUs.

NSP has additional functionality to roll over and configure the PSK. The rollover via NSP can be based on a configured timer.

2.3.2.3.9 MKA Hello timer

MKA uses a member identifier (MI) to identify each node in the CA domain.

A participant proves liveness to each of its peers by including their MI, together with an acceptably-recent message number (MN), in an MKPDU.

To avoid a new participant having to respond to each MKPDU from each partner as it is received, or trying to delay its reply until it is likely that MI MN tuples have been received from all potential partners, each participant maintains and advertises both of the following:

- **live peers list**

This list includes all the peers that have included the participant's MI and a recent MN in a recent MKPDU.

- **potential peers list**

This list includes all the other peers that have transmitted an MKPDU that has been directly received by the participant or that were included in the Live Peers List of a MKPDU transmitted by a peer that has proved liveness.

Peers are removed from each list when an interval of between MKA Life Time and MKA Life Time plus MKA Hello Time has elapsed since the participant's recent MN was transmitted. This time is sufficient to ensure that two or more MKPDUs are lost or delayed before the incorrect removal of a live peer.



Note: The specified use of the live and potential peer lists allows rapid removal of participants that are no longer active or attached to the LAN while reducing the number of MKPDUs transmitted during group formation. For example, a new participant is admitted to an established group after receiving, then transmitting, one MKPDU.

[Table 10: MKA participant timer values](#) describes the MKA participant timer values used on SR OS.

Table 10: MKA participant timer values

Timer use	Timeout (parameter)	Timeout (parameter)
Per participant periodic transmission, initialized on each transmission on expiry	MKA Hello Time or MKA Bounded Hello Time	2.0 0.5
Per peer lifetime, initialized when adding to or refreshing the Potential Peers List or Live	MKA Life Time	6.0

Timer use	Timeout (parameter)	Timeout (parameter)
Peers List, expiry causes removal from the list		
Participant lifetime, initialized when participant created or following receipt of an MKPDU, expiry causes participant to be deleted		
Delay after last distributing a SAK, before the Key Server distributes a fresh SAK following a change in the Live Peer List while the Potential Peer List is still not empty		

2.3.2.3.10 MACsec Capability, Desire, and encryption offset

802.1x-2010 had identified two fields in the MKA PDU. Those fields are:

- MACsec Capability
- Desire

MACsec Capability signals whether MACsec is capable of integrity and confidentiality. [Table 11: MACsec basic settings](#) describes the four basic settings for MACsec Capability.

Table 11: MACsec basic settings

Setting	Description
0	MACsec is not implemented
1	Integrity without confidentiality
2	The following are supported: <ul style="list-style-type: none"> • Integrity without confidentiality • Integrity and confidentiality with a confidentiality offset of 0
3 ¹	The following are supported: <ul style="list-style-type: none"> • Integrity without confidentiality • Integrity and confidentiality with a confidentiality offset of 0, 30, or 50

An encryption offset of 0, 30, or 50 starts from the byte after the SecTAG (802.1ae header). Ideally, the encryption offset should be configured for IPv4 (offset 30) and IPv6 (offset 50) to leave the IP header in the clear text. This allows routers and switches to use the IP header for LAG or ECMP hashing.

¹ SR OS supports setting (3): integrity without confidentiality and Integrity and confidentiality with a confidentiality offset of 0, 30, or 50.

2.3.2.3.11 Key server

The participants, in an MKA instance agree on a Key Server, are responsible for the following:

- deciding on the use of MACsec
- cipher suite selection
- SAK generation and distribution
- SA assignment
- identifying the CA when two or more CAs merge

Each participant in an MKA instance uses the Key Server priority (an 8-bit integer) encoded in each MKPDU to agree on the Key Server. Each participant selects the live participant advertising the highest priority as its Key Server whenever the Live Peers List changes, provided that highest priority participant has not selected another as its Key Server or is unwilling to act as the Key Server. If a Key Server cannot be selected, SAKs are not distributed. In the event of a tie for highest priority Key Server, the member with the highest priority SCI is chosen. For consistency with other uses of the SCI's MAC address component as a priority, numerically lower values of the Key Server Priority and SCI are accorded the highest priority.



Note: For SCI, each SC is identified by an SCI that comprises a globally unique MAC address and a Port Identifier, unique within the system that has been allocated that address.

2.3.2.3.12 SA limits and network design

Each MACsec device supports 64 TX-SAs and 64 RX-SAs. An SA (Security Association) is the key to encrypt or decrypt the data.

In accordance with the IEEE 802.1AE standard, each SecY contains a SC. An SC is a unidirectional concept; for example, Rx-SC or Tx-SC. Each SC contains at least one SA for encryption on Tx-SC and decryption on Rx-SC. Also, for extra security, each SC should be able to roll over the SA. Nokia recommends that each SC should have two SAs for rollover purposes.

Each MACsec phy, referred to as a MACsec security zone, supports 64Tx-SAs and 64 RX-SAs. Assuming two SAs per SC for SA rollover, then each security zone supports 32 RX-SC and 32 TX-SC.

[Table 12: Port mapping to security zone](#) describes the port mapping to security zones.

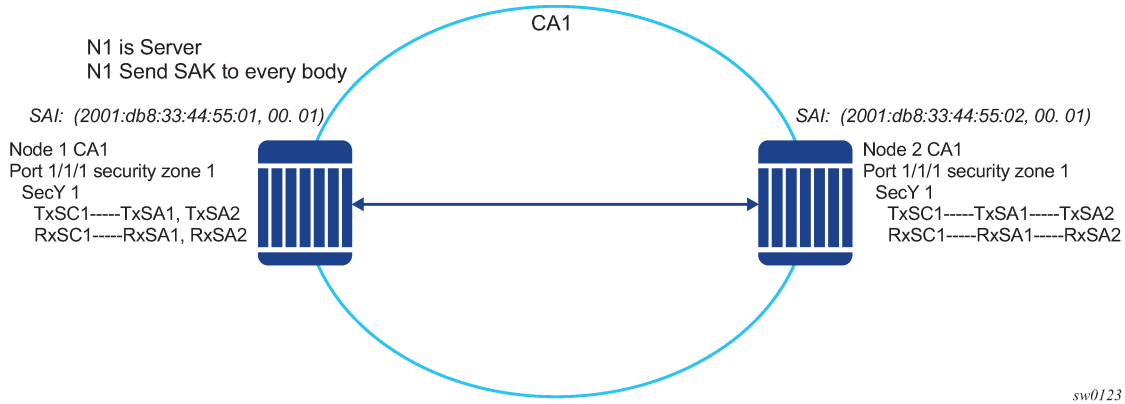
Table 12: Port mapping to security zone

MDA	Ports in security zone 1	Ports in security zone 2	Ports in security zone 3	SA limit per security zone
12-port SFP+/SFP MDA-e	Ports 1, 2, 3, 4	Ports 5, 6, 7, 8	Ports 9, 10, 11, 12	Rx-SA = 64 Tx-SA = 64

2.3.2.3.13 P2P (switch to switch) topology

In a point-to-point topology, each router needs a single security zone and single Tx-SC for encryption and a single Rx-SC for decryption. Each SC has two SAs. In total for point-to-point topology, four SAs are needed, two RxSA for RxSC1 and two TXSA for TxSC1. See [Figure 10: Switch point to switch point topology](#).

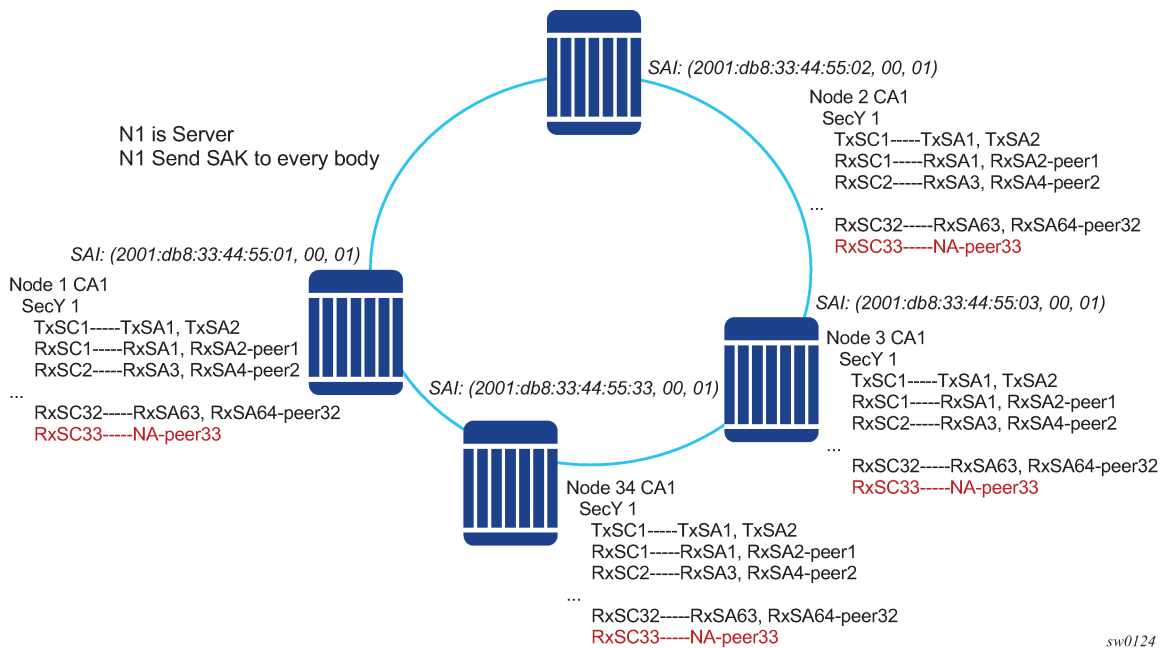
Figure 10: Switch point to switch point topology



2.3.2.3.14 P2MP (switch to switch) topology

In a multipoint topology with N nodes, each node needs a single TxSC and N RxSC, one for each one of the peers. As such, 64 max RX-SA per security zone translates to 32 Rx-SCs, which breaks down to only 32 peers (for example, only 33 nodes in the multipoint topology per security zone, from each node perspective there is one TxSC and 32 RxSC).

Figure 11: Switch multi-point to switch multi-point topology



In Figure 11: Switch multi-point to switch multi-point topology, when the 34th node joins the multi-point topology, all other 33 nodes that are already part of this domain do not have any SAs to create an RxSC for this 34th node. However, the 34th node has a TxSC and accepts 32 peers. The 34th node starts to transmit and encrypt the PDUs based on its TxSC. However, because all other nodes do not have a SC for this SAI, they drop all Rx PDUs.

It is recommended to ensure that a multicast domain, for a single security zone, does not exceed 32 peers or the summation of all the nodes, in a security zone's CA domain, do not exceed 33. This is the same if a security zone has four CAs, the summation of all nodes in the four CAs should be 33 or less.

2.3.2.3.15 SA exhaustion behavior

In [SA limits and network design](#), it was described that a security zone has 64 RxSAs and 64 TxSAs. Two RxSAs are used for each RxSC for rollover purposes and two TxSAs are used for TxSC for rollover purposes. This translates to 32 peers per security zone.

Under each port, a max-peer parameter can be configured. This parameter assigns the number of peers allowed on that port.



Caution: Nokia strongly recommends that the operator ensures the maximum peer value does not exceed the limit of maximum peers per security zone or maximum peers per port values (for example, exceed the **port max-peer** parameter). If the maximum peer is exceeded, the peer connectivity may be random in case of a node failure or packet loss. Peers may join the CA randomly, on a first-come first-served basis.

2.3.2.3.16 Clear tag mode

In most Layer 2 networks, MAC forwarding is done via destination MAC address. The 802.1AE standard dictates that any field after source and destination MAC address and after the SecTAG is required to be encrypted. This includes the 802.1Q tags. In some VLAN switching networks, it may be needed to leave the 802.1Q tag in clear text.

SR OS supports the configuration of 802.1Q tag, in clear text, by placing the 802.1Q tag before the SecTAG or encrypted, by placing it after the SecTAG.

[Table 13: MACsec encryption of 802.1Q tags with clear-tag configured](#) lists the MACsec encryption of 802.1Q tags when the **clear-tag** is configured on SR OS.

Table 13: MACsec encryption of 802.1Q tags with clear-tag configured

Unencrypted format	Clear-tag-mode configuration	Pre-encryption (Tx)	Pre-decryption (Rx)
Single tag (dot1q)	Single-tag	DA, SA, TPID, VID, Etype	DA, SA, TPID, VID, SecTAG
Single tag (dot1q)	Double-tag	DA, SA, TPID, VID, Etype	DA, SA, TPID, VID, SecTAG
Double tag (q-in-q)	Single-tag	DA, SA, TPID1, VID1, IPID2, VID2, Etype	DA, SA, TPID1, VID1, SecTAG
Double tag (q-in-q)	Double-tag	DA, SA, TPID1, VID1, IPID2, VID2, Etype	DA, SA, TPID1, VID1, IPID2, VID2, SecTAG

2.3.2.3.17 802.1X tunneling and multihop MACsec

MACsec is an Ethernet packet and, as with any other Ethernet packet, can be forwarded through multiple switches via Layer 2 forwarding. The encryption and decryption of the packets is performed via the 802.1x (MKA) capable ports.

To ensure that MKA is not terminated on any intermediate switch or router, the user can enable 802.1x tunneling on the corresponding port.

An example check to see if tunneling is enabled, is provided below.

```
*A:SwSim28>config>port>ethernet>dot1x# info
-----
tunneling
```

By enabling tunneling, the 802.1X MKA packets transit the port, without being terminated, therefore MKA negotiation does not occur on a port that has 802.1X tunneling enabled.

2.3.2.3.18 EAPoL destination address

The MKA packets are transported over EAPoL with a multicast destination MAC address. At some point, it may be needed to have the MKA have a point-to-point connection to a peer node over a Layer 2 multihop cloud. In this case, the EAPoL destination MAC address can be set to the peer MAC address. This forces the MKA to traverse multiple nodes and establish an MKA session with the specific peer.

2.3.2.3.19 Mirroring consideration

Mirroring is performed before the MACsec encryption engine. Therefore, if a port is MACsec-enabled and also, that port is mirrored, all the mirrored packets are in cleartext.

2.3.2.4 SONET/SDH port attributes

One OC-3/STM-1 ports are supported on the TDM satellite. The ports can be configured for either SONET or SDH operation. SONET ports are configured for channelized OC-3 operation. SDH ports can be configured for channelized STM-1 operation.

The port's transmit clock rate can be node or loop timed. The port's receive clock rate can be used as a synchronization source for the system. The Section Trace (C1) byte can be configured by the user to ensure correct physical cabling. The port can activate and deactivate local line and internal loopbacks.

All SONET/SDH line alarms are configurable to be either enabled (default) or disabled. Link hold timers can be configured in 100ms increments to control link up and link down indications. The line signal degradation bit error rate (ber-sd) threshold and the line signal failure bit error rate (ber-sf) threshold can be configured.

The TDM satellite support all standard SR OC-3/STM-1 SFP optics including multi-mode, intermediate reach, and long reach. Single fiber mode is not supported.

When an Ethernet port is configured in WAN mode (xgig wan), you can change specific SONET/SDH parameters to reflect the SONET/SDH requirements for this port.

2.3.2.5 SONET/SDH path attributes

Any CES path can only be configured to operate in access mode. Each path has a configurable text description. The SONET/SDH signal label byte (C2) is configurable. The SONET/SDH path trace string (J1) is configurable. Payload scrambling cannot be enabled on CES paths. The valid SONET and SDH path configurations are shown in [Table 14: Valid SONET and SDH path configurations](#).

Table 14: Valid SONET and SDH path configurations

Framing	Path configuration options per physical port	Max number of paths per TDM satellite port
SDH	STM1>AUG1>VC4>TUG3>TUG2>VC12>E1 STM1>AUG1>VC3>TUG2>VC12>E1	63 E1
SONET	OC3>STS1 SPE>VT GROUP>VT1.5 SPE>DS1	84 DS1
SDH	STM1>AUG1>VC4>TUG3>TUG2>TU11>VC11>DS1 STM1>AUG1>VC3>TUG2>VC11>DS1	84 DS1

All SONET/SDH path alarms are configurable to be either enabled (the default) or disabled. The MTU size is configurable per path in the range of 512 to 2092. The path uses a default MTU size set to equal the largest possible CES packet size.

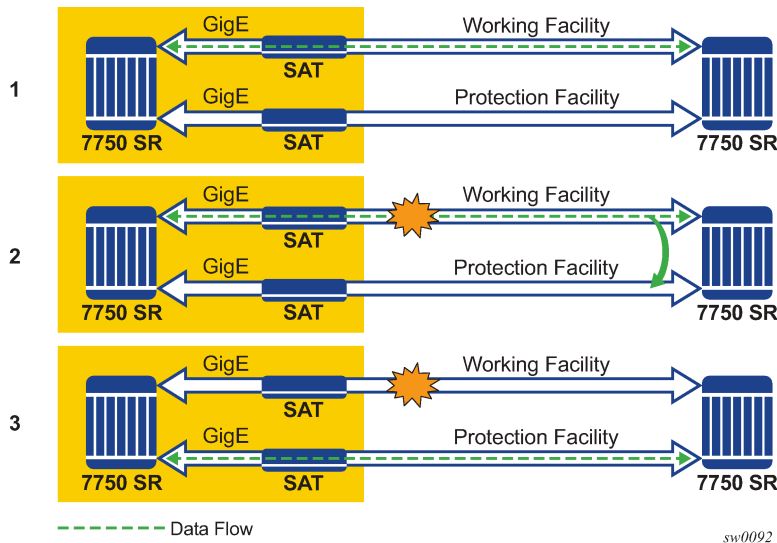
Load balancing options are not applicable to channelized CES paths.

When an Ethernet port is configured in WAN mode (xgig wan), you can change specific SONET/SDH parameters to reflect the SONET/SDH requirements for this port.

2.3.2.6 APS

APS is designed to protect SONET/SDH equipment from linear unidirectional or bidirectional failures. The Network Elements (NEs) in a SONET/SDH network constantly monitor the health of the network. When a failure is detected, the network proceeds through a coordinated pre-defined sequence of steps to transfer (or switchover) live traffic to the backup facility (protection facility). This happens very quickly to minimize lost traffic. Traffic remains on the protection facility until the primary facility (working facility) fault is cleared, at which time the traffic may optionally be reverted to the working facility. An example is shown in [Figure 12: APS protection \(single chassis APS\) and switchover](#).

Figure 12: APS protection (single chassis APS) and switchover



Note that “facility” in the router’s context refers to the physical line (including intermediate transport/switching equipment) and directly attached line terminating hardware (SFP module, MDA and IOM). “Circuit” is also a term used for a link/facility (working-circuit).

A 1+1 APS group contains two circuits.

APS is configured on a port by port basis. If all ports on an MDA or IOM need to be protected then each port on the MDA or IOM must be individually added into an APS group.

Working and protection circuits can be connected to a variety of types of network elements (ADMs, DACSes, ATM switches, routers) and serve as an access or network port providing one or more services or network interfaces to the router. APS-protected SONET/SDH ports may be further channelized. The ports may be one of a variety of encapsulation types as supported by the MDA including PPP, FR and more. For information about MDAs, port types, switching modes, bundles and encapsulations supported with APS, see [APS applicability, restrictions, and interactions](#).

2.3.2.6.1 Single chassis and multi-chassis APS

APS can operate in a single chassis configuration (SC-APS) or in a multi-chassis configuration (MC-APS).

An SC-APS group can span multiple ports, MDAs or IOMs within a single node whereas as MC-APS can span two separate nodes as shown in [Table 15: SC-APS versus MC-APS protection](#).

Table 15: SC-APS versus MC-APS protection

	Single chassis APS	Multi-chassis APS
Short form name	SC-APS	MC-APS
Link failure protection (including intermediate transmission equipment failure)	Yes	Yes

	Single chassis APS	Multi-chassis APS
Optical/electrical module (SPF, XFP) failure protection	Yes	Yes
MDA failure protection	Yes	Yes
IOM failure protection	Yes	Yes
Node failure protection	No	Yes

The support of SC-APS and MC-APS depends on switching modes, MDAs, port types and encaps. For a definitive description of the MDAs, port types, switching modes, bundles and encapsulations supported with APS, see [APS applicability, restrictions, and interactions](#).

2.3.2.6.1.1 APS on a single node (SC-APS)

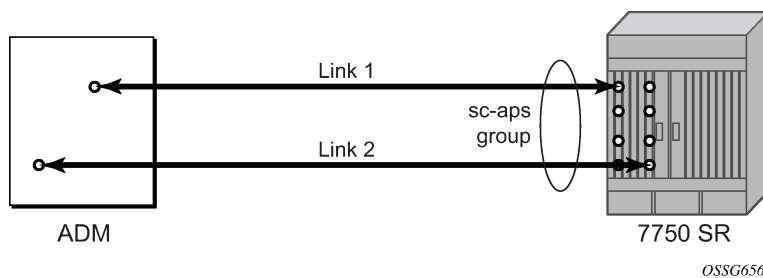
In a single chassis APS both circuits of an APS group are terminated on the same node.

The working and protect lines of a single chassis APS group can be:

- Two ports on the same MDA
- Two ports on different MDAs but on the same IOM
- Two ports on different MDAs on two different IOMs (installed in different slots)
- Two ports on two TDM satellites

If the working and protection circuits are on the same MDA, protection is limited to the physical port and the media connecting the two devices. If the working and protection circuits are on different IOMs then protection extends to MDA or IOM failure. [Figure 13: SC-APS group with MDA and IOM protection](#) shows a configuration that provides protection against circuit, port, MDA, or IOM failure on the 7750 SR connected to an Add-Drop-Multiplexer (ADM).

Figure 13: SC-APS group with MDA and IOM protection



2.3.2.6.1.2 APS across two nodes (MC-APS)

Multi-Chassis APS functionality extends the protection offered by SC-APS to include protection against nodal (7750 SR) failure by configuring the working circuit of an APS group on one 7750 SR node while configuring the protect circuit of the same APS group on a different 7750 SR node.

These two nodes connect to each other with an IP link to establish an MC-APS signaling path between the two 7750 SRs. Note that the working circuit and the protect circuit must have compatible configurations (such as the same speed, framing, and port-type). The relevant APS groups in both the working and protection routers must have same group ID, but they can have different names (for example, group port descriptions). Although the working and protection routers can be different platforms (7750 SR-7 and a 7750 SR-c12), switchover performance may be impacted so it is recommended to avoid a mix of platforms in the same MC-APS group where possible. The configuration consistency between the working circuit/router and the protection circuit/router is not enforced by the 7750 SR. Service or network-specific configuration data is not signaled nor synchronized between the two service routers.

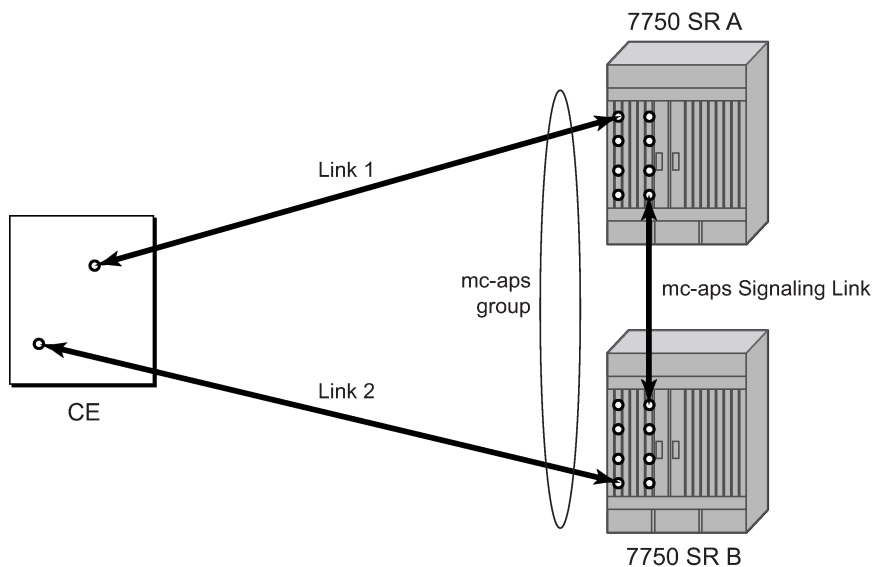
Signaling is provided using the direct connection between the two service routers. A heartbeat protocol can be used to add robustness to the interaction between the two routers. Signaling functionality includes support for:

- APS group matches 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 protect router in sync.
- Protect 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.

Note that external requests like lockout, force, and manual switches are allowed only on the APS group having the protection circuit.

The [Figure 14: MC-APS group protects against node failure](#) shows a Multi-Chassis APS group being used to protect against link, port, MDA, IOM or node failure.

Figure 14: MC-APS group protects against node failure



OSSG657

2.3.2.6.2 APS switching modes

APS behavior and operation differs based on the switching mode configured for the APS group as shown in [Table 16: APS switching modes](#). Several switching modes are supported in the router.

The switching mode affects how the two directions of a link behave during failure scenarios and how APS tx operates.

Unidirectional / Bidirectional configuration must be the same at both sides of the APS group. The APS protocol (K byte messages) exchange switching mode information to ensure that both nodes can detect a configuration mismatch.

- If one end of an APS group is configured in a Unidirectional mode (Uni 1+1 Sig APS or Uni 1+1 Sig +Data APS) then the other end must also be configured in a Unidirectional mode (Uni 1+1 Sig+Data APS).
- If one end of an APS group is configured in a Bidirectional mode then the other end must also be configured in Bidirectional mode.

Table 16: APS switching modes

	Bidirectional 1+1 signaling APS	Unidirectional 1+1 signaling APS	Unidirectional 1+1 signaling and datapath APS
Short form name	Bidir 1+1 Sig APS	Uni 1+1 Sig APS	Uni 1+1 Sig+Data APS
CLI	bidirectional	unidirectional	uni-1plus1
Interworks with a standards compliant APS implementation	Yes	Yes	Yes
Full 1+1 APS standards-based signaling	Yes	Yes	Yes
Data is transmitted simultaneously on both links/ circuits (1+1 Data)	No	No	Yes

The support of switching modes depends on SC-APS/MC-APS, MDAs, port types and encaps. For a definitive description of the MDAs, port types, switching modes, bundles and encapsulations supported with APS, see [APS applicability, restrictions, and interactions](#).

2.3.2.6.2.1 Bidirectional 1+1 signaling APS

In Bidir 1+1 Sig APS switching mode the Tx data is sent on the active link only (it is not bridged to both links simultaneously). 1+1 signaling, however, is used for full interoperability with signaling-compliant 1+1 architectures.

In the ingress direction (Rx), the decision to accept data from either the working or protection circuit is based on both locally detected failures/degradation and on what circuit the far-end is listening on (as indicated in the K bytes). If the far-end indicates that it has switched its active receiver, then the local node also switches its receiver (and Tx) to match the far-end. If the local Rx changes from one circuit to another it notifies the far end using the K bytes.

In the egress direction (Tx), the data is only transmitted on the active circuit. If the active Rx changes, then Tx also changes to the same circuit.

Bidirectional 1+1 Signaling APS ensures that both directions of active data flow (including both Rx) are using the same link/circuit (using the two directions of the same fiber pair) as required by the APS standards. If one end of the APS group changes the active receiver, it signals the far end using the K bytes. The far end then also changes its receiver to listen on the same circuit.

Because the router transmits on active circuits only and keeps active TX and RX on the same port, both local and remote switches are required to restore the service.

The APS channel (bytes K1 and K2 in the SONET header – K bytes) exchanges requests and acknowledgments for protection switch actions. In Bidirectional 1+1 Signaling APS switching mode, the router sends correct status on the K bytes and requires the far-end to also correctly update/send the K-bytes to ensure that data is transmitted on the circuit on which the far-end has selected as its active receiver.

Line alarms are processed and generated independently on each physical circuit.

In Bidirectional 1+1 Signaling APS mode, the highest priority local request is compared to the remote request (received from the far end node using an APS command in the K bytes), and whichever has the greater priority is selected. The relative priority of all events that affect APS 1+1 protection is listed in the [Table 17: K1 byte, bits 1 to 4: type of request](#) in descending order. The requests can be automatically initiated (such as signal failure or signal degrade), external (such as lockout, forced switch, request switch), and state requests (such as revert-time timers, and so on).

2.3.2.6.2.2 Unidirectional 1+1 signaling APS

In Uni 1+1 Sig APS switching mode the Tx data is sent on the active link only (it is not bridged to both links simultaneously). 1+1 signaling, however, is used for full interoperability with signaling-compliant 1+1 architectures.

In the ingress direction (Rx), the decision to accept data from either the working or protection circuit is based on both locally detected failures/degradation and on what circuit the far-end is listening on (as indicated in the K bytes). Although it is not required in the APS standards, the system's implementation of Unidirectional 1+1 Signaling APS uses standards based signaling to keep both the Rx and Tx on the same circuit / port. If the far-end indicates that it has switched its active receiver, then the local node also switches its receiver (and Tx) to match the far-end. If the local Rx changes from one circuit to another it notifies the far end using the K bytes.

In the egress direction (Tx), the data is only transmitted on the active circuit. If the active Rx changes, then Tx also changes to the same circuit.

Because the router transmits on active circuits only and keeps active TX and RX on the same port, both local and remote switches are required to restore the service. For a single failure a data outage is limited to a maximum of 100 milliseconds.

The APS channel (bytes K1 and K2 in the SONET header – K bytes) exchanges requests and acknowledgments for protection switch actions. In Unidirectional 1+1 Signaling APS switching mode, the router sends correct status on the K bytes and requires the far-end to also correctly update/send the K-bytes to ensure that data is transmitted on the circuit on which the far-end has selected as its active receiver.

Line alarms are processed and generated independently on each physical circuit.

In Unidirectional 1+1 Signaling APS switching mode:

- K-bytes are generated/transmitted based on local request/condition only (as required by the APS signaling).
- Local request priority is compliant to 1+1 U-APS specification.
- RX and TX are always forced on to the same (active) circuit (bidirectional). This has the following restrictions:
 - If an APS switch is performed because of a local condition, then the TX direction is moved as well to the newly selected RX circuit (old inactive). The router sends L-AIS on the old active TX circuit to force the remote end to APS switch to the newly active circuit. Note that some local request may not cause an APS switch when a remote condition prevents both RX and TX direction to be on the same circuit (for example an SD detected locally on a working circuit does not cause a switch if the protection circuit is locked out by the remote end).
 - If the remote end indicates an APS switch and the router can RX and TX on the circuit newly selected by the remote end, then the router moves its TX direction and performs an APS switch of its RX direction (unless the router already TX and RX on the newly selected circuit).
 - If the remote end indicates an APS switch and the router cannot RX and TX on the circuit newly selected by the remote end (for example because of a higher priority local request, like a force request or manual request, and so on), then L-AIS are sent on the circuit newly selected by the remote end to force it back to the previously active circuit.
 - The sent L-AIS in the above cases can be either momentary or persistent. The persistent L-AIS is sent under the following conditions:
 - On the protection circuit when the protection circuit is inactive and cannot be selected because of local SF or Lockout Request.
 - On the working circuit as long as the working circuit remains inactive because of a local condition. The persistent L-AIS is sent to prevent revertive switching at the other end.

In all other cases a momentary L-AIS is sent. The system provides debugging information that informs operators about the APS-induced L-AIS.

2.3.2.6.2.3 Unidirectional 1+1 signaling and datapath APS

Uni 1+1 Sig+Data APS supports unidirectional switching operations, 1+1 signaling and 1+1 datapath.

In the ingress direction (Rx) switching is done based on local requests only as per the APS specifications. K-bytes are used to signal the far end the APS actions taken.

In the egress direction (Tx), the data is transmitted on both active and protecting circuits.

Each end of the APS group may be actively listening on a different circuit.

The APS channel (bytes K1 and K2 in the SONET header) exchanges APS protocol messages.

In Uni 1+1 Sig+Data APS a received L-RDI signal on the active circuit does not cause that circuit (port) to be placed out of service. The APS group can continue to use that circuit as the active receiver. This behavior is not configurable.

Uni 1+1 Sig+Data APS also supports configurable:

- Debounce timers for signal failure and degradation conditions
- Suppression of L-RDI alarm generation

2.3.2.6.3 APS channel and SONET header K Bytes

The APS channel (bytes K1 and K2 in the SONET header) exchanges APS protocol messages for all APS modes.

2.3.2.6.3.1 K1 byte

The switch priority of a request is assigned as indicated by bits 1 through 4 of the K1 byte (as described in the rfc3498 APS-MIB); see [Table 17: K1 byte, bits 1 to 4: type of request](#).

Table 17: K1 byte, bits 1 to 4: type of request

Bit 1234	Condition
1111	Lockout of protection
1110	Force switch
1101	SF – High priority
1100	SF – Low priority
1011	SD – High priority
1010	SD – Low priority
1001	(not used)
1000	Manual switch
0111	(not used)
0110	Wait-to-restore
0101	(not used)
0100	Exercise
0011	(not used)
0010	Reverse request
0001	Do not revert
0000	No request

The channel requesting switch action is assigned by bits 5 through 8. When channel number 0 is selected, the condition bits show the received protection channel status. When channel number 1 is selected, the condition bits show the received working channel status. Channel values of 0 and 1 are supported.

[Table 18: K1 byte, bits 5 to 8 \(and K2 bits 1 to 4\), channel number code assignments](#) shows bits 5 to 8 of a K1 byte and K2 Bits 1 to 4 and the channel number code assignments.

Table 18: K1 byte, bits 5 to 8 (and K2 bits 1 to 4), channel number code assignments

Channel number Code	Channel and notes
0	Null channel. SD and SF requests apply to conditions detected on the protection line. For 1+1 systems, Forced and Request Switch requests apply to the protection line (for the 7750 SR only). Only code 0 is used with Lockout of Protection request.
1 to 14	Working channel. Only code 1 applies in a 1+1 architecture. Codes 1 through n apply in a 1:n architecture (for the 7750 SR only). SD and SF conditions apply to the corresponding working lines.
15	Extra traffic channel. May exist only when provisioned in a 1:n architecture. Only No Request is used with code 15.

2.3.2.6.3.2 K2 byte

The K2 byte indicates the bridging actions performed at the line-terminating equipment (LTE), the provisioned architecture and mode of operation.

The bit assignment for the K2 byte is listed in [Table 19: K2 byte functions](#).

Table 19: K2 byte functions

Bits 1 to 8	Function
1 to 4	Channel number. The 7750 SR supports only values of 0 and 1.
5	0 Provisioned for 1+1 mode 1 Provisioned for 1:n mode
6 to 8	111 Line AIS 110 Line RDI 101 Provisioned for bidirectional switching 100 Provisioned for unidirectional switching 011 (reserved for future use) 010 (reserved for future use) 001 (reserved for future use) 000 (reserved for future use)

2.3.2.6.3.3 Differences in SONET and SDH standards for K bytes

SONET and SDH standards are slightly different with respect to the behavior of K1 and K2 Bytes.

Table 20: Differences between SONET and SDH standards shows the differences between the two standards.

Table 20: Differences between SONET and SDH standards

	SONET	SDH	Comments
SONET/SDH standards use different codes in the transmitted K1 byte (bits 1-4) to notify the far-end of a signal fail/signal degrade detection.	1100 for signal fail 1010 for signal degrade 1101 unused 1011 unused	1101 for signal fail 1011 for signal degrade 1100 unused 1010 unused	None
SONET systems signal the switching mode in bits 5-8 of the K2 byte whereas SDH systems do not signal at all.	101 for bi-dir 100 for uni-dir	Not used. 000 is signaled in bits 5 to 8 of K2 byte for both bidirectional as well as unidirectional switching	SONET systems raise a mode mismatch alarm as soon as a mismatch in the TX and RX K2 byte (bits 5 to 8) is detected. SDH systems do not raise the mode mismatch alarm

2.3.2.6.3.4 Failures indicated by K bytes

The following sections describe failures indicated by K bytes.

2.3.2.6.3.4.1 APS protection switching byte failure

An APS Protection Switching Byte (APS-PSB) failure indicates that the received K1 byte is either invalid or inconsistent. An invalid code defect occurs if the same K1 value is received for 3 consecutive frames (depending on the interface type (framer) used, the 7750 SR may not be able to strictly enforce the 3 frame check per GR-253 and G.783/G.841) and it is either an unused code or irrelevant for the specific switching operation. An inconsistent APS byte defect occurs when no three consecutive received K1 bytes of the last 12 frames are the same.

If the failure detected persists for 2.5 seconds, a Protection Switching Byte alarm is raised. When the failure is absent for 10 seconds, the alarm is cleared. This alarm can only be raised by the active port operating in bidirectional mode.

2.3.2.6.3.4.2 APS channel mismatch failure

An APS channel mismatch failure (APS-CM) identifies that there is a channel mismatch between the transmitted K1 and the received K2 bytes. A defect is declared when the received K2 channel number differs from the transmitted K1 channel number for more than 50 ms after three identical K1 bytes are sent. The monitoring for this condition is continuous, not just when the transmitted value of K1 changes.

If the failure detected persists for 2.5 seconds, a channel mismatch failure alarm is raised. When the failure is absent for 10 seconds, the alarm is cleared. This alarm can only be raised by the active port operating in a bidirectional mode.

2.3.2.6.3.4.3 APS mode mismatch failure

An APS mode mismatch failure (APS-MM) can occur for two reasons. The first is if the received K2 byte indicates that 1:N protection switching is being used by the far-end of the OC-N line, while the near end uses 1+1 protection switching. The second is if the received K2 byte indicates that unidirectional mode is being used by the far-end while the near-end uses bidirectional mode.

This defect is detected within 100 ms of receiving a K2 byte that indicates either of these conditions. If the failure detected persists for 2.5 seconds, a mode mismatch failure alarm is raised. However, it continues to monitor the received K2 byte, and should it ever indicate that the far-end has switched to a bidirectional mode the mode mismatch failure clearing process starts. When the failure is absent for 10 seconds, the alarm is cleared, and the configured mode of 1+1 bidirectional is used.

2.3.2.6.3.4.4 APS far-end protection line failure

An APS far-end protection line (APS-FEPL) failure corresponds to the receipt of a K1 byte in 3 consecutive frames that indicates a signal fail (SF) at the far end of the protection line. This forces the received signal to be selected from the working line.

If the failure detected persists for 2.5 seconds, a far-end protection line failure alarm is raised. When the failure is absent for 10 seconds, the alarm is cleared. This alarm can only be raised by the active port operating in a bidirectional mode.

2.3.2.6.4 Revertive switching

The APS implementation also provides the revertive and non-revertive modes with non-revertive switching as the default option. In revertive switching, the activity is switched back to the working port after the working line has recovered from a failure (or the manual switch is cleared). In non-revertive switching, a switch to the protection line is maintained even after the working line has recovered from a failure (or if the manual switch is cleared).

A revert-time is defined for revertive switching so frequent automatic switches as a result of intermittent failures are prevented. A change in this value takes effect upon the next initiation of the wait to restore (WTR) timer. It does not modify the length of a WTR timer that has already been started. The WTR timer of a non-revertive switch can be assumed to be infinite.

In case of failure on both working and the protection line, the line that has less severe errors on the line is active at any point in time. If there is signal degrade on both ports, the active port that failed last stays active. When there is signal failure on both ports, the working port is always active. The reason is that the signal failure on the protection line is of a higher priority than on the working line.

2.3.2.6.5 Bidirectional 1+1 switchover operation example

[Table 21: Actions for the bidirectional protection switching process](#) describes the steps that a bidirectional protection switching process goes through during a typical automatic switchover.

Table 21: Actions for the bidirectional protection switching process

Status	APS commands sent in K1 and K2 bytes on protection line		Action	
	B -> A	A -> B	At site B	At site A
No failure (Protection line is not in use).	No request	No request	No action	No action
Working line Degraded in direction A->B.	SD on working channel 1	No request	Failure detected, notify A and switch to protection line	No action
Site A receives SD failure condition.	Same	Reverse request	No action	Remote failure detected, acknowledge and switch to protection line
Site B receives Reverse request.	Same	Same	No action	No action

2.3.2.6.6 Annex B (1+1 optimized) operation

Operation and behavior conformant with Annex B of ITU.T G.841 can be configured for an APS group. Characteristics of this mode include are the following:

- Annex B operates in non-revertive bidirectional switching mode only as defined in G.841.
- Annex B operates with 1+1 signaling, but 1:1 datapath whereby data is transmitted on the active link only.
- K bytes are transmitted on both circuits.

Because the request/reverse-request nature of an Annex B switchover, the data outage is longer than a typical (non Annex B single chassis) APS switchover. Nokia recommends using maintenance commands under the **tools>perform>aps** context for planned switchovers (not MDA or IOM shutdown) to minimize the outage.

2.3.2.6.6.1 Annex B APS outage reduction optimization

Typical standard Annex B behavior when a local SF is detected on the primary section (circuit), and this SF is the highest priority request on both the local side and from the remote side as per the APS specifications, is to send a request to the remote end and then wait until a reverse request is received before switching over to the secondary section. To reduce the recovery time for traffic, the router switches over to the secondary section immediately upon detecting the local SF on the primary section instead of waiting for the reverse request from the remote side. If the remote request is not received after a period of time then an "PSB Failure is declared" event is raised (Protection Switching Byte Failure – indicates an inconsistent or invalid Rx K1 Bytes), and the APS group on the local side switches back to the primary section.

When the remote side is in Lockout, and a local SF is detected then a reverse request is not received by the local side. In this case, the traffic no longer flows on the APS group because neither the primary nor secondary sections can carry traffic, and the outage reduction optimization causes a temporary switchover from the primary to the secondary and then back again (which causes no additional outage or traffic issue because neither section is usable). If this temporary switchover is not wanted then it is recommended to either perform Lockout from the router side, or to Lockout from both sides, which avoids the possibility of the temporary switchover.

Failures detected on the secondary section cause immediate switch over as per the Annex B specification. There is no outage reduction optimization in the router for this case as it is not needed.

Some examples of events that can cause a local SF to be detected include: a cable being cut, laser transmitter or receiver failure, a port administratively "shutdown", MDA failure or shutdown, IOM failure or shutdown.



Note: In Annex B operation, all switch requests are for a switch from the primary section to the secondary section. After a switch request clears normally, traffic is maintained on the section to which it was switched by making that section the primary section. The primary section may be working circuit 1 or working circuit 2 at any particular moment.

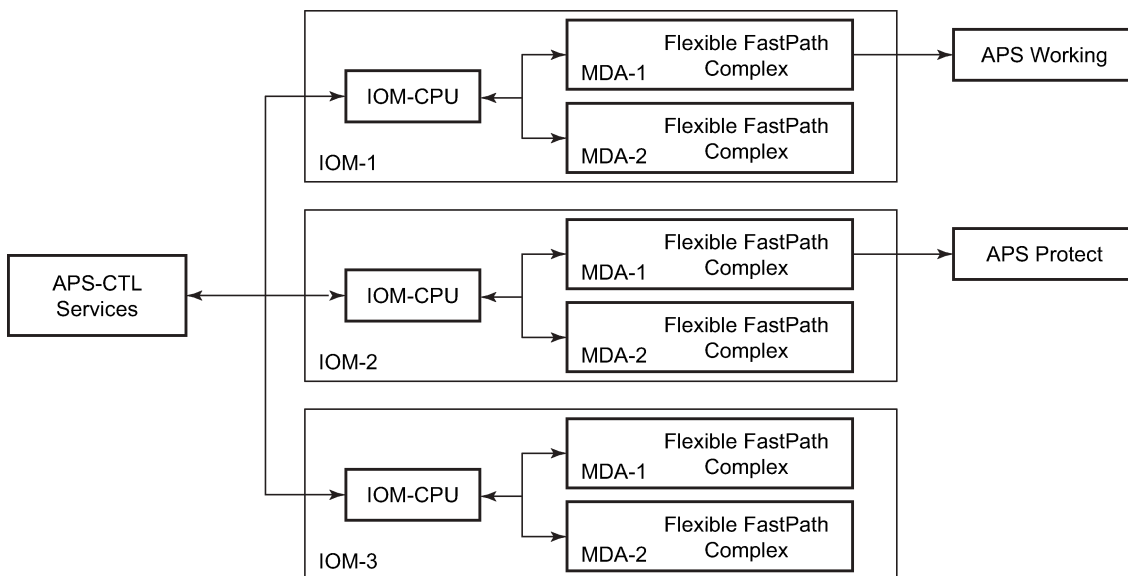
2.3.2.6.7 Protection of upper layer protocols and services

APS prevents upper layer protocols and services from being affected by the failure of the active circuit.

The following example with figures and description illustrate how services are protected during a single-chassis APS switchover.

[Figure 15: APS working and protection circuit example](#) shows an example in which the APS working circuit is connected to IOM-1/MDA-1 and the protection circuit is connected to IOM-2/MDA-1. In this example, assume that the working circuit is currently used to transmit and receive data.

Figure 15: APS working and protection circuit example



Fig_4

2.3.2.6.7.1 Switchover process for transmitted data

For packets arriving on all interfaces that need to be transmitted over APS protected interfaces, the next hop associated with all these interfaces are programmed in all Flexible Fast-Path complexes in each MDA with a logical next-hop index. This next hop-index identifies the actual next-hop information used to direct traffic to the APS working circuit on IOM-1/MDA-1.

All Flexible Fast-Path complexes in each MDA are also programmed with next hop information used to direct traffic to the APS protect circuit on IOM-2/MDA-1. When the transmitted data needs to be switched from the working to the protect circuit, only the relevant next hop indexes need to be changed to the pre-programmed next-hop information for the protect circuit on IOM-2/MDA-1.

Although the control CPM on the SF/CPM blade initiates the changeover between the working to protect circuit, the changeover is transparent to the upper layer protocols and service layers that the switchover occurs.

Physical link monitoring of the link is performed by the CPU on the relevant IOM for both working and protect circuits.

2.3.2.6.7.2 Switchover process for received data

The Flexible Fast-Path complexes for both working and protect circuits are programmed to process ingress. The inactive (protect) circuit however is programmed to ignore all packet data. To perform the switchover from working circuit to the protect circuit the Flexible Fast-Path complex for the working circuit is set to ignore all data while the Flexible Fast-Path complex of the protect circuit is changed to accept data.

The ADM or compatible head-end transmits a valid data signal to both the working and protection circuits. The signal on the protect line is ignored until the working circuit fails or degrades to the degree that requires a switchover to the protect circuit. When the switchover occurs all services including all their QoS and filter policies are activated on the protection circuit.

2.3.2.6.8 APS user-initiated requests

The following subsections describe APS user-initiated requests.

2.3.2.6.8.1 Lockout protection

The lockout of protection disables the use of the protection line. Because the **tools>perform>aps>lockout** command has the highest priority, a failed working line using the protection line is switched back to itself even if it is in a fault condition. No switches to the protection line are allowed when locked out.

2.3.2.6.8.2 Request switch of active to protection

The request or manual switch of active to protection command switches the active line to use the protection line unless a request of equal or higher priority is already in effect. If the active line is already on the protection line, no action takes place.

2.3.2.6.8.3 Request switch of active to working

The request or manual switch of active to working command switches the active line back from the protection line to the working line unless a request of equal or higher priority is already in effect. If the active line is already on the working line, no action takes place.

2.3.2.6.8.4 Forced switching of active to protection

The forced switch of active to protection command switches the active line to the protection line unless a request of equal or higher priority is already in effect. When the forced switch of working to protection command is in effect, it may be overridden either by a lockout of protection or by detecting a signal failure on the protection line. If the active line is already on the protection line, no action takes place.

2.3.2.6.8.5 Forced switch of active to working

The forced switch of active to working command switches the active line back from the protection line to the working unless a request of equal or higher priority is already in effect.

2.3.2.6.8.6 Exercise command

The exercise command is only supported in the bidirectional mode of the 1+1 architecture. The exercise command is specified in the **tools>perform>aps>force>exercise** context and exercises the protection line by sending an exercise request over the protection line to the tail-end and expecting a reverse request response back. The switch is not actually completed during the exercise routine.

2.3.2.6.9 APS and SNMP

SNMP Management of APS uses the APS-MIB (from rfc3498) and the TIMETRA-APS-MIB.

[Table 22: Switching mode to MIB mapping](#) shows the mapping between APS switching modes and MIB objects.

Table 22: Switching mode to MIB mapping

switching-mode	TIMETRA-APS-MIB tApsProtectionType	APS-MIB apsConfigDirection
Bidir 1+1 Sig APS (bidirectional)	onePlusOneSignalling (1)	bidirectional (2)
Uni 1+1 Sig APS (unidirectional)	onePlusOneSignalling (1)	unidirectional (1)
Uni 1+1 Sig+Data APS (uni-1plus1)	onePlusOne (2)	unidirectional (1)

apsConfigMode in the APS-MIB is set to onePlusOneOptimized for Annex B operation.

2.3.2.6.10 APS applicability, restrictions, and interactions



Note: The Release Notes for the relevant SR OS release should be consulted for details about APS restrictions.

Table 23: Supported APS mode combinations shows the supported APS mode combinations.

Table 23: Supported APS mode combinations

	Bidirectional 1+1 signaling APS	Unidirectional 1+1 signaling APS	Unidirectional 1+1 signaling and datapath APS
Single Chassis APS (SC-APS)	✓ ²	✓	Supported for 7750 SR-c4/12 platforms only
Multi-Chassis APS (MC-APS)	✓ ²		

2.3.2.6.10.1 APS switchover impact on statistics

All SAP-level statistics are retained with an APS switch. A SAP reflects the data received regardless of the number of APS switches that has occurred. ATM statistics, however, are cleared after an APS switch. Therefore, any ATM statistics viewed on an APS port are only the statistics since the current active member port became active.

Physical layer packet statistics on the APS group reflect what is currently on the active member port.

Port and path-level statistics follow the same behavior as described above.

Any SONET physical-layer statistics (for example, B1,B2,B3,...) on the APS port are only what is current on the active APS member port.

2.3.2.6.10.2 APS switchover during CPM switchover

An APS switchover immediately before, during or immediately after a CPM switchover may cause a longer outage than normal.

2.3.2.6.10.3 Removing or failure of a protect MDA

The detection of an MDA removal or an MDA failure can take additional time. This can affect the APS switchover time upon the removal or failure of a protection MDA. If the removal is scheduled during maintenance, it is recommended that the port and, or protect circuit be shutdown first to initiate an APS switchover before the MDA maintenance is performed.

² TDM satellite supports these modes only.

2.3.2.6.10.4 Mirroring support

Mirroring parameters configured on a specific port or service, are maintained during an APS failover.

2.3.2.6.11 Example APS applications

The following subsections provide APS application examples.

2.3.2.6.11.1 Example APS application: MC-APS on channelized interfaces

The 7750 SR supports APS on channelized interfaces. This allows the router to be deployed as the radio access network (RAN) aggregation router which connects the base transceiver station (BTS) and the radio network controller (RNC).

2.3.2.7 E-LMI

The Ethernet Local Management Interface (E-LMI) protocol is defined in Metro Ethernet Forum (MEF) technical specification MEF16. This specification defines the protocol and procedures that convey the information for auto-configuration of a CE device and provides the means for EVC status notification. MEF16 does not include link management functions. In the Ethernet context that role is already accomplished with Clause 57 Ethernet OAM (formerly 802.3ah).

The SR OS currently implements the User Network Interface-Network (UNI-N) functions for status notification supported on Ethernet access ports with dot1q encapsulation type. Notification related to status change of the EVC and CE-VLAN ID to EVC mapping information is provided as a one to one between SAP and EVC.

The E-LMI frame encapsulation is based on IEEE 802.3 untagged MAC frame format using an ether-type of 0x88EE. The destination MAC address of the packet 01-80-C2-00-00-07 is dropped by any 802.1d compliant bridge that does not support or have the E-LMI protocol enabled. This means the protocol cannot be tunneled.

Status information is sent from the UNI-N to the UNI-C, either because a status inquiry was received from the UNI-C or unsolicited. The Active and Not Active EVC status are supported. The Partially Active state is left for further study.

The bandwidth profile sub-information element associated with the EVC Status IE does not use information from the SAP QoS policy. A value of 0 is used in this release as MEF 16 indicates the bandwidth profile sub-IE is mandatory in the EVC Status IE. The EVC identifier is set to the description of the SAP and the UNI identifier is set to the description configured on the port. Further, the implementation associates each SAP with an EVC. Currently, support exists for CE-VLAN ID/EVC bundling mode.

The E-LMI the UNI-N can participate in the OAM fault propagation functions. This is a unidirectional update from the UNI-N to the UNI-C and interacting with service manager of VLL, VPLS, VPRN and IES services.

2.3.2.8 LLDP

The IEEE 802.1ab Link Layer Discovery Protocol (LLDP) standard defines protocol and management elements that are suitable for advertising information to stations attached to the same IEEE 802 LAN (emulation) for the purpose of populating physical or logical topology and device discovery management

information databases. The protocol facilitates the identification of stations connected by IEEE 802 LANs/ MANs, their points of interconnection, and access points for management protocols.

Note that LAN emulation and logical topology wording is applicable to customer bridge scenarios (enterprise/carrier of carrier) connected to a provider network offering a transparent LAN emulation service to their customers. It helps the customer bridges detect misconnection by an intermediate provider by offering a view of the customer topology where the provider service is represented as a LAN interconnecting these customer bridges.

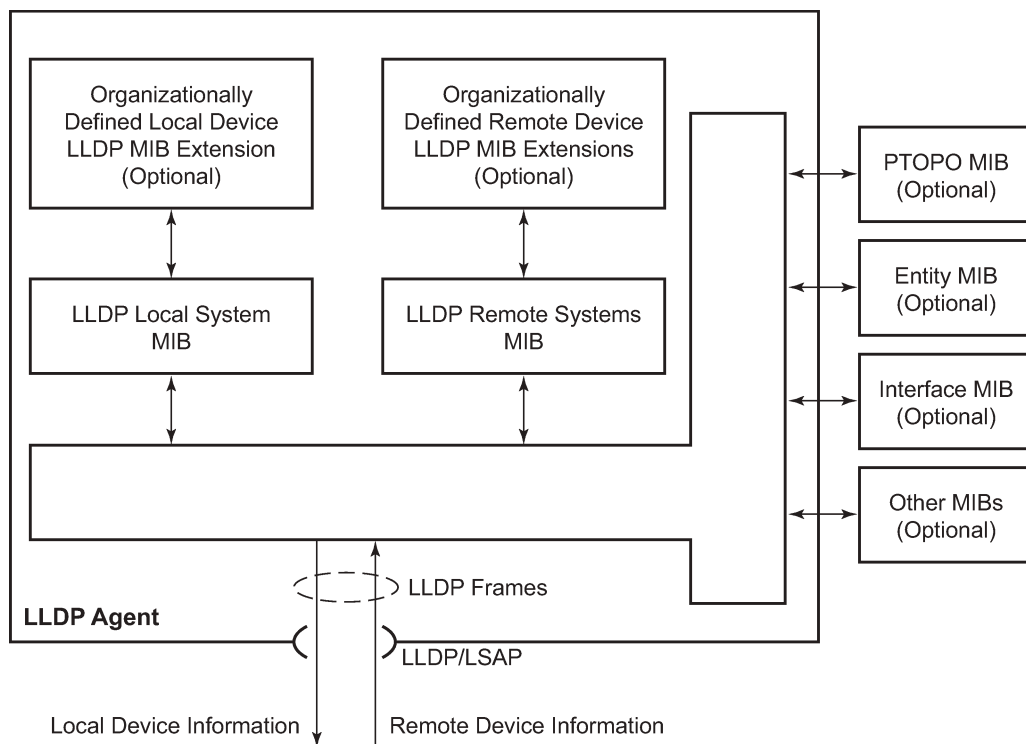
The IEEE 802.1ab standard defines a protocol that:

1. Advertises connectivity and management information about the local station to adjacent stations on the same IEEE 802 LAN.
2. Receives network management information from adjacent stations on the same IEEE 802 LAN.
3. Establishes a network management information schema and object definitions that are suitable for storing connection information about adjacent stations.

Attributes of the IEEE 802.1ab standard include:

- Operating with all IEEE 802 access protocols and network media.
- Providing compatibility with a number of MIBs as shown in [Figure 16: LLDP internal architecture for a network node](#).

Figure 16: LLDP internal architecture for a network node

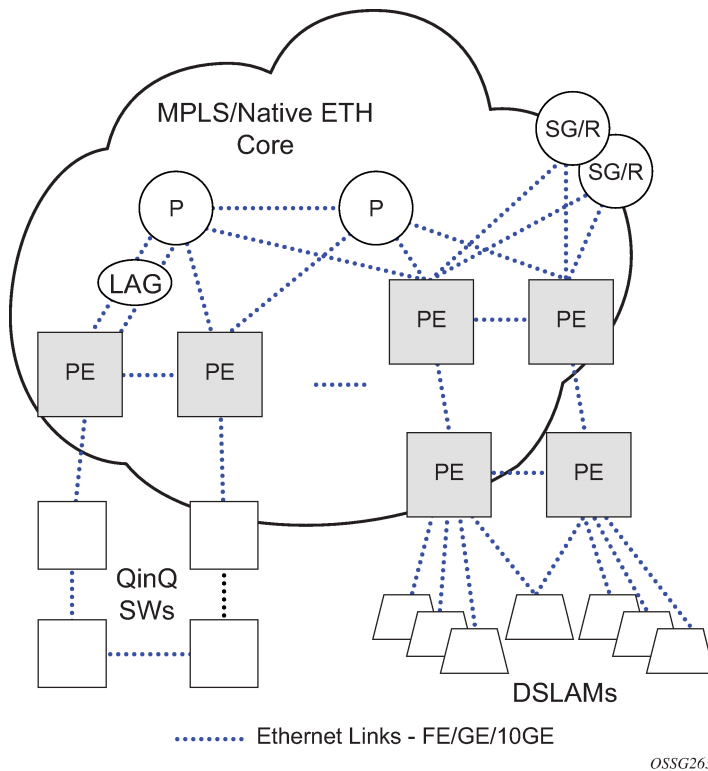


OSSG262

Network operators must be able to discover the topology information to detect and address network problems and inconsistencies in the configuration. Moreover, standard-based tools can address the complex network scenarios where multiple devices from different vendors are interconnected using Ethernet interfaces.

The example shown in [Figure 17: Generic customer use case For LLDP](#) depicts a MPLS network that uses Ethernet interfaces in the core or as an access/handoff interfaces to connect to different kind of Ethernet enabled devices such as service gateway/routers, QinQ switches, DSLAMs or customer equipment.

Figure 17: Generic customer use case For LLDP



IEEE 802.1ab LLDP running on each Ethernet interfaces in between all the above network elements may be used to discover the topology information.

Operators who are utilizing IOM/IMM can tunnel the nearest-bridge at the port level using the **tunnel-nearest-bridge** command under the **config>port>ethernet>lldp>destmac** (nearest-bridge) hierarchy. The dest-mac nearest-bridge must be disabled for tunneling to occur.

2.3.2.8.1 LLDP protocol features

LLDP is a unidirectional protocol that uses the MAC layer to transmit specific information related to the capabilities and status of the local device. Separately from the transmit direction, the LLDP agent can also receive the same kind of information for a remote device which is stored in the related MIBs.

LLDP itself does not contain a mechanism for soliciting specific information from other LLDP agents, nor does it provide a specific means of confirming the receipt of information. LLDP allows the transmitter and the receiver to be separately enabled, making it possible to configure an implementation so the local LLDP agent can either transmit only or receive only, or can transmit and receive LLDP information.

The information fields in each LLDP frame are contained in an LLDP Data Unit (LLDPDU) as a sequence of variable length information elements, that each include type, length, and value fields (known as TLVs), where:

- Type identifies what kind of information is being sent.
- Length indicates the length of the information string in octets.
- Value is the actual information that needs to be sent (for example, a binary bit map or an alphanumeric string that can contain one or more fields).

Each LLDPDU contains four mandatory TLVs and can contain optional TLVs as selected by network management:

- Chassis ID TLV
- Port ID TLV
- Time To Live TLV
- Zero or more optional TLVs, as allowed by the maximum size of the LLDPDU
- End Of LLDPDU TLV

The chassis ID and the port ID values are concatenated to form a logical identifier that is used by the recipient to identify the sending LLDP agent/port. Both the chassis ID and port ID values can be defined in a number of convenient forms. After selected however, the chassis ID/port ID value combination remains the same as long as the particular port remains operable.

A non-zero value in the TTL field of the Time To Live TLV tells the receiving LLDP agent how long all information pertaining to this LLDPDU's identifier is valid so that all the associated information can later be automatically discarded by the receiving LLDP agent if the sender fails to update it in a timely manner. A zero value indicates that any information pertaining to this LLDPDU's identifier is to be discarded immediately.

A TTL value of zero can be used, for example, to signal that the sending port has initiated a port shutdown procedure. The End Of LLDPDU TLV marks the end of the LLDPDU.

The implementation defaults to setting the port-id field in the LLDP OAMPDU to **tx-local**. This encodes the port-id field as ifIndex (sub-type 7) of the associated port. Some network management systems use the ifIndex value to properly build the Layer Two Topology Network Map. However, this numerical value is difficult to interpret or readily identify the LLDP peer. Configuration options are available to control the encoding of the port-id information and the associated subtype using the **port-id-subtype** option. Three options are supported for the port-id-subtype:

tx-if-alias

transmits the ifAlias String (subtype 1) that describes the port as stored in the IF-MIB, either user configured description or the default entry (10/100/Gig Ethernet SFP)

tx-if-name

transmits the ifName string (subtype 5) that describes the port as stored in the IF-MIB, ifName info

tx-local

the interface ifIndex value (subtype 7)

IPv6 (address subtype 2) and IPv4 (address subtype 1) LLDP System Management addresses are supported. The IP addresses can be selected from the system IP addressing, the out-of-band management address (BOF), or both.

If the **port-desc** TLV is enabled it has a maximum of 255 byte limit. However, truncation of the port description information occurs if the combination of the **port-id** and the **ifDesc** (for example 1/1/c1/2, 10-Gig Ethernet) exceeds the 255 byte maximum. The truncation of the port description information is equal to the number of bytes required to include the **ifDesc**.

2.3.2.9 Exponential Port Dampening

Exponential Port Dampening (EPD) provides the ability to automatically block a port from reuse for a period of time after physical link-down and physical link-up events. If a series of down-up events occur close together, EPD keeps the port's operational state down for a longer period than if only one down-up event has occurred. The router avoids using that port if external events are causing the link state to fluctuate. The more events that occur, the longer the port is kept down and avoided by the routing protocols.

EPD behavior uses a fixed penalty amount per link-down event and a half-life decay equation to reduce these penalties over time. The following equation defines exponential decay:

$$N(t) = N_0 \left(\frac{1}{2}\right)^{\frac{t}{t_{1/2}}}$$

sw0109

where:

$N(t)$ is the quantity that still remains after a time t

N_0 is the initial quantity

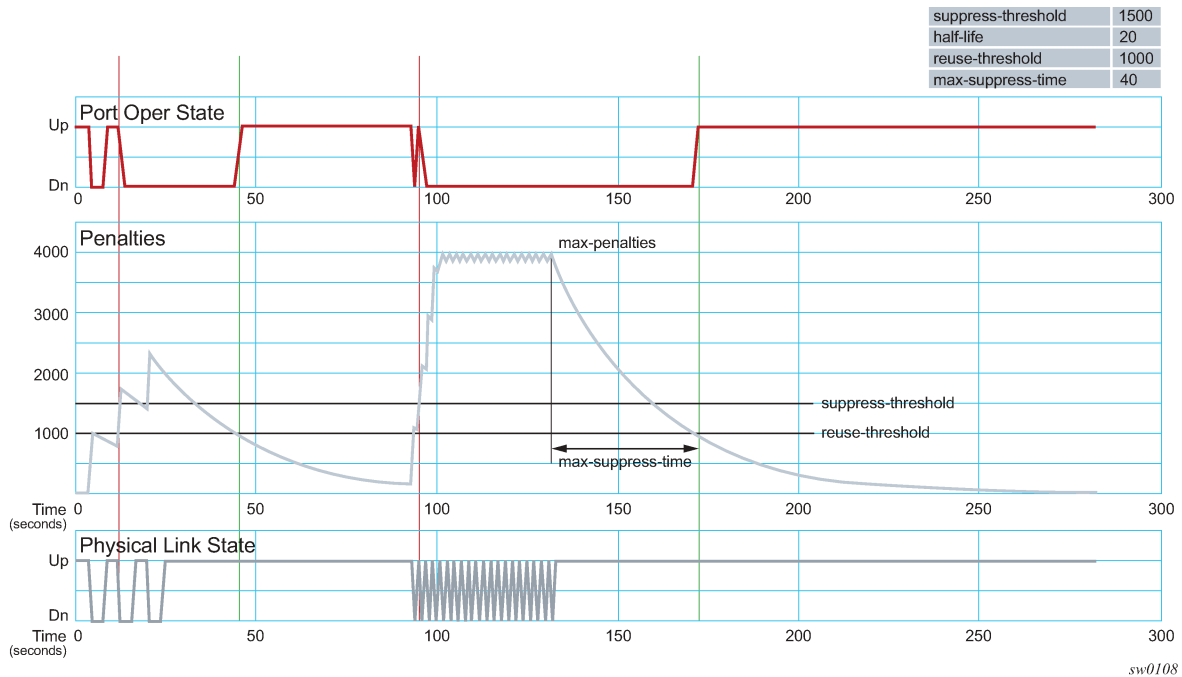
$t_{1/2}$ is the half-life

In dampening, N_0 refers to the starting penalties from the last link-down event. The quantity $N(t)$ refers to the decayed penalties at a specific time, and is calculated starting from the last link-down event (that is, from the time when N_0 last changed).

This equation can also be used on a periodic basis by updating the initial quantity value N_0 each period and then computing the new penalty over the period (t).

The following figure shows an example usage of the EDP feature.

Figure 18: EPD example



At time ($t = 0$) in the preceding figure, the initial condition has the link up, the accumulated penalties are zero, the dampening state is idle, and the port operational state is up. The following series of events and actions occur.

1. $t = 5$: link-down event
 - the accumulated penalties are incremented by 1000
 - the accumulated penalties now equal 1000, which is less than the suppress threshold (of 1500), so the dampening state is idle
 - because the dampening state is idle, link-down is passed to the upper layer
 - link-down triggers the port operational state to down
2. $t = 9$: link-up event
 - the accumulated penalties equal 869, which is less than the suppress threshold, so the dampening state remains as idle
 - because the dampening state is idle, link-up is passed to the upper layer
 - link-up triggers the port operational state to up
3. $t = 13$: link-down event
 - the accumulated penalties are incremented by 1000
 - the accumulated penalties now equal 1755, which is greater than the suppress threshold, so the dampening state is changed to active
 - because the dampening state just transitioned to active, link-down is passed to the upper layer
 - link-down triggers the port operational state to down
4. $t = 17$: link-up event

- the accumulated penalties equal 1527, which is above the reuse threshold (of 1000) and greater than the suppress threshold, so the dampening state remains as active
 - because the dampening state is active, link-up is not passed to the upper layer
 - the port operational state remains down
- 5.** $t = 21$: link-down event
- the accumulated penalties are incremented by 1000
 - the accumulated penalties now equal 2327, which is above the reuse threshold, so the dampening state remains as active
 - because the dampening state is active, link-down is not passed to the upper layer
 - the port operational state remains down
- 6.** $t = 25$: link-up event
- the accumulated penalties equal 2024, which is above the reuse threshold, so dampening state remains as active
 - because the dampening state is active, link-up is not passed to the upper layer
 - the port operational state remains down
- 7.** $t = 46$: accumulated penalties drop below the reuse threshold
- the accumulated penalties drop below the reuse threshold, so the dampening state changes to idle
 - because the dampening state is idle and the current link state is up, link-up is passed to the upper layer
 - the port operational state changes to up
- 8.** $t = 94$ to 133 : link-down and link-up events every second
- similar to previous events, the accumulated penalties increment on every link-down event
 - the dampening state transitions to active at $t = 96$, and link state events are not sent to the upper layer after that time
 - the upper layer keeps the port operational state down after $t = 96$
 - the accumulated penalties increment to a maximum of 4000
- 9.** $t = 133$: final link event of link-up
- the accumulated penalties equal 3863
 - the dampening state remains active and link state events are not sent to the upper layer
 - the upper layer keeps the port operational state down
- 10.** $t = 172$: accumulated penalties drop below the reuse threshold
- the accumulated penalties drop below the reuse threshold, so the dampening state changes to idle
 - because the dampening state is idle and the current link state is up, link-up is passed to the upper layer
 - the port operational state changes to up

2.3.3 Per port aggregate egress queue statistics monitoring

Monitoring the aggregate egress queue statistics per port provides in-profile, out-of-profile, and total statistics for both forwarded and dropped packets and octets on a specific port.

When enabled, all queues on the port are monitored, including SAP egress, network egress, subscriber egress, and egress queue group queues, as well as system queues which can be used, for example, to send port-related protocol packets (LACP, EFM, and so on).

This is enabled and disabled using the following command:

```
config port <port-id> [no] monitor-agg-egress-queue-stats
```

When enabled, the line card polls the related queues to derive the aggregates which provide the delta of the queue statistics since turning on the monitoring. This means that the reported statistics are not reduced by those from a deleted queue and so the aggregates correctly represent the forwarded/dropped statistics since the start of monitoring.

The aggregates can be shown with the following command:

```
show port [<port-id>] [statistics [egress-aggregate]] [detail]
```

As an example, the output below enables monitoring of aggregate egress queue statistics on port 2/1/1 and then shows the monitored statistics:

```
*A:PE# configure port 2/1/1 monitor-agg-egress-queue-stats
*A:PE# show port 2/1/1 statistics egress-aggregate

=====
Port 2/1/1 Egress Aggregate Statistics on Slot 2
=====
```

	Forwarded	Dropped	Total
PacketsIn	144	0	144
PacketsOut	0	0	0
OctetsIn	12353	0	12353
OctetsOut	0	0	0

```
=====
*A:PE#
```

To clear the aggregate statistics, the monitoring must be disabled and then re-enabled. The aggregate statistics are also cleared when the card is cleared (using a **clear card slot-number** command) or power-cycled (with the **tools perform card slot-id** command). Additionally, aggregate statistics related to MDA are cleared when the MDA is cleared (using the **clear mda mda-id** command) or the MDA is inserted into an IOM. The aggregate statistics are not cleared when a **shutdown/no shutdown** is performed on the card and, or MDA.

There is no specific limit on the number of queues that can be monitored, but the amount of each line card's CPU resources allocated to the monitoring is bounded; consequently, when more queues on a card's ports are monitored, the aggregate statistics are updated the less frequently.

Monitoring of aggregate statistics is supported on **PXC** sub-ports but not on a **PXC** physical port. It is also not supported on satellite ports.

2.3.4 Forward Error Correction

Users can use Forward Error Correction (FEC) on some ports to improve either the transmission reliability or reach, or both. FEC must always be used on some interface types while it is optional for other interface types. Also, some interface types allow more than one type of FEC. No matter what the setting of the FEC attributes, the transmitter and the receiver must have the same configuration, or the link will not work. The setting of FEC on a specific port is dependent on the interface type and the specific optical transceiver in use. Contact your Nokia representative for information about the options based on the transceiver in use.

2.4 Datapath mapping

The mapping between a card and its MDAs, FPs, MACs, connectors, and ports on hardware can be displayed using the **show datapath** command.

```
# show datapath 1
```

Card	[X10M/]MDA	FP	TAP	MAC	Chip Num	Connector	Port
1	x1/1	1	N/A	1		c1	
1	x1/1	1	N/A	1		c2	
1	x1/1	1	N/A	1		c3	
1	x1/1	1	N/A	2		c4	
1	x1/1	1	N/A	2		c5	
1	x1/1	1	N/A	2		c6	
1	x2/1	5	N/A	1		c1	
1	x2/1	5	N/A	1		c2	
1	x2/1	5	N/A	1		c3	
1	x2/1	5	N/A	1		c4	
1	x2/1	5	N/A	1		c5	
1	x2/1	5	N/A	1		c6	
1	x2/1	5	N/A	2		c7	
1	x2/1	5	N/A	2		c8	
1	x2/1	5	N/A	2		c9	
1	x2/1	5	N/A	2		c10	
1	x2/1	5	N/A	2		c11	
1	x2/1	5	N/A	2		c12	
1	x2/1	5	N/A	3		c13	
1	x2/1	5	N/A	3		c14	
1	x2/1	5	N/A	3		c15	
1	x2/1	5	N/A	3		c16	
1	x2/1	5	N/A	3		c17	
1	x2/1	5	N/A	3		c18	

2.5 Port Cross-Connect

2.5.1 PXC terminology

The following describes Port Cross-Connect (PXC) terminology:

- **Port Cross-Connect (PXC)**

PXC is a software concept representing a pair of logical ports interconnecting egress and ingress forwarding paths within the same forwarding complex.

The physical underpinning of a PXC can be either of the following:

- **a faceplate (physical) port in a loopback mode**

The PXC is referred to as a port-based PXC. Multiple PXCs can be created per a faceplate port.

- **a loopback configuration in the MAC chip**

The PXC is referred to as an internal or MAC-based PXC. Multiple PXCs can be created per MAC loopback.

- **PXC sub-port**

PXC sub-port is a logical port that is created under the PXC. Two interconnected PXC sub-ports are created per PXC. This is further described in [Port-based PXC](#).

- **Forwarding Complex (FC)**

FC is a chipset connected to a set of faceplate ports that processes traffic in the ingress direction (the ingress path) and the egress direction (the egress path). A line card can contain multiple FCs for increased throughput, while the inverse is not true, a single FC cannot be distributed over multiple line cards.

The terms cross-connect and loopback can be used interchangeably.

2.5.2 Overview

This section describes the Port Cross-Connect (PXC) feature implementation. PXC is a software concept representing a pair of logical ports interconnecting egress and ingress forwarding paths within the same forwarding complex (FC). In cross-connect functionality, an egress forwarding path is looped back to the ingress forwarding path on the same forwarding complex instead of leading out of the system. The FC is a chipset connected to a set of faceplate ports that processes traffic in the ingress direction (the ingress path) and the egress direction (the egress path). A line card can contain multiple FCs for increased throughput, but a single FC cannot be distributed over multiple line cards. The most common use for a cross-connect configuration is to process traffic entering the node. In this case, traffic passes through the ingress path twice. The first ingress pass is always on the FC on which traffic enters the node (an ingress line card), while the second ingress pass, achieved through the cross-connect, can be on any forwarding complex. The operator can select to co-locate the ingress line card and the line card hosting the cross-connect. In this co-located case, traffic is looped through the same ingress forwarding path twice.

The reasons for dual-stage ingress processing are related to the manipulation of multi-layer headers in the frame within the service termination context. This operation is, in some instances, too complex to perform in a single stage. Feeding the traffic from the first ingress stage to the second through the cross-connect is shown in [Figure 19: Traffic preprocessing using PXC](#). A cross-connect can be created in two ways:

- a faceplate (physical) port in a loopback mode
- a loopback configuration in the MAC chip. This implementation does not require a faceplate port

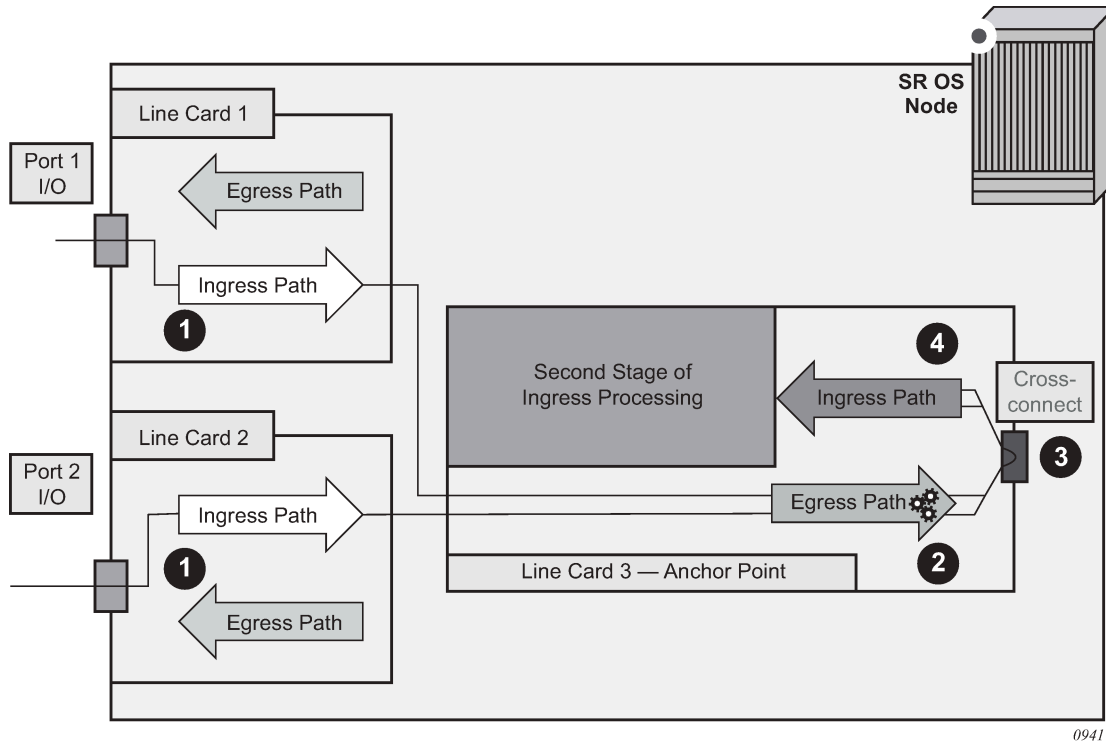
In both cases, the cross-connect is modeled in the system and in the CLI as a port, appropriately naming the feature Port Cross-Connect (PXC) software concept representing a pair of logical ports interconnecting egress and ingress forwarding paths within the same forwarding complex.

Conceptually, PXC functionality is similar to the functionality provided by two externally interconnected faceplate ports where traffic exits the system through one port (the egress path) and is immediately looped back into another port (the ingress path) through a cable.

Figure 19: Traffic preprocessing using PXC shows the traffic flow from the first to the second stage through a cross-connect in a system with PXC:

1. Traffic entering a node through a faceplate port is processed by the local ingress forwarding path (1) on the line cards 1 and 2. Traffic is then directed toward the PXC (3) on the line card 3.
2. The PXC (3) loops the traffic from the local egress path (2) into the local ingress forwarding path (4) where it is further processed.

Figure 19: Traffic preprocessing using PXC

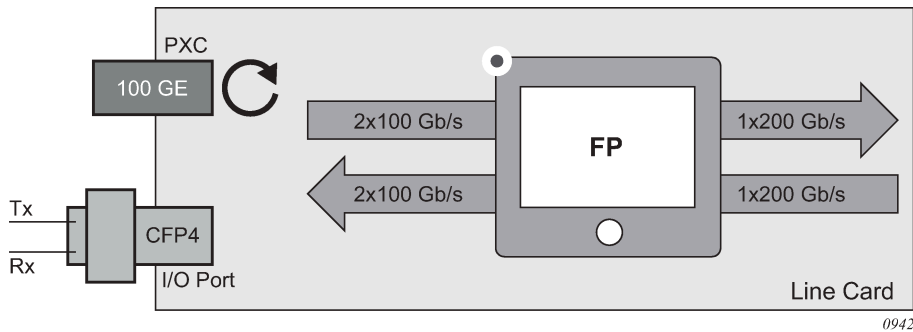


0941

2.5.3 Port-based PXC

The concept of a port-based PXC (a PXC based on a faceplate port in loopback mode) is shown in [Figure 20: Port-based PXC](#). This PXC does not require an optical transceiver.

Figure 20: Port-based PXC



The faceplate port is placed in a cross-connect mode with the following MD-CLI commands:

MD-CLI commands

```
configure {
  port-xc {
    pxc 1 {
      admin-state enable
      port-id 1/x1/1/c1/1
    }
  }
}
```

Multiple PXC's can be created on the same underlying cross-connect:

MD-CLI commands

```
configure {
  port-xc {
    pxc 1 {
      admin-state enable
      port-id 1/x1/1/c1/1
    }
    pxc 2 {
      admin-state enable
      port-id 1/x1/1/c1/1
    }
    pxc 3 {
      admin-state enable
      port-id 1/x1/1/c1/1
    }
  }
}
```

A faceplate port that has been placed in the loopback mode for PXC use, supports only hybrid mode of operation and dot1q encapsulation. The recommendation is that the MTU value be configured to the maximum value. dot1x tunneling is enabled and cannot be changed.

The pre-set dot1q Ethernet encapsulation on the faceplate port is irrelevant from the operator's perspective and there is no need to change it. The relevant encapsulation carrying service tags defined on PXC subports and that encapsulation is configurable. For more information, see [PXC sub-ports](#).

The following guidelines apply to a PXC configuration based on faceplate ports:

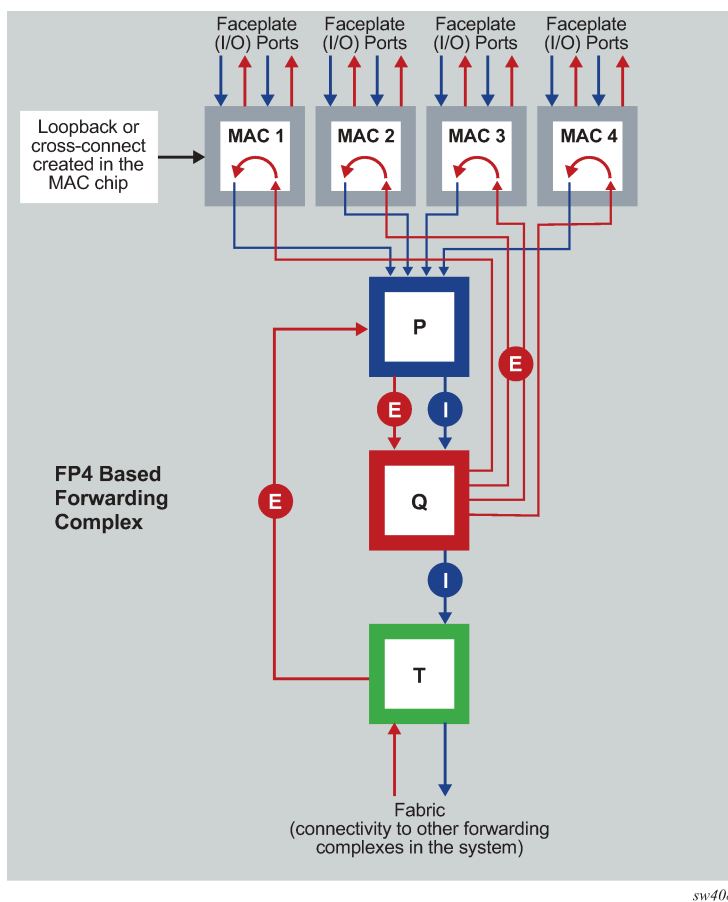
- Only unused faceplate ports (not associated with an interface or SAP) can be referenced within a PXC ID configuration.

- When the faceplate port is allocated to a PXC, it cannot be used outside of the PXC context. For example, an IP interface cannot use the faceplate port directly, or a SAP under a such port cannot be associated with an Epipe or VPLS service.

2.5.4 Internal PXC

With internal (or MAC-based) PXC, the egress path is cross-connected to the ingress path in the MAC chip, without the need to consume a faceplate port, as shown in [Figure 21: Internal cross-connect \(loopback\) in a MAC chip](#). The number of the MAC chips on a line card varies with the line card type. The **show datapath** command shows the MAC chip related connectivity information in the datapath (forwarding complex). This information is essential for proper configuration of the cross-connect.

Figure 21: Internal cross-connect (loopback) in a MAC chip



The cross-connect in the MAC chip is represented in the CLI as a loopback and is created with the following commands:

MD-CLI commands

```
configure card <id>{
  mda <id> {
    xconnect {
      mac <id> {
```

```

        description <string>
        loopback <id> {
            description <string>
            bandwidth <enum>
        }
    }
}

```

On the cards with IOM-s modules, an addition of xiom node is used:

MD-CLI commands

```

configure card <id>{
    xiom <xiom-slot> {
        mda <id> {
            xconnect {
                mac <id> {
                    description <string>
                    loopback <id> {
                        description <string>
                        bandwidth <enum> //enum: 10G, 25G, 40G, 100G, 400G//
                    }
                }
            }
        }
    }
}

```

The loopback created on the MAC chip at location **card**, **xiom**, **mda**, or **mac** is assigned a bandwidth in discrete steps. This is a Layer 2 bandwidth, which includes the Ethernet Layer 2 header, but excludes the 20 bytes of Ethernet preamble and the inter-packet gap.

The cross-connect cannot be used in this form, but instead, it must be represented as a port in the system so other software components, such as services, can access it.

The port associated with the loopback is automatically created in the classic CLI. In MD-CLI, such port must be manually created:

MD-CLI commands

```

configure {
    port <port-id> {
        admin-state enable
    }
}

```

where: the format for the cross-connect is:

<slotNum>/<mdaNum>/m<MACchipNum>/<loopback-id>

slotNum

the slot number

xiom

the xiom-slot

mdaNum

the MDA number

m

the keyword indicating that this is a cross-connect type port created in a MAC chip

MACchipNum

the MAC chip number, the physical location of the loopback on a forwarding complex

loopback-id

the loopback ID

For example, a cross-connect (loopback) port 1 on the card 1, xiom 'x1', mda 1, MAC chip 1 is created with the following commands:

MD-CLI commands

```
configure {
  port 1/x1/1/m1/1 {
    admin-state enable
  }
}
```

This port is ready to be used as a PXC, without having to reserve external ports. Multiple PXC's can be created on the same underlying cross-connect:

MD-CLI commands

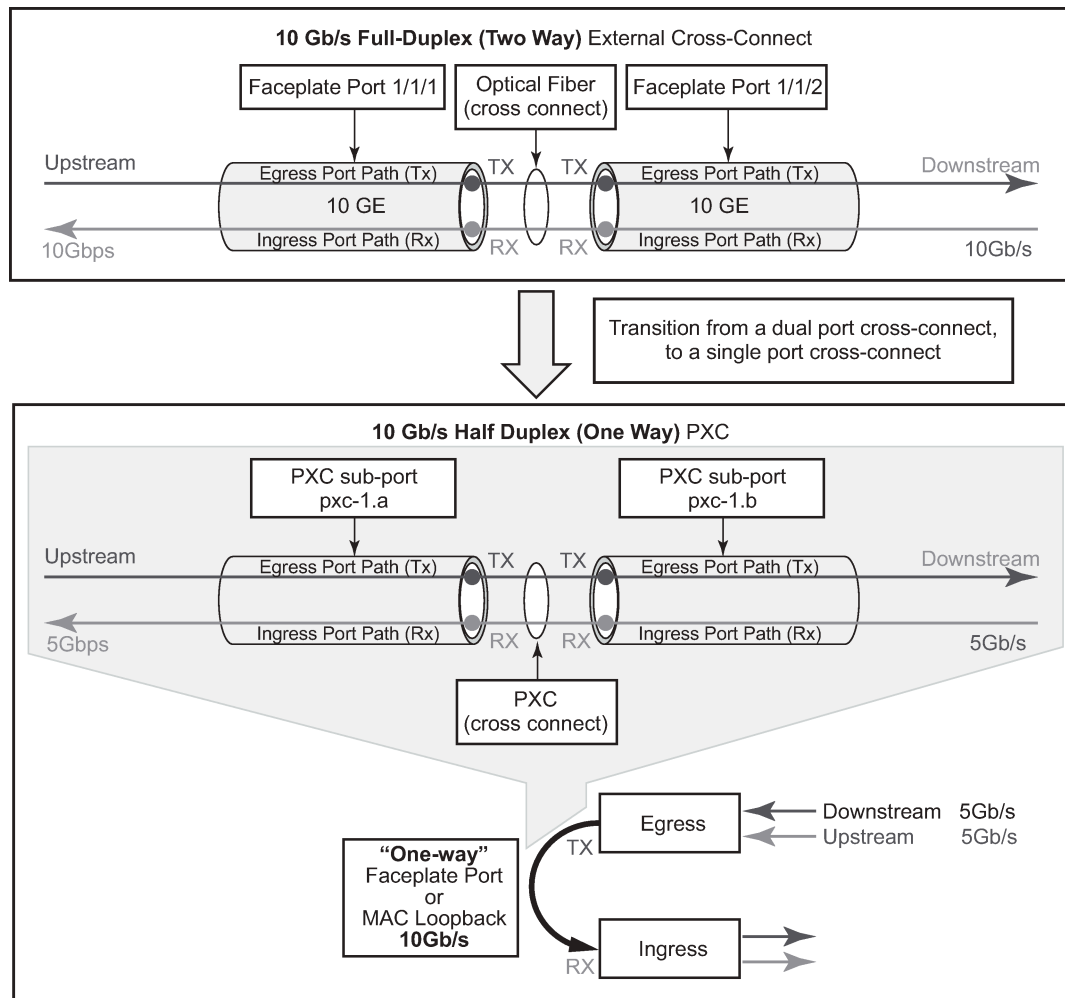
```
configure {
  port-xc {
    pxc 1 {
      admin-state enable
      port-id 1/x1/1/m1/1
    }
    pxc 2 {
      admin-state enable
      port-id 1/x1/1/m1/1
    }
    pxc 3 {
      admin-state enable
      port-id 1/x1/1/m1/1
    }
  }
}
```

The loopback created in the MAC chip does not have an administrative state, but the port created on top of it does have an administrative state.

2.5.5 PXC sub-ports

[Figure 22: Two cross-connected external ports versus a single cross-connect](#) displays the benefit of PXC sub-ports on top of the cross-connect, which is analogous to two distinct faceplate ports that are connected by a fiber cable.

Figure 22: Two cross-connected external ports versus a single cross-connect



0943

Bidirectional connectivity provided by PXC requires two sub-ports, one in each direction. The router uses these PXC sub-ports as logical configurations to transmit traffic in both directions over a half-duplex (one-way) cross-connect created in the system. As a result, the total bandwidth capacity supported by the mated PXC sub-ports is limited by the bandwidth capacity of the underlying cross-connect (a single faceplate port or a MAC loopback).

For example, if a 10 Gb/s faceplate port is allocated for PXC functions, the sum of downstream and upstream traffic on the mated PXC sub-ports is always less than or equal to 10 Gb/s. The bandwidth distribution is flexible; it can be symmetric (5 Gb/s downstream and 5 Gb/s upstream), or asymmetric (9 Gb/s downstream and 1 Gb/s upstream, 8 Gb/s downstream and 2 Gb/s upstream, or any other downstream and upstream distribution combination). Therefore, the faceplate port speed from the PXC perspective is half-duplex.

Similar logic can be followed for MAC-based PXC, with two key differences:

- The bandwidth (for example, 100 Gb/s) is configured under the MAC loopback and there is no need to allocate an additional faceplate port.

- PXC traffic is not reserved as part of the faceplate port bandwidth, as it is in the port-based PXC where a faceplate port is reserved only for PXC traffic. Instead, the PXC traffic is added to the traffic from the faceplate ports even in situations where all faceplate ports are 100% used, potentially oversubscribing the forwarding complex.

After the faceplate port or the port based on MAC loopback is associated with a PXC ID, a pair of mated PXC sub-ports is automatically created in the classic CLI by the SR OS.

In MD-CLI, the user must manually create the sub-ports.

In classic and MD-CLI, the sub-ports must be explicitly enabled with the **no shutdown** command in classic CLI and the **admin-state enable** command in MD-CLI.

The two PXC sub-ports are distinguishable by ".a" and ".b" suffixes. They transmit traffic toward each other, simulating two ports that are interconnected.

Although, the most PXC sub-ports parameters are configurable, specific parameters are fixed and cannot be changed. For example, PXC sub-ports are created in a hybrid mode and this cannot be modified.

Each PXC sub-port is internally (within the system) represented by an internal four-byte VLAN tag which is not visible to the operator. Therefore, traffic carried over the PXC contains four extra bytes, which must be accounted for in the QoS configured on PXC sub-ports.

Example: Classic CLI

```
configure
  port-xc
    pxc 1 create
      port 1/1/1
      [no] shutdown
    pxc 2 create
      port 1/1/2
      [no] shutdown
```

The preceding configuration automatically creates the following PXC sub-ports in CLI:

```
configure
  port
    pxc-1.a # cross-connected with pxc-1.b
    pxc-1.b # cross-connected with pxc-1.a

    pxc-2.a # cross-connected with pxc-2.b
    pxc-2.b# cross-connected with pxc-2.a
```

2.5.6 Bandwidth considerations and QoS

Bandwidth consumed by PXC based on faceplate ports correlates with the faceplate's port capacity. Because each PXC allocates a faceplate port for exclusive use, the PXC capacity cannot exceed the card capacity that is already allocated for the faceplate ports. In other words, a PXC based on a faceplate port does not add any additional bandwidth to the forwarding complex. This is in contrast to the MAC-based PXC, where bandwidth consumed on each PXC is added to the faceplate port capacity. If bandwidth is not carefully managed on cards with MAC-based PXC, extensive periods of oversubscription can occur. Operating under extended periods of congestion is not recommended and should be avoided. Therefore, practical use of a MAC-based PXC makes sense in an environment where the utilization of faceplate ports is relatively low so the remaining bandwidth can be used for traffic flowing through the MAC-based PXC.

The bandwidth management in the PXC environment is performed through existing QoS mechanisms, and in addition, for MAC-based PXC by careful selection of the MAC chip and in some instances in fabric taps. The user can control the selection of these entities.

2.5.6.1 Location selection for internal PXC

The following are general guidelines for MAC chip selection and the loopback naming:

- A suitable card candidate has lower actual traffic volumes through faceplate ports so the unused bandwidth in the forwarding complex can be used for PXC.
- A suitable MAC chip candidate is connected to the faceplate ports with lower bandwidth utilization.
- SR-s platforms with IOM-s cards (XIOM configuration), a loopback ID influences the selection of the fabric tap to which the PXC traffic is mapped. In these platforms, loopback should be distributed over the fabric taps in a way that avoids congestion. This is described in [Internal PXC and source fabric taps](#).

The connectivity layout between the MAC chips, faceplate ports, and fabric taps is shown in the output of the following command:

```
A:admin@cse-B04# /show datapath 1 detail
=====
Card    [XIOM/]MDA  FP  TAP  MAC Chip Num  Connector  Port
-----
1      x1/1        1   1    1             c1         1/x1/1/c1/1
1      x1/1        1   1    1             c1         1/x1/1/c1/2
1      x1/1        1   1    1             c2         1/x1/1/c2/1
1      x1/1        1   1    1             c2         1/x1/1/c2/2
1      x1/1        1   1    1             c3         1/x1/1/c3/1
1      x1/1        1   1    1             c3         1/x1/1/c3/2
1      x1/1        1   N/A  2             c4
1      x1/1        1   N/A  2             c5
1      x1/1        1   N/A  2             c6
1      x1/1        1   1    1             N/A        1/x1/1/m1/1
1      x1/2        2   1    1             c1         1/x1/2/c1/1
1      x1/2        2   1    1             c1         1/x1/2/c1/2
1      x1/2        2   1    1             c1         1/x1/2/c1/3
1      x1/2        2   1    1             c1         1/x1/2/c1/4
1      x1/2        2   1    1             c1         1/x1/2/c1/5
1      x1/2        2   1    1             c1         1/x1/2/c1/6
1      x1/2        2   1    1             c1         1/x1/2/c1/7
1      x1/2        2   1    1             c1         1/x1/2/c1/8
1      x1/2        2   1    1             c1         1/x1/2/c1/9
1      x1/2        2   1    1             c1         1/x1/2/c1/10
1      x1/2        2   N/A  1             c2
1      x1/2        2   N/A  1             c3
1      x1/2        2   N/A  1             c4
1      x1/2        2   N/A  1             c5
1      x1/2        2   N/A  1             c6
1      x1/2        2   N/A  2             c7
1      x1/2        2   N/A  2             c8
1      x1/2        2   N/A  2             c9
1      x1/2        2   2    2             c10        1/x1/2/c10/1
1      x1/2        2   2    2             c10        1/x1/2/c10/2
1      x1/2        2   2    2             c10        1/x1/2/c10/3
1      x1/2        2   2    2             c10        1/x1/2/c10/4
1      x1/2        2   2    2             c10        1/x1/2/c10/5
1      x1/2        2   2    2             c10        1/x1/2/c10/6
1      x1/2        2   2    2             c10        1/x1/2/c10/7
1      x1/2        2   2    2             c10        1/x1/2/c10/8
1      x1/2        2   2    2             c10        1/x1/2/c10/9
```

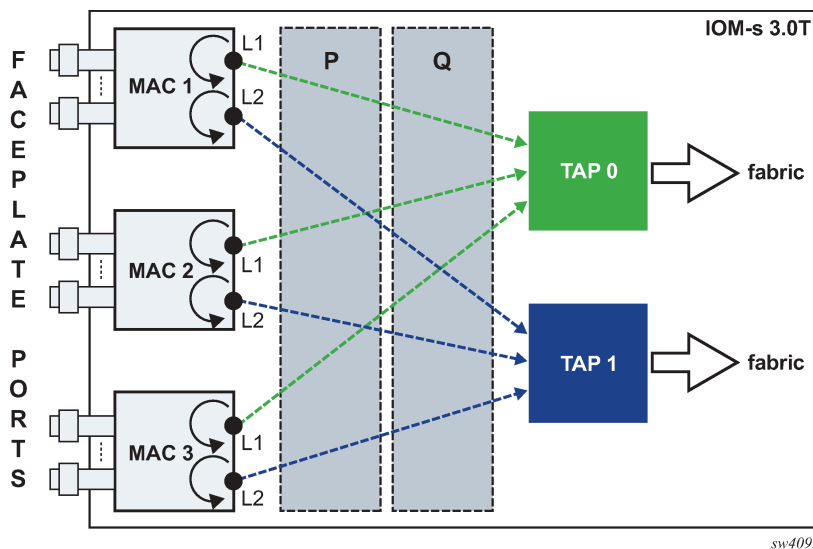
1	x1/2	2	2	2	c10	1/x1/2/c10/10
1	x1/2	2	N/A	2	c11	
1	x1/2	2	N/A	2	c12	
1	x1/2	2	N/A	3	c13	
1	x1/2	2	N/A	3	c14	
1	x1/2	2	N/A	3	c15	
1	x1/2	2	N/A	3	c16	
1	x1/2	2	N/A	3	c17	
1	x1/2	2	N/A	3	c18	

2.5.6.1.1 Internal PXC and source fabric taps

PXC traffic passing through the MAC loopback is mapped to a specific source fabric tap that moves the traffic from the local source forwarding complex into the fabric and toward the destination forwarding complex. A fabric tap represents a chip that connects a forwarding complex to the system fabric. Traffic that is on its way from an ingress port (any port, including a PXC port) to the destination port, is always mapped to the same fabric tap (source fabric tap) on the ingress forwarding complex. If the source forwarding complex has two fabric taps, the fabric tap selection plays a role in optimal bandwidth distribution. An example of these forwarding complexes can be found on IOM-s cards in SR-s platforms.

On IOM-s 3.0T, the source tap selection is based on the loopback ID. The mapping scheme is simple; loopbacks with even IDs are mapped to one source tap while loopbacks with odd IDs are mapped to the other. This is shown in [Figure 23: Mapping of internal loopbacks to source taps](#). On IOM-s 1.5T, the mapping is based on the MDA number.

Figure 23: Mapping of internal loopbacks to source taps

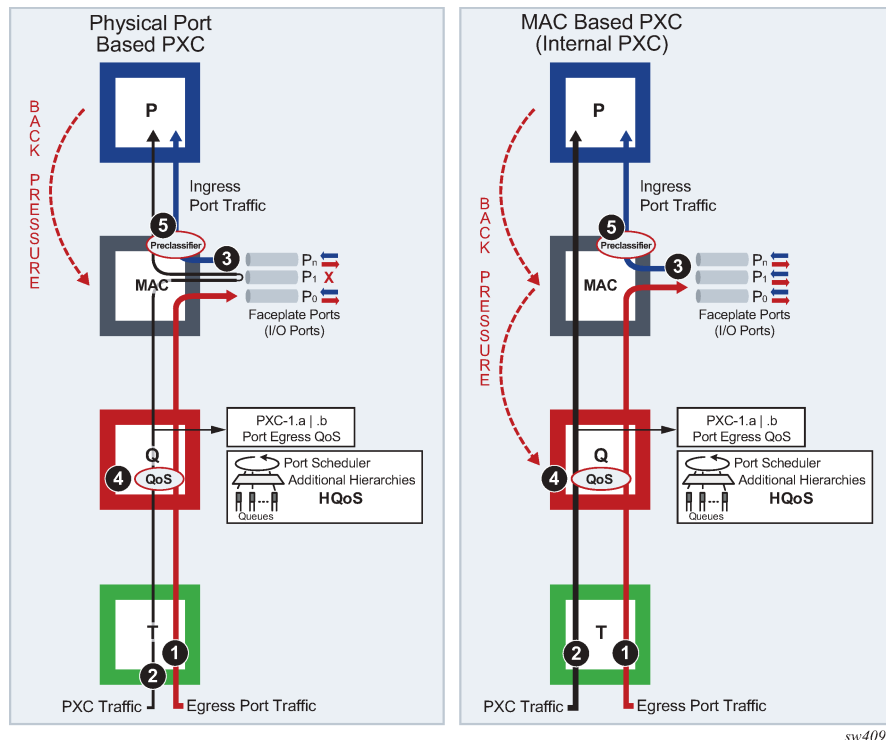


2.5.6.2 QoS

Interaction points between the PXC traffic and non-PXC traffic in the FC depends on the configured PXC type. [Figure 24: Interaction between PXC and non-PXC traffic](#) shows this interaction as the traffic enters the egress forwarding path from the fabric tap (T). This traffic consists of non-PXC traffic (1) destined for the egress faceplate ports and PXC traffic (2) that is sent (cross-connected) to the ingress forwarding path

(P) within the same forwarding complex. Regular ingress traffic from the faceplate ports (3) is added to the stream and merged into the same ingress forwarding path as the PXC traffic.

Figure 24: Interaction between PXC and non-PXC traffic



The physical port-based PXC configuration on the left side of [Figure 24: Interaction between PXC and non-PXC traffic](#), shows interaction of the three traffic streams on the forwarding complex with a PXC based on the faceplate ports. To manage congestion, the user-configured input can be exerted in points 4 and 5.

Point 4 represents regular egress QoS in the traffic manager (Q) applied to an egress port. In this setup, the faceplate port P1 is reserved for PXC traffic which is represented by the two sub ports (PXC sub-ports **pxc-id.a** and **pxc-id.b**). Egress QoS is applied to each PXC subport.

```
pxc-id      - pxc-<id>.<sub-port>
             pxc          - keyword
             id           - [1..64]
             sub-port     - [a..b]
```

Point 5 represents a pre-classifier in the MAC chip that manages ingress bandwidth if transient bursts occur in the ingress datapath (P), which then exerts back pressure toward the MAC. During congestion, the pre-classifier arbitrates between regular ingress traffic from the faceplate ports and the PXC traffic.

The MAC-based PXC configuration on the right side of [Figure 24: Interaction between PXC and non-PXC traffic](#) shows the traffic flow in the FC with the MAC-based PXC. Similar to the previous case, the existing egress QoS in the traffic manager (Q) in point 4 is applied to the egress ports. However, the egress port in the PXC case is not a faceplate port but instead it is represented by the configured loopback in the MAC chip. This loopback is configured with the maximum bandwidth using the **configure card id xiom xiom-slot mda id xconnect mac id loopback id** command. The congestion management logic with internal PXC diverges from the previous scenario because the PXC traffic is moved through the MAC chip straight to the ingress datapath (P), bypassing the egress faceplate ports. Therefore, the pre-classifier in point 5 has no

effect on PXC traffic. Any congestion in the (P) and MAC is managed in the traffic manager (Q) at point 4, as a result of the back pressure from the (P) and MAC toward the (Q).

2.5.6.2.1 QoS on PXC sub-ports

The network operator must understand the concept of the PXC sub-ports described in [Port-based PXC](#) for correct egress QoS configuration in the traffic manager (Q).

The following summarizes key points for the PXC sub ports:

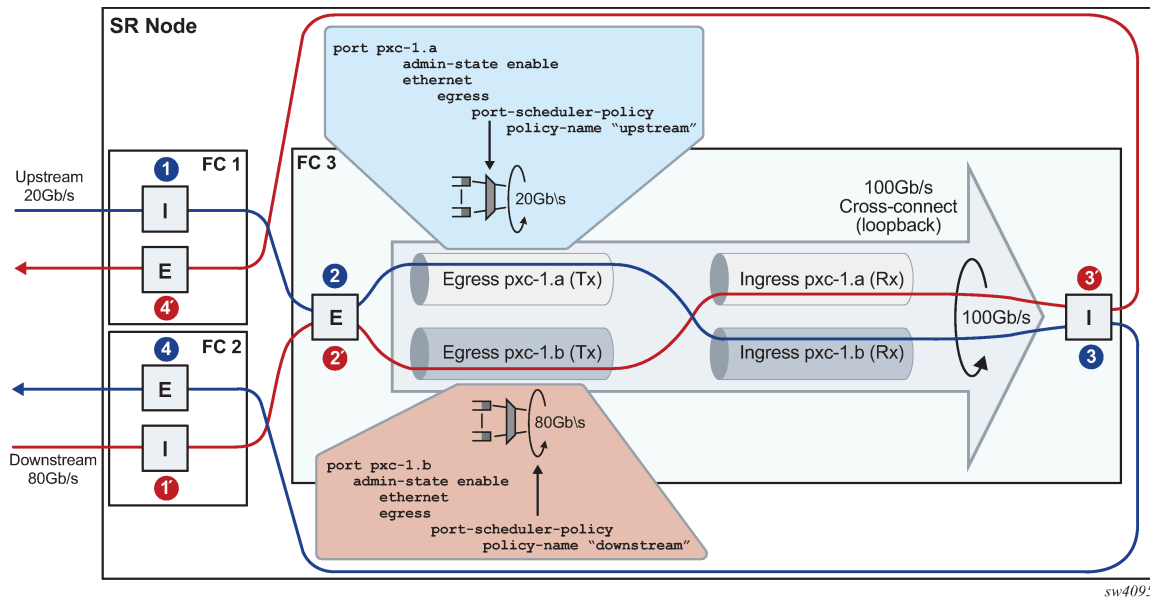
- Each subport (**pxc-id.a** and **pxc-id.b**) in a PXC is, in the context of egress QoS, treated as a separate port with its own port scheduler policy.
- Both sub-ports are created on top of the same loopback configuration (port-based or MAC-based). For faceplate ports, this bandwidth is determined by the port capabilities (for example, a 100 Gb/s port versus a 400 Gb/s port) and for the MAC loopback, this bandwidth is configurable.

Funneling traffic from two PXC sub-ports through the same loopback requires separate bandwidth management for each PXC sub-ports. The sum of the configured bandwidth caps for the Egress Port Scheduler (EPS) under the two PXC sub-ports should not exceed the bandwidth capacity of the underlying loopback. [Figure 25: Bandwidth management on PXC sub-ports](#) shows an example of this concept where each PXC sub-port is divided into two parts, the Tx or the egress part and the Rx or the ingress part. [Figure 25: Bandwidth management on PXC sub-ports](#) shows bidirectional traffic entering and exiting the SR node at forwarding complex 1 and 2, with PXC processing on forwarding complex 3. In the upstream direction, traffic enters SR node at the ingress forwarding complex 1 at point (1) and is redirected to the PXC for additional processing, points (2) and (3). From there, traffic is sent by the egress forwarding complex 2 out of the node, at point (4).

Similar logic can be followed in the downstream (opposite) direction where the traffic enters the ingress forwarding complex 2 at point (1'), it is redirected to the same PXC on forwarding complex 3 and exists the node on forwarding complex 1 at point (4').

In this example with the maximum loopback bandwidth of 100 Gb/s, port-schedulers under the PXC egress subports must be configured to support their respective anticipated bandwidth in each direction (20 Gb/s upstream and 80 Gb/s downstream), for the total bandwidth of 100 Gb/s supported on the cross-connect.

Figure 25: Bandwidth management on PXC sub-ports



Traffic traversing PXC contains an overhead of 4 bytes per packet that are attributed to the internal VLAN tag used for PXC sub-port identification within the SR node. However, these 4 bytes are not accounted for in the configured QoS rates. Therefore, the operator should take this into consideration when configuring rates on QoS objects under PXC ports.

2.5.6.3 Queue allocation on PXC sub-ports

PXC sub-ports are auto-configured in hybrid mode and this cannot be changed by configuration. The PXC sub-ports each have a set of queues on the network egress side and a set of queues on the access egress and ingress (per SAP or ESM subscriber). Queues on network ingress are shared per FP or per MDA, as they are on non-PXC ports in hybrid mode.

Queue groups are allocated per PXC sub-ports.

2.5.6.4 Pool allocations on PXC ports

Queue buffers are created in buffer pools and are used for traffic buffering when queues are congested. Buffer pools are allocated per forwarding complex or per cross-connect.

Each cross-connect has three associated buffer pools:

- access ingress
- access egress
- network egress

The network ingress pool is shared between all faceplate ports on a forwarding complex. The size of the buffer pools is automatically determined by the system based on the forwarding complex type and cross-connect configuration.

2.5.7 Operational states

A port under a PXC (for example, port 1/1/1 or port 1/x1/1/m1/1), the PXC itself (PXC ID represented by the cross-connect port configuration port-xc pxc 1), and PXC sub-ports (for example, port pxc-1.a and pxc-1.b) all have administrative and operational states.

For a port-based PXC, when all layers of a PXC (PXC port, PXC ID, and PXC sub-ports) are operationally up, the faceplate port status LED on the faceplate blinks amber. The port activity LED lights green in the presence of traffic on PXC ports and turns off in the absence of traffic on PXC ports. The presence of the optical transceiver on the PXC has no effect on its operational state. Traffic cannot be sent out through the transceiver or be received through the transceiver from the outside. However, the existing traps related to insertion or removal of a transceiver (SFF Inserted/Removed) are supported. The "Signal-Fail" alarm on the PXC is suppressed.

The operational state of the PXC ID is derived from its administrative state and the operational state of the sub-ports.

The operational state of the PXC sub-ports is dependent on the operational state of the underlying port (faceplate port or MAC loopback) and the administrative state of the corresponding PXC ID.

2.5.8 PXC statistics

Two types of statistics can be collected on a regular, non-PXC Ethernet port:

- Low-level port statistics which provide information about conditions on the data-link layer and physical port, for example, the aggregate number of forwarded and dropped octets or bytes on the data-link layer (Layer 2 MAC), FCS errors, number of collisions, and so on. These statistics can be viewed with the **show port port-id** command.
- Network-level statistics provide information about forwarded and dropped octets or packets on a per-queue level on network ports. These statistics can be viewed with the **show port port-id detail** command.

Statistics collection ability on the PXC port depends on whether the PXC is port-based or MAC-based (internal).

2.5.8.1 Statistics on faceplate PXC ports

The statistics on the faceplate PXC ports are maintained only on the data-link layer (Layer 2 MAC). The internal q-tag used for PXC sub-port identification within the router is included in the displayed octet count. The collected statistics represent the combined upstream and downstream traffic carried by the corresponding PXC sub-ports.

For example, in port level statistics output for a faceplate PXC port, the output count represents the upstream and downstream traffic flowing out of the faceplate port while the input count represents the same looped traffic returning into the same port.

```

=====
Traffic Statistics
=====
-----
                                Input          Output
-----
Octets                          290164703      290164703
Packets                          2712661        2712661
    
```

Errors	0	0
--------	---	---

Statistics are cleared when a faceplate port is added or removed from the PXC.

Statistics collection to a local file is not supported for faceplate PXC ports.

Queues are not instantiated on the faceplate PXC ports, therefore, the network level (queue) statistics are not maintained in that location.

2.5.8.2 Statistics collection on internal (MAC-based) PXC

Internal ports created in the MAC chip (for example port 1/x1/1/m1/1) do not have Layer 2 MAC addresses, therefore, statistics based on the data-link layer (Layer 2 MAC) are not available.

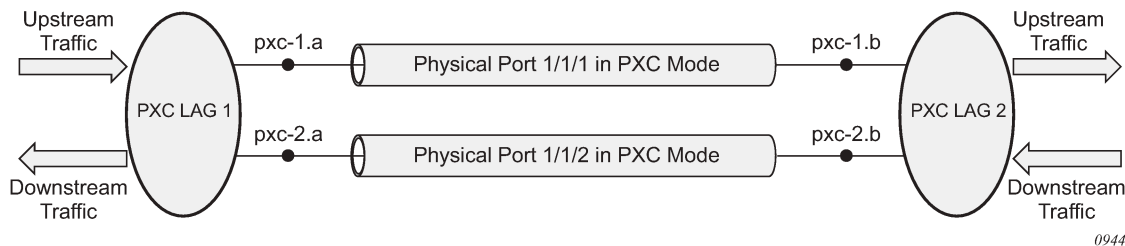
2.5.8.3 Statistics collection on PXC sub-ports

PXC sub-ports provide only network-level statistics (queue statistics). MAC related statistics are not available on PXC sub-ports.

2.5.9 PXC LAG

PXC sub-ports can be aggregated into a PXC LAG for increased capacity and card redundancy. A logical concept of a PXC LAG is shown in [Figure 26: Logical concept of a LAG on PXC ports](#).

Figure 26: Logical concept of a LAG on PXC ports



Although the configuration allows for a mix of port-based PXC and MAC-based PXC in a LAG, the configuration should be used in a production network only during a short migration period when transitioning from one type of PXC to the other. Outside of the migration, the PXC in a LAG should be of the same type, for example, a LAG should contain only port-based PXC or only MAC-based PXC but not both.

The LAGs on PXC ports must be configured in pairs as shown in this example:

MD-CLI commands

```
configure {
  lag 1 {
    description "lag in the up direction"
    port pxc-1.a {
    }
    port pxc-2.a {
    }
  }
}
```

```

lag 2 {
  description "lag in the down direction"
  port pxc-1.b {
  }
  port pxc-2.b {
  }
}

```

Within the router, the two sides of the PXC LAG (LAG 1 and LAG 2 in the example configuration) are not aware of their interconnection. As a result, the operational state of one side of the PXC LAG is not influenced by the state of the PXC LAG on the other side.

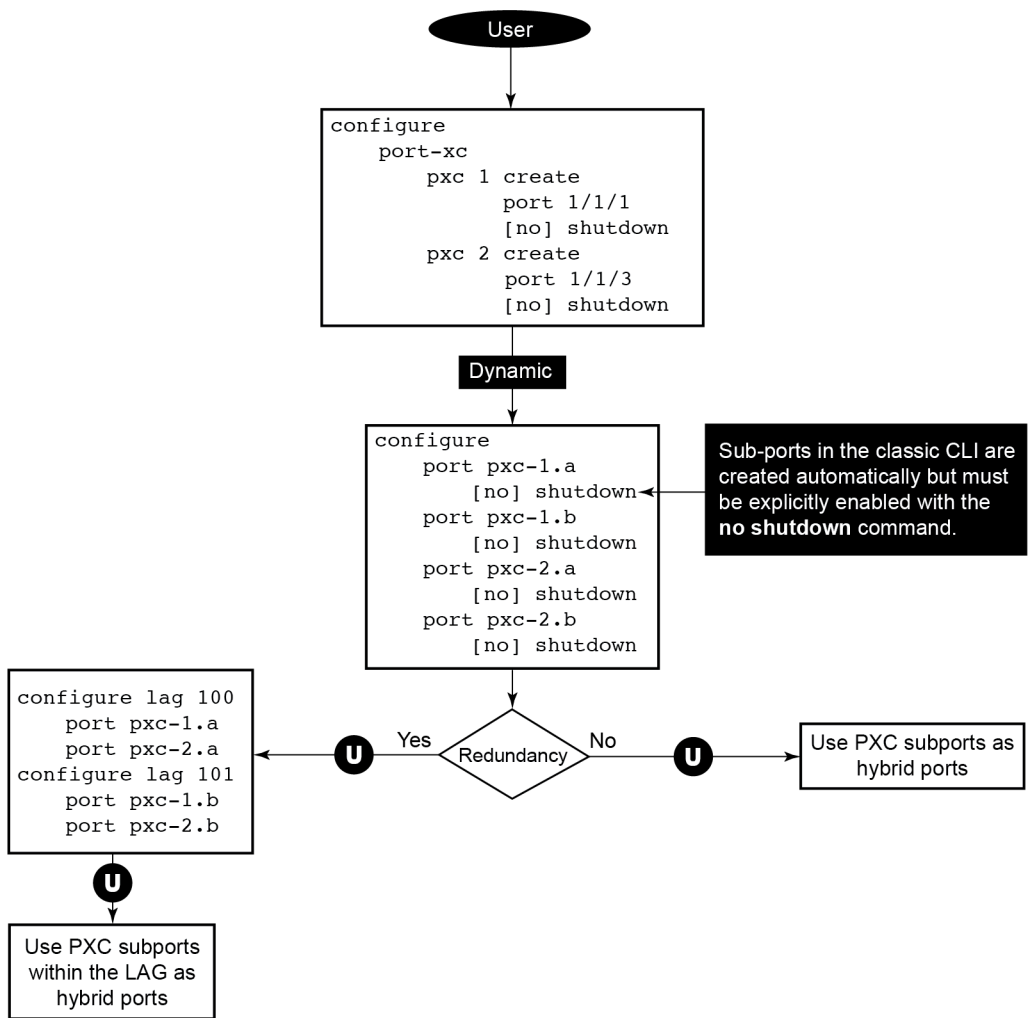
PXC sub-ports in a LAG must have the same properties (such as the same speed). Mixing PXC sub-ports and non-PXC ports is not allowed. The first port added to a LAG determines the type of LAG (PXC or non-PXC).

Statistics in the output of the **show lag id statistics** command represent combined traffic carried over the referenced **lag id** and its pair (lag 1 and lag 2 in the above example).

2.5.10 Basic PXC provisioning

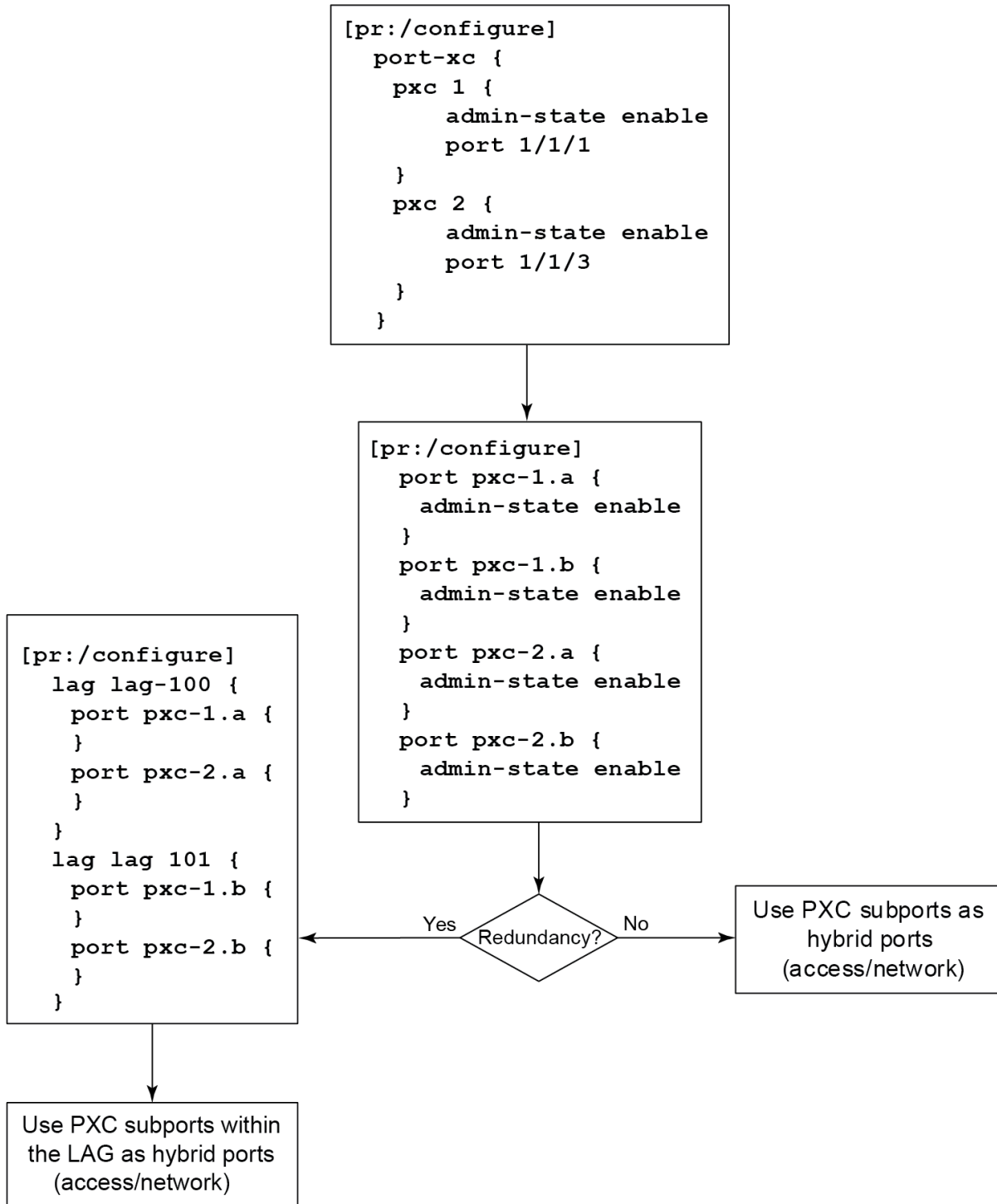
The CLI configuration flow example shown in [Figure 27: Classic CLI flow](#) represents a PXC configuration based on the faceplate port. The oval marked "User" represents a configuration step that the user must perform. The block marked "Dynamic" represents a step that the system performs automatically without a user's assistance.

Figure 27: Classic CLI flow



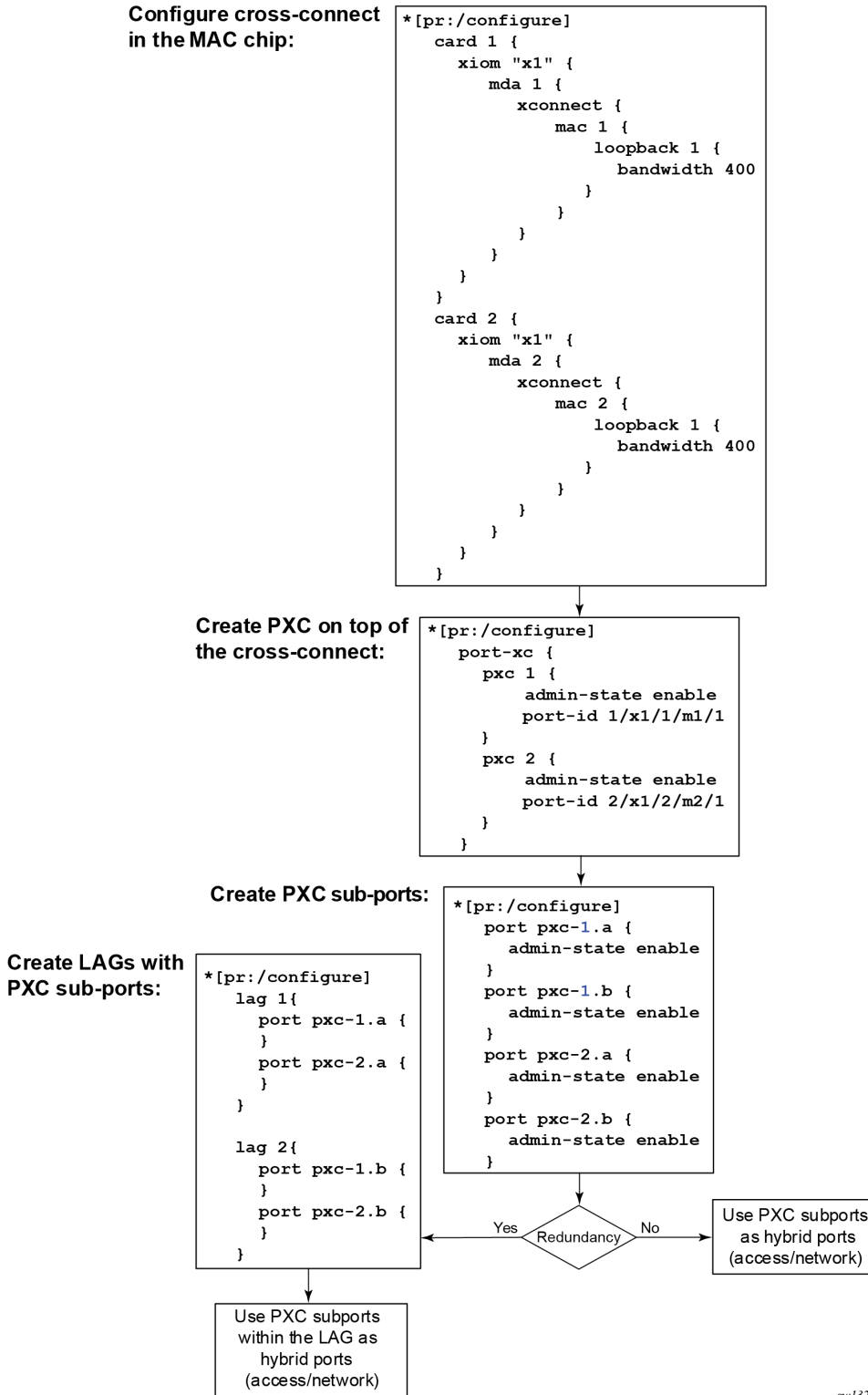
0945

Figure 28: MD-CLI flow



sw1326

Figure 29: MD-CLI PXC configuration on internal loopback



sw1327

2.5.11 PXC mirroring and LI

Traffic on a PXC sub-port can be mirrored or lawfully intercepted (LI). For example, subscriber "Annex1" traffic arriving on a PXC sub-port is mirrored if "Annex1" is configured as a mirror or LI source. A PXC sub-port can also be used to transmit mirror and LI traffic out from a mirror-destination service (such as a mirror-dest SAP or SDP can egress out a PXC sub-port, or a routable LI encapsulated packet can be forwarded and transmitted out a PXC sub-port).

A mirror destination can be configured to transmit mirrored and LI traffic out of a SAP on a PXC sub-port that is then cross connected into a VPLS service where a VXLAN encapsulation is added to the mirrored packets before transmission out of the node.

The internal q-tag that represents the PXC sub-port within the system is included in the lawfully intercepted copy of the packet for traffic intercepted (mirrored) on the ingress side of a PXC sub-port ,when the associate mirror-dest service is of type **ether** (the default) with routable lawful interception encapsulation (**config>mirror-dest>encap**).

See the *7450 ESS, 7750 SR, 7950 XRS, and VSR OAM and Diagnostics Guide* for information about LI.

2.5.12 Multi-chassis redundancy

Multi-Chassis Synchronization (MCS) configuration (**config>redundancy>multi-chassis>peer>sync**) is supported for entities using PXC. However, MC-LAG is not supported directly on PXC because PXC are not directly connected to external equipment.

2.5.13 Health monitoring on the PXC

Health monitoring of the PXC sub-ports is based on EFM OAM where the Information OAMPDUs are transmitted by each peer (pxc sub-port) at the configured intervals. Their purpose is to perform keepalive and critical notification functions.

For PXC with underlying faceplate ports, status monitoring can be enabled for either or both of the faceplate ports.

- **crc-monitoring** (link quality) on the RX side of the port (**config>port>ethernet>crc-monitor**)
- **crc-monitoring** (link quality) on the path from IOM toward MDA (**config>port>ethernet>down-on-internal-error**)

The **tx-disable** flag (disable remote laser on error) is not supported on PXC ports because PXC ports are looped.

CRC monitoring on the RX side of the faceplate ports has the following characteristics:

- monitor ingress error conditions
- compare error counts against configurable thresholds
- CRC errors are only recorded if frames are transmitted
- crossing the signal degrade (SD) threshold raises log event
- crossing the signal failure (SF) threshold takes down the port's operational state
- error rate thresholds uses format m·10-n; both the threshold (n) and multiplier (m) are configurable

Health monitoring on the faceplate ports level is disabled by default.

In addition to the explicitly configured aforementioned health monitoring mechanisms, PXC operational state transitions are by default reported by a port UP/DOWN trap:

```
478 2015/10/22 14:08:15.86 UTC WARNING: SNMP #2004 Base pxc-1.b Interface pxc-1.b is not operational
```

```
478 2015/10/22 14:08:15.86 UTC WARNING: SNMP #2004 Base pxc-1.b Interface pxc-1.b is operational
```

2.5.14 Configuration example

The following example is based on a PXC on a faceplate port. Subscriber traffic with QinQ encapsulation arriving on two different line cards (3 and 4) is terminated on the PXC LAG on line cards 1 and 2. With this method, if one of the ingress line cards (3 or 4) fails, the subscriber traffic remains unaffected (continues to be terminated on line cards 1 and 2) provided that the correct protection mechanism is implemented in the access part of the network. This protection mechanism in the access part of the network must ensure that traffic arriving on card 3 can be rerouted to card 4 if card 3 fails. The opposite must be true as well (the path to card 4 must be protected by a path to card 3).

PXC can be on any card, independent of ingress ports.

The following is an example of faceplate (physical) port configuration on cards 3 and 4:

MD-CLI commands

```
[pr:/configure]
port 3/1/1 {
    description "access I/O port on card 3; ecap is null which means that all
    VLAN tagged and untagged traffic will be accepted"
    ethernet {
        mode access
        encap-type null
    }
}
port 4/1/1 {
    description "access I/O port on card 4; ecap is null which means that all
    VLAN tagged and untagged traffic will be accepted"
    ethernet {
        mode access
        encap-type null
    }
}
```

The following is an example of a PXC configuration on cards 1 and 2:

MD-CLI commands

```
[pr:/configure]
port-xc {
    pxc 1 create {
        admin-state enable
        description "PXC on card 1"
        port 1/1/1
    }
    pxc 2 create {
        admin-state enable
        description "PXC on card 2"
        port 2/1/1
    }
}
```

The user must manually configure the sub-port encapsulation (the default is dot1q). PXC sub-ports transparently pass traffic with preserved QinQ tags from the .b side of the PXC to the .a side of the PXC where a *.* capture SAP is configured.

MD-CLI commands

```
[pr:/configure]
port pxc-1.a {
    admin-state enable
    description "termination PXC side; *.* capture SAP will be configured here"
    encap-type qinq
}

port pxc-1.b {
    admin-state enable
    description "transit PXC side; all VLAN tags (*) will be transparently
    passed via this side"
    encap-type dot1q
}

port pxc-2.a
    admin-state enable
    description "together with pxc-1.a, this sub-port is a member of LAG 1"
    encap-type qinq
}

port pxc-2.b {
    admin-state enable
    description "together with pxc-1.b, this sub-port is a member of LAG 2"
    encap-type dot1q
}
```

The following is an example of a PXC LAG configuration:

MD-CLI commands

```
[pr:/configure]
lag 1 {
    admin-state enable
    description "terminating side of the cross-connect"
    port pxc-1.a {
    }
    port pxc-2.a {
    }
}
lag 2 {
    admin-state enable
    description "transient side of the cross-connect"
    port pxc-1.b {
    }
    port pxc-2.b {
    }
}
```

Passing traffic from the ingress side on access (ports 3/1/1 and 4/1/1) via the transient PXC sub-ports pxc-1.b and pxc-2.b to the termination side of the PXC is performed through VPLS.

MD-CLI commands

```
[pr:/configure]
service vpls 1 {
    admin-state enable
}
```

```

customer 1
description "stitching access side to the anchor"
split-horizon-group "access (I/O) side" {
}
sap 3/1/1 {
admin-state enable
description "I/O port"
split-horizon-group "access"
}

sap 4/1/1 {
admin-state enable
description "I/O port"
split-horizon-group "access"
}
sap lag-2:* {
admin-state enable
description "transit side od PXC"
}
}

```

The following is an example of capture SAPs on the anchor:

MD-CLI commands

```

[pr:/configure]
service vpls 3 {
admin-state enable
customer 1
description "VPLS with capture SAPs"
capture-sap lag-1:10.* {
admin-state enable
description "termination side of PXC; traffic with S-tag=10 will be extracted here"
trigger-packet {
dhcp true
dhcpv6 true
pppoe true
}
}

capture-sap lag-1:11.* {
admin-state enable
description "termination side of PXC; traffic with S-tag=11 will be extracted
here".
}

```

2.6 FPE

Certain applications in the SR OS require extra traffic processing in the forwarding plane. Such additional traffic processing is facilitated by an internal cross-connect that uses PXC ports (described in the [Port Cross-Connect](#)). Application-specific use of the cross-connect is built on the common premise that the traffic must be steered from the input ports to the PXC ports where the traffic can be looped for additional processing in the forwarding plane. To shield the user from the intricacies involved when configuring application-specific cross-connect attributes, a CLI construct referred to as Forwarding Path Extensions (FPE) simplifies provisioning of various applications which rely on PXC functionality. The following are examples of applications that rely on PXC and FPE:

- anchored PW-ports where PW payload termination in Layer 3 services is disjointed from I/O ports in the system

- VXLAN termination on non-system IPv4 addresses and VXLAN IPv6 underlay
- origination or termination of a service with an SRv6 tunnel
- GTP-U tunnel termination for Fixed Wireless Access (FWA)

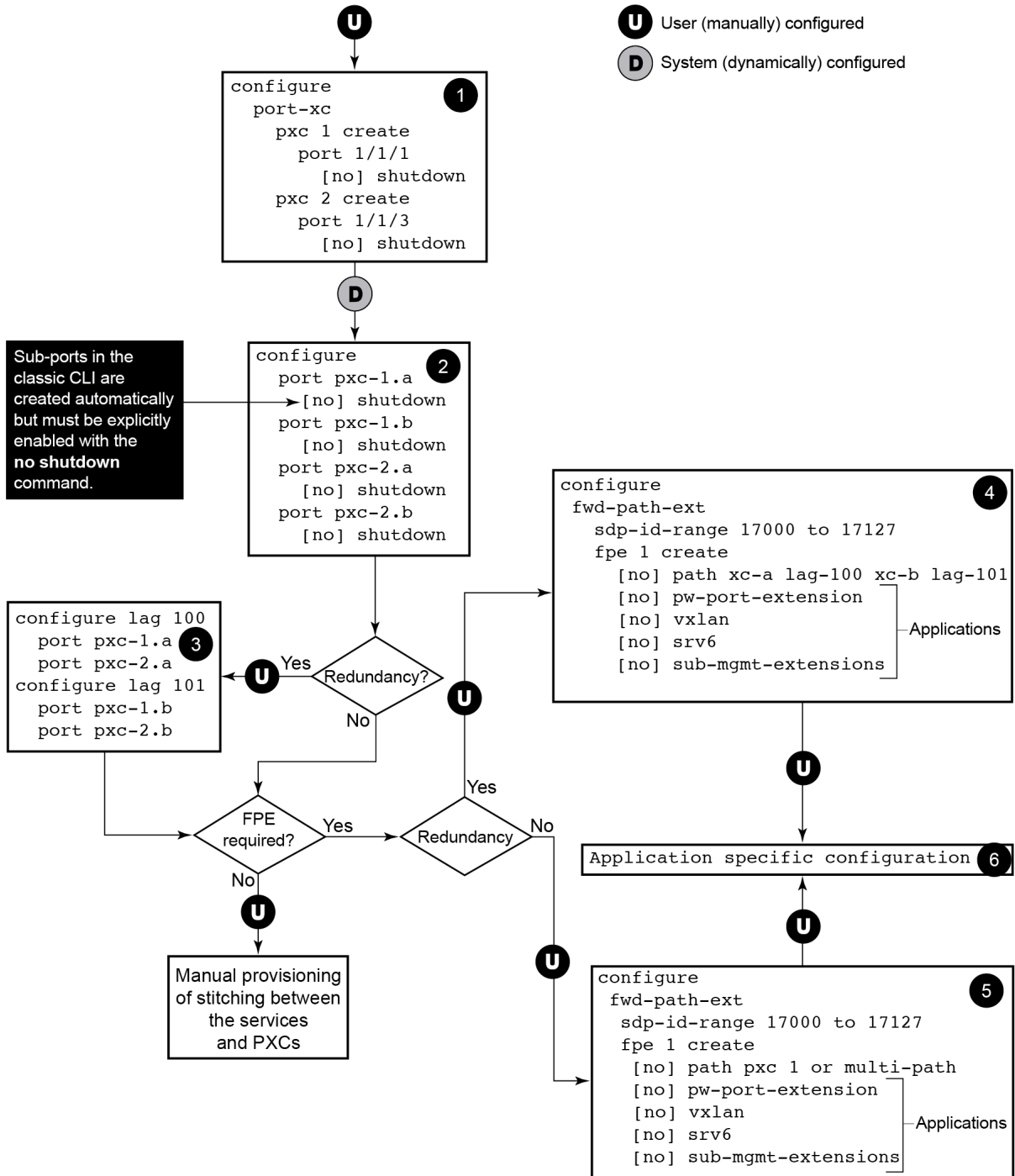
Application-specific uses of PXC ports and FPEs are described in the respective service guides which may include, but are not limited to the *7450 ESS, 7750 SR, and VSR Triple Play Service Delivery Architecture Guide* and *7450 ESS, 7750 SR, 7950 XRS, and VSR Layer 2 Services and EVPN Guide*.

The FPE configuration provides information to the SR OS node necessary to associate the application with the PXC (paired PXC sub-ports, multipath PXC sub-ports, or PXC based LAG IDs). Consequently, the SR OS node sets up the internal logic using PXC as required by the application.

An example of FPE provisioning in classic and MD-CLI are provided in [Figure 30: FPE – classic CLI provisioning steps](#) and [Figure 31: FPE – MD-CLI provisioning steps](#).

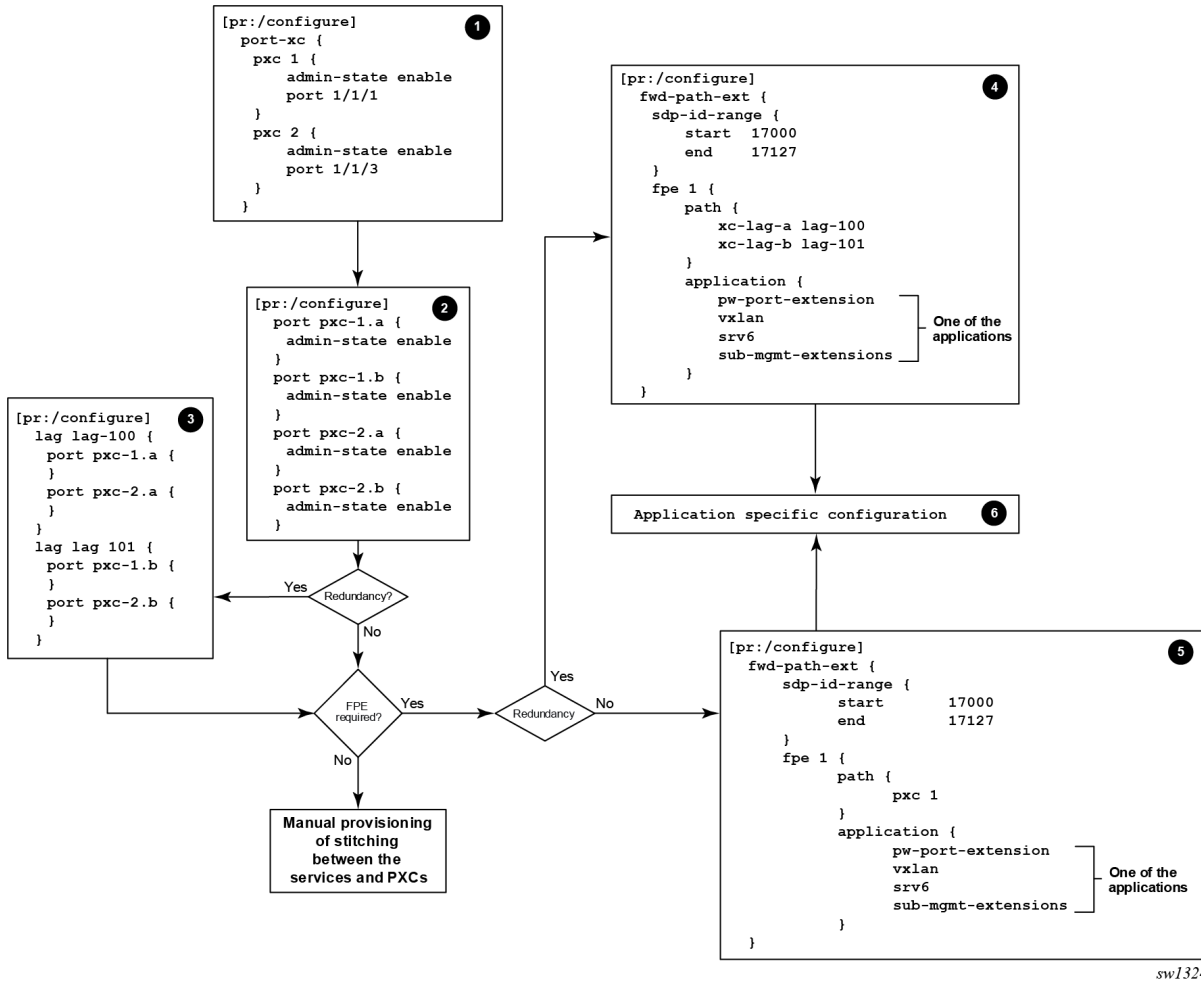
- The first three steps in the classic CLI example are applicable to PXC port provisioning. In MD-CLI, the user must explicitly create sub-ports as described in [Port Cross-Connect](#).
- Association between the application and the PXC is performed in steps 4 and 5. These applications require internal configuration of SDPs and their IDs are allocated from the user configurable range. To prevent conflict between the user-provisioned SDP IDs and internally configured SDP IDs in FPE case, a range of SDP IDs that are used by FPE is reserved by the **sdp-id-range** commands under **config>fwd-path-ext**.
- Application-specific configuration is performed in step 6, partially by the user and partially by the system. This is described in the application-specific user guides.

Figure 30: FPE – classic CLI provisioning steps



No3479

Figure 31: FPE – MD-CLI provisioning steps



sw1324

After the PXC sub-port or LAG is associated with an FPE object, the user cannot use CLI to create IP interfaces manually and SAPs under these PXC sub-ports or LAGs. Only the SR OS system is allowed to reference these PXC sub-ports or LAGs in internal IP interfaces and SAPs, as required by each application.

The user can modify PXC sub-port and LAG parameters (QoS, LAG profiles, and so on). To remove PXC sub-ports or LAGs from the FPE object, they must not be associated with an application.

Multipath FPE

Some applications require load balancing over multiple PXC ports or LAGs of PXC ports; for example, the subscriber management application has limited forwarding and QoS resources per line card. To optimize resource usage, Nokia recommends load-balance sessions over multiple line cards. However, the subscriber-management application cannot use a LAG of PXC's for load balancing because a LAG reserves subscriber-management resources on each line card that has LAG members. To solve this issue, use the **configure fwd-path-ext fpe multi-path** command to configure an FPE with multiple paths, each of which can be a PXC port or a LAG of PXC ports.

Support for multipath FPEs is application specific.

2.7 LAG

Based on the IEEE 802.1ax standard (formerly 802.3ad), Link Aggregation Groups (LAGs) can be configured to increase the bandwidth available between two network devices, depending on the number of links installed. LAG also provides redundancy if one or more links participating in the LAG fail. All physical links in a specific LAG links combine to form one logical interface.

Packet sequencing must be maintained for any session. The hashing algorithm deployed by the Nokia routers is based on the type of traffic transported to ensure that all traffic in a flow remains in sequence while providing effective load sharing across the links in the LAG.

LAGs must be statically configured or formed dynamically with Link Aggregation Control Protocol (LACP). The optional marker protocol described in IEEE 802.1ax is not implemented. LAGs can be configured on network and access ports.

The LAG load sharing is executed in hardware, which provides line rate forwarding for all port types.

The LAG implementation supports LAG with all member ports of the same speed and LAG with mixed port-speed members (see the sections that follow for details).

2.7.1 LACP

Under normal operation, all non-failing links in a LAG become active and traffic is load balanced across all active links. In some circumstances, however, this is not wanted. Instead, it may be needed that only some of the links are active (for example, all links on the same IOM) and the other links be kept in stand-by condition.

LACP enhancements allow active LAG-member selection based on particular constrains. The mechanism is based on the IEEE 802.1ax standard so interoperability is ensured.

To use LACP on a LAG, operator must enable LACP on the LAG including, if wanted, selecting non-default LACP mode: active/passive and configuring administrative key to be used (**config>lag lacp**). In addition, an operator can configure a needed LACP transmit interval (**config>lag lacp-xmit-interval**).

When LACP is enabled, an operator can see LACP changes through traps and log messages logged against the LAG. See the TIMETRA-LAG-MIB.mib for more details.

2.7.1.1 LACP multiplexing

The router supports two modes of multiplexing RX/TX control for LACP: coupled and independent.

In coupled mode (default), both RX and TX are enabled or disabled at the same time whenever a port is added or removed from a LAG group.

In independent mode, RX is first enabled when a link state is UP. LACP sends an indication to the far-end that it is ready to receive traffic. Upon the reception of this indication, the far-end system can enable TX. Therefore, in independent RX/TX control, LACP adds a link into a LAG only when it detects that the other end is ready to receive traffic. This minimizes traffic loss that may occur in coupled mode if a port is added into a LAG before notifying the far-end system or before the far-end system is ready to receive traffic. Similarly, on link removals from LAG, LACP turns off the distributing and collecting bit and informs the far-end about the state change. This allows the far-end side to stop sending traffic as soon as possible.

Independent control provides for lossless operation for unicast traffic in most scenarios when adding new members to a LAG or when removing members from a LAG. It also reduces loss for multicast and broadcast traffic.

Note that independent and coupled mode are interoperable (connected systems can have either mode set).

Independent and coupled modes are supported when using PXC ports, however, independent mode is recommended as it provides significant performance improvements.

2.7.1.2 LACP tunneling

LACP tunneling is supported on Epipe and VPLS services. In a VPLS, the Layer 2 control frames are sent out of all the SAPs configured in the VPLS. This feature should only be used when a VPLS emulates an end-to-end Epipe service (an Epipe configured using a three-point VPLS, with one access SAP and two access-uplink SAP/SDPs for redundant connectivity). The use of LACP tunneling is not recommended if the VPLS is used for multipoint connectivity. When a Layer 2 control frame is forwarded out of a dot1q SAP or a QinQ SAP, the SAP tags of the egress SAP are added to the packet.

The following SAPs can be configured for tunneling the untagged LACP frames (the corresponding protocol tunneling needs to be enabled on the port).

- If the port encapsulation is null, a null SAP can be configured on a port to tunnel these packets.
- If the port encapsulation is dot1q, either a dot1q explicit null SAP (for example, 1/1/10:0) or a dot1q default SAP (for example, 1/1/11:*) can be used to tunnel these packets.
- If the port encapsulation is QinQ, a 0.* SAP (for example, 1/1/10:0.*) can be used to tunnel these packets.

LAG port states may be impacted if LACP frames are lost because of incorrect prioritization and congestion in the network carrying the tunnel.

2.7.2 Active-standby LAG operation

An active-standby LAG provides redundancy by logically dividing LAG into subgroups. The LAG is divided into subgroups by either assigning each LAG's ports to an explicit subgroup (1 by default), or by automatically grouping all LAG's ports residing on the same line card into a unique sub-group (auto-iom) or by automatically grouping all LAG's ports residing on the same MDA into a unique sub-group (auto-mds). When a LAG is divided into sub-groups, only a single sub-group is elected as active. Which sub-group is selected depends on selection criterion chosen.

The active-standby decision for LAG member links is a local decision driven by pre-configured selection-criteria. When LACP is configured, this decision was communicated to remote system using LACP signaling.

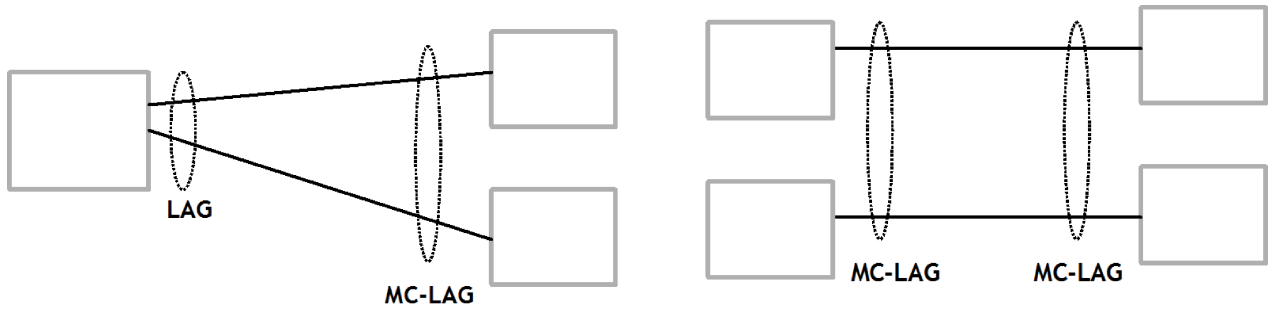
To allow non-LACP operation, an user must disable LACP on a specific LAG and select transmitter-driven standby signaling (**config>lag standby-signaling power-off**). As a consequence, the transmit laser is switched off for all LAG members in standby mode. On switch over (active-links failed) the laser is switched on all standby LAG members so they can become active.

When the power-off is selected as the standby-signaling, the selection-criteria **best-port** can be used.

It is not possible to have an active LACP in power-off mode before the correct selection criteria is selected.

Figure 32: Active-standby LAG operation without deployment examples shows how LAG in Active/Standby mode can be deployed toward a DSLAM access using sub-groups with auto-iom sub-group selection. LAG links are divided into two sub-groups (one per line card).

Figure 32: Active-standby LAG operation without deployment examples



In case of a link failure, as shown in Figure 33: LAG on access interconnection and Figure 34: LAG on access failure switchover, the switchover behavior ensures that all LAG-members connected to the same IOM as failing link become standby and LAG-members connected to other IOM become active. This way, QoS enforcement constraints are respected, while the maximum of available links is used.

Figure 33: LAG on access interconnection

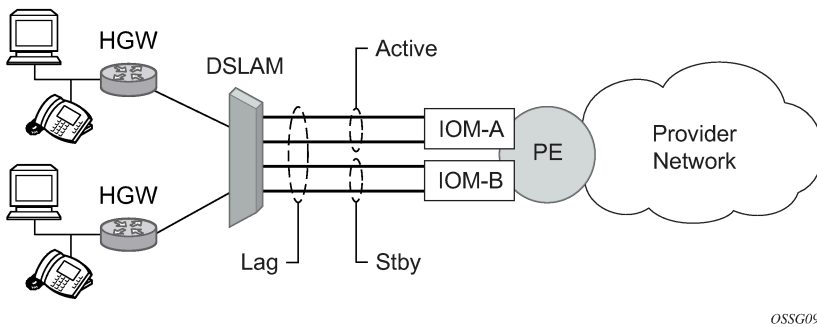
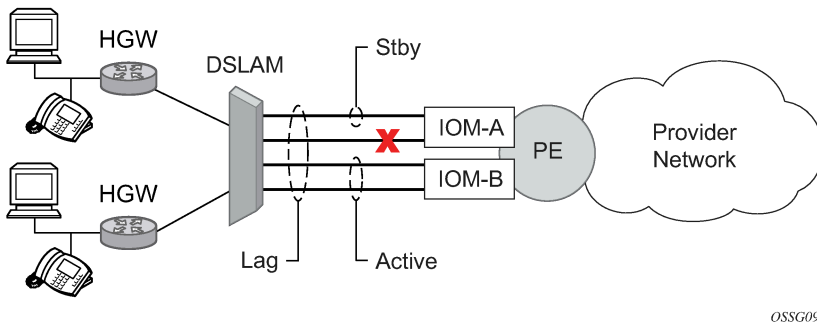


Figure 34: LAG on access failure switchover



2.7.3 LAG on access QoS consideration

The following section describes various QoS related features applicable to LAG on access.

2.7.3.1 Adapt QoS modes

Link Aggregation is supported on the access side with access or hybrid ports. Similarly to LAG on the network side, LAG on access aggregates Ethernet ports into all active or active-standby LAG. The difference with LAG on networks lies in how the QoS or H-QoS is handled. Based on hashing configured, a SAP's traffic can be sprayed on egress over multiple LAG ports or can always use a single port of a LAG. There are three user-selectable modes that allow the user to best adapt QoS configured to a LAG the SAPs are using:

- **adapt-qos distributed (default)**

In a distributed mode the SLA is divided among all line cards proportionally to the number of ports that exist on that line card for a specific LAG. For example a 100 Mb/s PIR with 2 LAG links on IOM A and 3 LAG links on IOM B would result in IOM A getting 40 Mb/s PIR and IOM B getting 60 Mb/s PIR. Because of this distribution, SLA can be enforced. The disadvantage is that a single flow is limited to IOM's share of the SLA. This mode of operation may also result in underrun because of hashing imbalance (traffic not sprayed equally over each link). This mode is best suited for services that spray traffic over all links of a LAG.

- **adapt-qos link**

In a link mode the SLA is provided to each port of a LAG. With the example above, each port would get 100 Mb/s PIR. The advantage of this method is that a single flow can now achieve the full SLA. The disadvantage is that the overall SLA can be exceeded, if the flows span multiple ports. This mode is best suited for services that are guaranteed to hash to a single egress port.

- **adapt-qos port-fair**

Port-fair distributes the SLA across multiple line cards relative to the number of active LAG ports per card (in a similar way to distribute mode) with all LAG QoS objects parented to scheduler instances at the physical port level (in a similar way to link mode). This provides a fair distribution of bandwidth between cards and ports whilst ensuring that the port bandwidth is not exceeded. Optimal LAG utilization relies on an even hash spraying of traffic to maximize the use of the schedulers' and ports' bandwidth. With the example above, enabling port-fair would result in all five ports getting 20 Mb/s.

When port-fair mode is enabled, per-Vport hashing is automatically disabled for subscriber traffic such that traffic sent to the Vport no longer uses the Vport as part of the hashing algorithm. Any QoS object for subscribers, and any QoS object for SAPs with explicitly configured hashing to a single egress LAG port, are given the full bandwidth configured for each object (in a similar way to link mode). A Vport used together with an egress port scheduler is supported with a LAG in port-fair mode, whereas it is not supported with a distribute mode LAG.

- **adapt-qos distributed include-egr-hash-cfg**

This mode can be considered a mix of link and distributed mode. The mode uses the configured hashing for LAG/SAP/service to choose either link or distributed adapt-qos modes. The mode allows:

- SLA enforcement for SAPs that through configuration are guaranteed to hash to a single egress link using full QoS per port (as per link mode)
- SLA enforcement for SAPs that hash to all LAG links proportional distribution of QoS SLA amongst the line cards (as per distributed mode)

- SLA enforcement for multi service sites (MSS) that contain any SAPs regardless of their hash configuration using proportional distribution of QoS SLA amongst the line cards (as per distributed mode)

The following restrictions apply to `adapt-qos distributed include-egr-hash-cfg`:

- LAG mode must be access or hybrid.
- The user cannot change from **`adapt-qos distribute include-egr-hash-cfg`** to **`adapt-qos distribute`** when `link-map-profiles` or `per-link-hash` is configured.
- The user cannot change from **`adapt-qos link`** to **`adapt-qos distribute include-egr-hash-cfg`** on a LAG with any configuration.

Table 24: Adapt QoS bandwidth/rate distribution shows examples of rate/BW distributions based on the **`adapt-qos`** mode used.

Table 24: Adapt QoS bandwidth/rate distribution

	distribute	link	port-fair	distribute include-egr-hash-cfg
SAP Queues	% # local links ³	100% rate	100% rate (SAP hash to one link) or %# all links ⁴ (SAP hash to all links)	100% rate (SAP hash to one link) or % # local linksa (SAP hash to all links)
SAP Scheduler	% # local linksa	100% bandwidth	100% rate (SAP hash to one link) or %# all linksb (SAP hash to all links)	100% bandwidth (SAP hash to a one link) or % # local linksa (SAP hash to all links)
SAP MSS Scheduler	% # local linksa	100% bandwidth	% # local linksa	% # local linksa

2.7.3.2 Per-fp-ing-queuing

Per-fp-ing-queuing optimization for LAG ports provides the ability to reduce the number of hardware queues assigned on each LAG SAP on ingress when the flag at LAG level is set for per-fp-ing-queuing.

When the feature is enabled in the `config>lag>access` context, the queue allocation for SAPs on a LAG are optimized and only one queuing set per ingress forwarding path (FP) is allocated instead of one per port.

The following rules apply for configuring the per-fp-ing-queuing at LAG level:

- To enable per-fp-ing-queuing, the LAG must be in access mode

³ % # local links = X * (number of local LAG members on a line card/ total number of LAG members)

⁴ %# all links = X * (link speed)/(total LAG speed)

- The LAG mode cannot be set to network mode when the feature is enabled
- Per-fp-ing-queuing can only be set if no port members exists in the LAG

2.7.3.3 Per-fp-egr-queuing

Per-fp-egr-queuing optimization for LAG ports provides the ability to reduce the number of egress resources consumed by each SAP on a LAG, and by any encap groups that exist on those SAPs.

When the feature is enabled in the **config>lag>access** context, the queue and virtual scheduler allocation are optimized. Only one queuing set and one H-QoS virtual scheduler tree per SAP/encap group is allocated per egress forwarding path (FP) instead of one set per each port of the LAG. In case of a link failure/recovery, egress traffic uses failover queues while the queues are moved over to a newly active link.

Per-fp-egr-queuing can be enabled on existing LAG with services as long as the following conditions are met.

- The mode of the LAG must be **access** or **hybrid**.
- The port-type of the LAGs must be **standard**.
- The LAG must have either **per-link-hash** enabled or all SAPs on the LAG must use **per-service-hashing** only and be of a type: VPLS SAP, i-VPLS SAP, or e-Pipe VLL or PBB SAP.

To disable per-fp-egr-queuing, all ports must first be removed from a specific LAG.

2.7.3.4 Per-fp-sap-instance

Per-fp-sap-instance optimization for LAG ports provides the ability to reduce the number of SAP instance resources consumed by each SAP on a lag.

When the feature is enabled, in the **config>lag>access** context, a single SAP instance is allocated on ingress and on egress per each forwarding path instead of one per port. Thanks to an optimized resource allocation, the SAP scale on a line card increases, if a LAG has more than one port on that line card. Because SAP instances are only allocated per forwarding path complex, hardware reprogramming must take place when as result of LAG links going down or up, a SAP is moved from one LAG port on a specific line card to another port on a specific line card within the same forwarding complex. This results in an increased data outage when compared to per-fp-sap-instance feature being disabled. During the reprogramming, failover queues are used when SAP queues are reprogrammed to a new port. Any traffic using failover queues is not accounted for in SAPs statistics and is processed at best-effort priority.

The following rules apply when configuring a per-fp-sap-instance on a LAG:

- Per-fp-ing-queuing and per-fp-egr-queuing must be enabled.
- The functionality can be enabled/disabled on LAG with no member ports only. Services can be configured.

Other restrictions:

- SAP instance optimization applies to LAG-level. Whether a LAG is sub-divided into sub-groups or not, the resources are allocated per forwarding path for all complexes LAG's links are configured on (that is irrespective of whether a sub-group a SAP is configured on uses that complex or not).
- Egress statistics continue to be returned per port when SAP instance optimization is enabled. If a LAG links are on a single forwarding complex, all ports but one have no change in statistics for the last interval – unless a SAP moved between ports during the interval.

- Rollback that changes per-fp-sap-instance configuration is service impacting.

2.7.4 Traffic load balancing options

When a requirement exists to increase the available bandwidth for a logical link that exceeds the physical bandwidth or add redundancy for a physical link, typically one of two methods is applied: equal cost multi-path (ECMP) or Link Aggregation (LAG). A system can deploy both at the same time using ECMP of two or more Link Aggregation Groups (LAG) and, or single links.

Different types of hashing algorithms can be employed to achieve one of the following objectives:

- ECMP and LAG load balancing should be influenced solely by the offered flow packet. This is referred to as *per-flow* hashing.
- ECMP and LAG load balancing should maintain consistent forwarding within a specific service. This is achieved using *consistent per-service* hashing.
- LAG load balancing should maintain consistent forwarding on egress over a single LAG port for a specific network interface, SAP, and so on. This is referred as *per link* hashing (including explicit per-link hashing with LAG link map profiles). Note that if multiple ECMP paths use a LAG with per-link hashing, the ECMP load balancing is done using either *per flow* or *consistent per service* hashing.

These hashing methods are described in the following subsections. Although multiple hashing options may be configured for a specific flow at the same time, only one method is selected to hash the traffic based on the following decreasing priority order:

For ECMP load balancing:

1. Consistent per-service hashing
2. Per-flow hashing

For LAG load balancing:

1. LAG link map profile
2. Per-link hash
3. Consistent per-service hashing
4. Per-flow hashing

2.7.4.1 Per-flow hashing

Per-flow hashing uses information in a packet as an input to the hash function ensuring that any specific flow maps to the same egress LAG port/ECMP path. Note that because the hash uses information in the packet, traffic for the same SAP/interface may be sprayed across different ports of a LAG or different ECMP paths. If this is not wanted, other hashing methods described in this section can be used to change that behavior. Depending on the type of traffic that needs to be distributed into an ECMP and, or LAG, different variables are used as input to the hashing algorithm that determines the next hop selection. The following describes default per-flow hashing behavior for those different types of traffic:

- VPLS known unicast traffic is hashed based on the IP source and destination addresses for IP traffic, or the MAC source and destination addresses for non-IP traffic. The MAC SA/DA are hashed and then, if the Ethertype is IPv4 or IPv6, the hash is replaced with one based on the IP source address/destination address.
- VPLS multicast, broadcast and unknown unicast traffic.

- Traffic transmitted on SAPs is not sprayed on a per-frame basis, but instead, the service ID selects ECMP and LAG paths statically.
- Traffic transmitted on SDPs is hashed on a per packet basis in the same way as VPLS unicast traffic. However, per packet hashing is applicable only to the distribution of traffic over LAG ports, as the ECMP path is still chosen statically based on the service ID.

Data is hashed twice to get the ECMP path. If LAG and ECMP are performed on the same frame, the data is hashed again to get the LAG port (three hashes for LAG). However, if only LAG is performed, then hashing is only performed twice to get the LAG port.

- Multicast traffic transmitted on SAPs with IGMP snooping enabled is load-balanced based on the internal multicast ID, which is unique for every (s,g) record. This way, multicast traffic pertaining to different streams is distributed across different LAG member ports.
 - The hashing procedure that used to be applied for all VPLS BUM traffic would result in PBB BUM traffic being sent out on BVPLS SAP to follow only a single link when MMRP was not used. Therefore, traffic flooded out on egress BVPLS SAPs is now load spread using the algorithm described above for VPLS known unicast.
- Unicast IP traffic routed by a router is hashed using the IP SA/DA in the packet.
 - MPLS packet hashing at an LSR is based on the whole label stack, along with the incoming port and system IP address. Note that the EXP/TTL information in each label is not included in the hash algorithm. This method is referred to as *Label-Only Hash* option and is enabled by default, or can be re-instated in CLI by entering the *lbl-only* option. A few options to further hash on the headers in the payload of the MPLS packet are also provided. For more details, see [Changing default per-flow hashing inputs](#).
 - VLL traffic from a service access point is not sprayed on a per-packet basis, but as for VPLS flooded traffic, the service ID selects one of the ECMP/LAG paths. The exception to this is when shared-queuing is configured on an Epipe SAP, or lpipe SAP, or when H-POL is configured on an Epipe SAP. In those cases, traffic spraying is the same as for VPLS known unicast traffic. Packets of the above VLL services received on a spoke SDP are sprayed the same as for VPLS known unicast traffic.
 - Note that Cpipe VLL packets are always sprayed based on the service-id in both directions.
 - Multicast IP traffic is hashed based on an internal multicast ID, which is unique for every record similar to VPLS multicast traffic with IGMP snooping enabled.

If the ECMP index results in the selection of a LAG as the next hop, then the hash result is hashed again and the result of the second hash is input to the modulo like operation to determine the LAG port selection.

When the ECMP set includes an IP interface configured on a spoke SDP (IES/VP RN spoke interface), or a Routed VPLS spoke SDP interface, the unicast IP packets—which is sprayed over this interface—is not further sprayed over multiple RSVP LSPs/LDP FEC (part of the same SDP), or GRE SDP ECMP paths. In this case, a single RSVP LSP, LDP FEC next-hop or GRE SDP ECMP path is selected based on a modulo operation of the service ID. In case the ECMP path selected is a LAG, the second round of the hash, hashes traffic based on the system, port or interface load-balancing settings.

In addition to the above described per-flow hashing inputs, the system supports multiple options to modify default hash inputs.

2.7.4.1.1 Changing default per-flow hashing inputs

For some traffic patterns or specific deployments, per-flow hashing is needed but the hashing result using default hash inputs as described above may not produce the wanted distribution. To alleviate this issue, the system allows users to modify the load balancing algorithm at the system, interface or port level.

2.7.4.1.1.1 LSR hashing

By default, the LSR hash routine operates on the label stack only. However, there is also the ability to hash on the IP header if a packet is IP. An LSR considers a packet to be IP if the first nibble following the bottom of the label stack is either 4 (IPv4) or 6 (IPv6). This allows the user to include an IP header in the hashing routine at an LSR for the purpose of spraying labeled IP packets over multiple equal cost paths in ECMP in an LSP and, or over multiple links of a LAG group in all types of LSPs.

The user enables the LSR hashing on label stack and, or IP header by entering the following system-wide command: **config>system>load-balancing>lsrc-load-balancing [lbi-only | lbi-ip | ip-only]**

By default, the LSR falls back to the hashing on label stack only. This option is referred to as lbi-only and the user can revert to this behavior by entering one of the two commands:

config>system>load-balancing>lsrc-load-balancing lbi-only

config>system>load-balancing>no lsrc-load-balancing

The user can also selectively enable or disable the inclusion of label stack and IP header in the LSR hash routine for packets received on a specific network interface by entering the following command:

config>router>if>load-balancing>lsrc-load-balancing [lbi-only | lbi-ip | ip-only | eth-encap-ip | lbi-ip-l4-teid]

This provides some control to the user such that this feature is disabled if labeled packets received on a specific interface include non IP packets that can be confused by the hash routine for IP packets. These could be VLL and VPLS packets without a PW control word.

When the user performs the **no** form of this command on an interface, the interface inherits the system level configuration.

2.7.4.1.1.2 LSR default hash routine—label-only hash option

The following is the behavior of ECMP and LAG hashing at an LSR in the existing implementation. These are performed in two rounds.

The first round starts with the ECMP hash, which consists of an initial hash based on the source port/system IP address. Each label in the stack is then hashed separately with the result of the previous hash, up to a maximum of 16 labels. The net result is used to select which LSP next-hop to send the packet to using a modulo operation of the net result with the number of next-hops.

This same net result feeds to a second round of hashing if there is LAG on the egress port where the selected LSP has its NHLFE programmed.

2.7.4.1.1.3 LSR label-IP hash option enabled

In the first hash round for ECMP, the algorithm parses down the label stack and after it reaches the bottom, it checks the next nibble. If the nibble value is 4, it assumes it is an IPv4 packet. If the nibble value is 6, it assumes it is an IPv6 packet. In both cases, the result of the label hash is fed into another hash along

with source and destination address fields in the IP packet header. Otherwise, it uses the label stack hash already calculated for the ECMP path selection.

If there are more than five labels in the stack, the algorithm also uses the result of the label hash for the ECMP path selection.

The second round of hashing for LAG re-uses the net result of the first round of hashing. This means IPv6 packets continue to be hashed on label stack only.

2.7.4.1.1.4 LSR IP-only hash option enabled

This option behaves like the label-IP hash option, except that when the algorithm reaches the bottom of the label stack in the ECMP round and finds an IP packet, it throws the outcome of the label hash and only uses the source and destination address fields in the IP packet header.

2.7.4.1.1.5 LSR Ethernet encapsulated IP hash only option enabled

This option behaves like LSR IP-only hash, except for how the IP SA/DA information is found.

After the bottom of the MPLS stack is reached, the hash algorithm verifies that what follows is an Ethernet II untagged or tagged frame. For untagged frames, the system determines the value of Ethertype at the expected packet location and checks whether it contains an Ethernet-encapsulated IPv4 (0x0800) or IPv6 (0x86DD) value. The system also supports Ethernet II tagged frames with up to two 802.1Q tags, provided that the Ethertype value for the tags is 0x8100.

When the Ethertype verification passes, the first nibble of the expected IP packet location is then verified to be 4 (IPv4) or 6 (IPv6).

2.7.4.1.1.6 LSR hashing of MPLS-over-GRE encapsulated packet

When the router removes the GRE encapsulation, pops one or more labels and then swaps a label, it acts as an LSR. The LSR hashing for packets of a MPLS-over-GRE SDP or tunnel follows a different procedure, which is enabled automatically and overrides the LSR hashing option enabled on the incoming network IP interface (**lsr-load-balancing {lbi-only | lbi-ip | ip-only | eth-encap-ip | lbi-ip-l4-teid}**).

On a packet-by-packet basis, the new hash routine parses through the label stack and the new hash routine hashes on the SA/DA fields and the Layer 4 SRC/DST Port fields of the inner IPv4/IPv6 header.



Note:

- If the GRE header and label stack sizes are such that the Layer4 SRC/DST Port fields are not read, it hashes on the SA/DA fields of the inner IPv4/IPv6 header.
- If the GRE header and label stack sizes are such that the SA/DA fields of the inner IPv4/IPv6 header are not read, it hashes on the SA/DA fields of the outer IPv4/IPv6 header.

2.7.4.1.1.7 LSR hashing when an Entropy Label is present in the packet's label stack

The LSR hashing procedures are modified as follows:

- If the **lbi-only** hashing option is enabled, or if one of the other LSR hashing options are enabled but an IPv4 or IPv6 header is not detected below the bottom of the label stack, the LSR hashes on the Entropy Label (EL) only.

- If the **lbl-ip** option is enabled, the LSR hashes on the EL and the IP headers.
- If the **ip-only** or **eth-encap-ip** is enabled, the LSR hashes on the IP headers only.

2.7.4.1.1.8 Layer 4 load balancing

Operators can enable Layer 4 load balancing to include TCP/UDP source/destination port numbers in addition to source/destination IP addresses in per-flow hashing of IP packets. By including the Layer 4 information, a SA/DA default hash flow can be sub-divided into multiple finer-granularity flows if the ports used between a specific SA/DA vary.

Layer 4 load balancing can be enabled/disabled on system and interface levels. When enabled, the extra Layer 4 port inputs apply to per-flow hashing for unicast IP traffic and multicast traffic (if **mc-enh-load-balancing** is enabled).

2.7.4.1.1.9 System IP load balancing

This enhancement adds an option to add the system IP address into the hash algorithm. This adds a per system variable so that traffic being forward through multiple routers with similar ECMP paths have a lower chance of always using the same path to a destination.

Currently, if multiple routers have the same set of ECMP next hops, traffic uses the same next hop at every router hop. This can contribute to the unbalanced utilization of links. The new hash option avoids this issue.

This feature when enabled, enhances the default per-flow hashing algorithm described earlier. It however does not apply to services which packets are hashed based on service-id or when per-service consistent hashing is enabled. This hash algorithm is only supported on IOM3-XP/IMMs or later generations of hardware. The System IP load balancing can be enabled per-system only.

2.7.4.1.1.10 TEID hash for GTP-encapsulated traffic

This options enables TEID hashing on Layer 3 interfaces. The hash algorithm identifies GTP-C or GTP-U by looking at the UDP destination port (2123 or 2152) of an IP packet to be hashed. If the value of the port matches, the packet is assumed to be GTP-U/C. For GTPv1 packets TEID value from the expected header location is then included in hash. For GTPv2 packets the TEID flag value in the expected header is additionally checked to verify whether TEID is present. If TEID is present, it is included in hash algorithm inputs. TEID is used in addition to GTP tunnel IP hash inputs: SA/DA and SPort/DPort (if Layer 4 load balancing is enabled). If a non-GTP packet is received on the GTP UDP ports above, the packets are hashed as GTP.

2.7.4.1.1.11 Source-only/destination-only hash inputs

This option allows an operator to only include source parameters or only include destination parameters in the hash for inputs that have source/destination context (such as IP address and Layer 4 port). Parameters that do not have source/destination context (such as TEID or System IP for example) are also included in hash as per applicable hash configuration. The functionality allows, among others, to ensure that both upstream and downstream traffic hash to the same ECMP path/LAG port on system egress when traffic is sent to a hair-pinned appliance (by configuring source-only hash for incoming traffic on upstream interfaces and destination-only hash for incoming traffic on downstream interfaces).

2.7.4.1.1.12 Enhanced multicast load balancing

Enhanced multicast load balancing allows users to replace the default multicast per-flow hash input (internal multicast ID) with information from the packet. When enabled, multicast traffic for Layer 3 services (such as IES, VPRN, r-VPLS) and ng-MVPN (multicast inside RSVP-TE, LDP LSPs) are hashed using information from the packet. Which inputs are chosen depends on which per-flow hash inputs options are enabled based on the following:

- **IP replication**

The hash algorithm for multicast mimics unicast hash algorithm using SA/DA by default and optionally TCP/UDP ports (Layer 4 load balancing enabled) and/or system IP (System IP load balancing enabled) and, or source/destination parameters only (Source-only/Destination-only hash inputs).

- **MPLS replication**

The hash algorithm for multicast mimics unicast hash algorithm is described in the [LSR hashing](#) section.



Note: Enhanced multicast load balancing is not supported with Layer 2 and ESM services. It is supported on all platforms except for the 7450 ESS in standard mode.

2.7.4.1.1.13 SPI load balancing

IPsec tunneled traffic transported over LAG typically falls back to IP header hashing only. For example, in LTE deployments, TEID hashing cannot be performed because of encryption, and the system performs IP-only tunnel-level hashing. Because each SPI in the IPsec header identifies a unique SA, and therefore flow, these flows can be hashed individually without impacting packet ordering. In this way, SPI load balancing provides a mechanism to improve the hashing performance of IPsec encrypted traffic.

The system allows enabling SPI hashing per Layer 3 interface (this is the incoming interface for hash on system egress)/Layer 2 VPLS service. When enabled, an SPI value from ESP/AH header is used in addition to any other IP hash input based on per-flow hash configuration: source/destination IPv6 addresses, Layer 4 source/dest ports in case NAT traversal is required (Layer 4 load balancing is enabled). If the ESP/AH header is not present in a packet received on a specific interface, the SPI is not part of the hash inputs, and the packet is hashed as per other hashing configurations. SPI hashing is not used for fragmented traffic to ensure first and subsequent fragments use the same hash inputs.

SPI hashing is supported for IPv4 and IPv6 tunnel unicast traffic and for multicast traffic (mc-enh-load-balancing must be enabled) on all platforms and requires Layer 3 interfaces or VPLS service interfaces with SPI hashing enabled to reside on IOM3-XP or newer line-cards.

2.7.4.2 Adaptive load balancing

Adaptive load balancing can be enabled per LAG to resolve traffic imbalance dynamically between LAG member ports. The following can cause traffic distribution imbalance between LAG ports:

- hashing limitations in the presence of large flows
- flow bias or service imbalance leading to more traffic over specific ports

Adaptive load balancing actively monitors the traffic rate of each LAG member port and identifies if an optimization can be made to distribute traffic more evenly between LAG ports. The traffic distribution remains flow based with packets of the same flow egressing a single port of the LAG. The traffic rate

of each LAG port is polled at regular intervals and an optimization is only executed if the adaptive load balancing tolerance threshold is reached and the minimum bandwidth of the most loaded link in the LAG exceeds the defined bandwidth threshold. The pooling interval, the tolerance, and the bandwidth threshold can be configured.

The interval for polling LAG statistics from the line cards is configurable in seconds. The system optimizes traffic distribution after two polling intervals.

The tolerance is a configurable percentage value corresponding to the difference between the most and least loaded ports in the LAG. The following formula is used to calculate the tolerance:

$$\text{Tolerance} = (\text{rate of the most loaded link} - \text{rate of the least loaded link}) / \text{rate of the most loaded link} * 100$$

Using a LAG of two ports as an example, where port A = 10 Gb/s and port B = 8 Gb/s, the difference between the most and least loaded ports in the LAG is equal to the following: $(10 - 8) / 10 * 100 = 20\%$.

The bandwidth threshold defines the minimum bandwidth threshold, expressed in percentage, of the most loaded LAG port egress before adaptive load balancing optimization is performed.



Note:

- The bandwidth threshold default value is 10% for PXC LAG and 30% for other LAG.
- Adaptive load balancing is not supported in combination with **per-link hash**, mixed-speed LAG, customized **hash-weight** values, **per-fp-egr-queuing**, **per-fp-sap-instance**, MC-LAG, or ESM.
- When using port cross-connect or satellite ports, adaptive load balancing rates and the tolerance threshold do not consider egress BUM traffic of the LAG and the adaptive load balance algorithm does not take into account **link-map-profile** traffic.
- Contact your Nokia representative for more information about scaling when:
 - More than 16 ports per LAG are used in combination with LAG ID 1 to 64 in classic CLI or **max-ports 64** in MD-CLI.
 - More than 8 ports per LAG are used in combination with LAG ID 65 to 800 in classic CLI or **max-ports 32** in MD-CLI.

The following example shows an adaptive load balancing configuration:

Example: MD-CLI

```
*[ex:/configure lag "lag-1"]
A:admin@node-2# info
  encap-type dot1q
  mode access
  adaptive-load-balancing {
    tolerance 20
  }
  port 1/1/1 {
  }
  port 1/1/2 {
  }
```

Example: classic CLI

```
*A:node-2>config>lag# info
-----
```

```

mode access
encap-type dot1q
port 1/1/1
port 1/1/2
adaptive-load-balancing tolerance 20
no shutdown
-----

```

2.7.4.3 LAG port weight speed

Nokia routers support mixing ports of different speeds in a single LAG using either **port-weight-speed** or **hash-weight**. See section [LAG port hash-weight](#) for details related to LAG port hash-weight.

The LAG **port-weight-speed** command enables mixed-speed LAG and defines the lowest port speed for a member port in that LAG as well as the type of speed ports allowed when mixed in the same LAG:

- port-weight-speed 1 supports port speed of 1GE and 10GE
- port-weight-speed 10 supports port speed of 10GE, 40GE, and 100GE

For mixed-speed LAGs:

- Both LACP and non-LACP configurations are supported. With LACP enabled, LACP is unaware of physical port differences.
- QoS is distributed proportionally to port-speed, unless explicitly configured not to do so (see **internal-scheduler-weight-mode**).
- User data traffic is hashed proportionally to port speed when any per-flow hash is deployed.
- CPM-originated OAM control traffic that requires per LAG hashing is hashed per physical port.
- Nokia recommends that users use **weight-threshold** or **hash-weight-threshold** instead of **port-threshold** to control the LAG operational status as **port-threshold** is unaware of the port speed. When 1GE and 10GE or 10GE, 40GE, and 100GE ports are mixed in the same LAG then the weight assigned to each port is respectively 1 and 10 or 1, 4, and 10 thereby allowing the user to control the LAG threshold based on the effective port weight.

The **weight-threshold** or **hash-weight-threshold** commands can also be used for LAGs with all ports of equal speed to allow for a common operational model. For example, each port has a weight of 1 to mimic **port-threshold** and its related configuration.

- Nokia recommends that users use weight-based thresholds for other system configurations that react to operational change of LAG member ports, like MCAC (see **use-lag-port-weight**) and VRRP (see **weight-down**).
- When sub-groups are used, the following behavior should be noted for selection criteria:
 - **highest-count**
highest-count continues to operate on physical link counts. Therefore, a sub-group with lower speed links is selected even if its total bandwidth is lower. For example: a 4 * 10GE subgroup is selected over a 100GE + 10 GE sub-group).
 - **highest-weight**
highest-weight continues to operate on operator-configured priorities. Therefore, it is expected that configured weights take into account the proportional bandwidth difference between member ports to achieve the wanted behavior. For example, to favor sub-groups with higher bandwidth capacity but lower link count in a 10GE/100GE LAG, 100GE ports need to have their priority set to a value that is at least 10 times that of the 10GE ports priority value.

– **best-port**

best-port continues to operate on operator-configured priorities. Therefore, it is expected that the configured weights take into account proportional bandwidth difference between member ports to achieve the wanted behavior.

Migrating a LAG to higher speed ports is done following the process below:



Note: LAG must be operational with same speed ports before performing the steps below.

1. Turn on mixed-speed LAG using **port-weight-speed**
2. Add compatible higher speed ports to the LAG
3. Optionally use **weight-threshold** or **hash-weight-threshold** instead of **port-threshold**
4. Remove lower speed links
5. Turn off mixed-speed LAG using **no port-weight-speed**

Feature limitations in mixed-speed LAGs:

- PIM **lag-usage-optimization** is not supported and must not be configured
- LAG member links must have the default configuration for **config>port>ethernet>egress-rate/ingress-rate**
- Not supported for ESM
- The following ports or cards are not supported:
 - VSM ports
 - 10/100 FE ports
 - ESAT ports
 - PXC ports
- LAN and WAN port combinations in the same LAG:
 - support for 100GE LAN with 10GE WAN
 - no support for 100GE LAN with both 10GE LAN and 10GE WAN
 - no support for mixed 10GE LAN and 10GE WAN

2.7.4.4 LAG port hash-weight

The LAG port **hash-weight** allows the customization of the flow hashing distribution between LAG ports by adjusting the weight of each port independently for both same-speed and mixed-speed LAGs.

The LAG port **hash-weight** capabilities combined with the support for additional mixed-speed LAG combinations as compared to **port-weight-speed** make the LAG port **hash-weight** a more flexible feature.

Common rules for LAG port **hash-weight**:

- The configured **hash-weight** value per port is ignored until all the ports in the LAG have a **hash-weight** configured.
- The **hash-weight** value can be set to **port-speed** or an integer value from 1 to 100000:
 - **port-speed**

This allows the user to implicitly assign a **hash-weight** value based on the physical port speed.

– **1 to 100000**

This value range allows for control of flow hashing distribution between LAG ports.

- The LAG port **hash-weight** value is normalized internally to distribute flows between LAG ports. The minimum value returned by this normalization is 1.
- The LAG port **hash-weight** defaults to the port-speed value when unconfigured.

The **hash-weight** values using **port-speed** per physical port types are shown in [Table 25: Port types and speeds](#).

Table 25: Port types and speeds

Port type	Port speed
FE port	port-speed value 1
1GE port	port-speed value 1
10GE port	port-speed value 10
25GE port	port-speed value 25
40GE port	port-speed value 40
100GE port	port-speed value 100
400GE port	port-speed value 400
Other ports	port-speed value 1

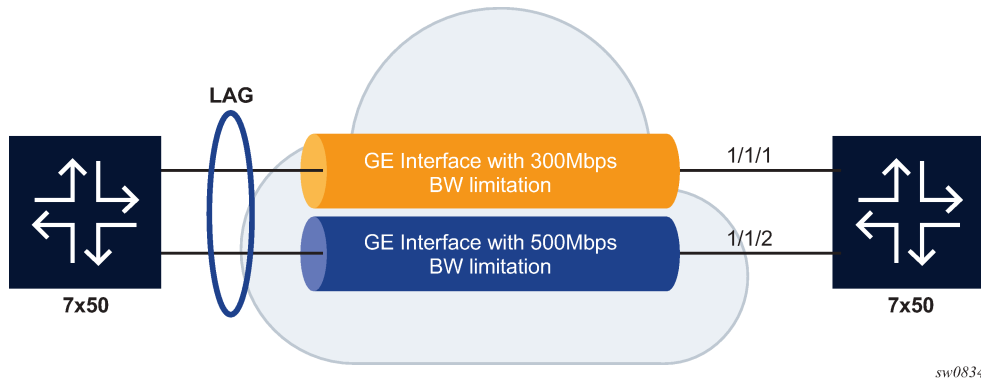
2.7.4.4.1 Configurable hash weight to control flow distribution

The user can use the LAG port **hash-weight** to further control the flow distribution between LAG ports.

This capability can be especially useful when LAG links on Nokia routers are rate limited by a 3rd party user providing the connectivity between two sites as shown in the [Figure 35: Same speed LAG with ports of different hash-weight](#):

- LAG links 1/1/1 and 1/1/2 are GE
- LAG link 1/1/1 is rate limited to 300 Mb/s by a 3rd party transport operator
- LAG Link 1/1/2 is rate limited to 500 Mb/s by a 3rd party transport operator

Figure 35: Same speed LAG with ports of different hash-weight



In this context, the user can configure the LAG to adapt the flow distribution between LAG ports according to the bandwidth restrictions on each port using customized hash-weight values as shown in the following configuration example:

```
A:Dut-A>config>lag# info
-----
port 1/1/1 hash-weight 300
port 1/1/2 hash-weight 500
no shutdown
-----
```

The user can also display the resulting flow-distribution between active LAG ports using the following **show** command:

```
A:Dut-A# show lag 3 flow-distribution
=====
Distribution of allocated flows
=====
Port                Bandwidth (Gbps) Hash-weight  Flow-share (%)
-----
1/1/1                10.000           300          37.50
1/1/2                10.000           500          62.50
-----
Total operational bandwidth: 20.000
=====
```



Note: hash-weight and same-speed LAG:

- If all ports have a **hash-weight** configured other than **port-speed**, then the configured value is used and normalized to modify the hashing between LAG ports.
- If the LAG ports are all configured to **port-speed**, or if only some of the ports have a customized **hash-weight** value, then the system uses a hash weight of 1 for every port in case of same-speed LAG. For mixed-speed LAG, the system uses the **port-speed** value.

2.7.4.4.2 Mixed-speed LAGs

Combining ports of different speeds in the same LAG is supported by default and the user does not need to use the **port-weight-speed** or a customized **hash-weight** values.

The user can modify a LAG, in service, by adding or removing ports of different speeds. In cases where some, but not all, ports in a mixed-speed LAG have a **hash-weight** value configured, all LAG ports use the **port-speed** value.

The different combinations of physical port speeds supported in the same LAG are:

- 1GE and 10GE
- 10GE, 25GE, 40GE, and 100GE
- 100GE and 400GE

For mixed-speed LAGs:

- Both LACP and non-LACP configurations are supported. With LACP enabled, LACP is unaware of physical port differences.
- QoS is distributed according to the **internal-scheduler-weight-mode** command. By default, the **hash-weight** value is taken into account.
- User data traffic is hashed proportionally to the **hash-weight** value.
- CPM-originated OAM control traffic that requires per LAG hashing is hashed per physical port.
- When sub-groups are used, the following behavior should be noted for selection criteria:
 - **highest-count**
highest-count continues to operate on physical link counts. Therefore, a sub-group with lower speed links is selected even if its total bandwidth is lower. For example, a 4 * 10GE subgroup is selected over a 100GE + 10 GE sub-group).
 - **highest-weight**
highest-weight continues to operate on user-configured priorities. Therefore, it is expected that configured weights take into account the proportional bandwidth difference between member ports to achieve the wanted behavior. For example, to favor sub-groups with higher bandwidth capacity but lower link count in a 10GE/100GE LAG, 100GE ports need to have their priority set to a value that is at least 10 times that of the 10GE ports priority value.
 - **best-port**
best-port continues to operate on user-configured priorities. Therefore, it is expected that the configured weights take into account proportional bandwidth difference between member ports to achieve the wanted behavior.

Feature limitations in mixed-speed LAG:

- PIM **lag-usage-optimization** is not supported and must not be configured
- LAG member links must have the default configuration for **config>port>ethernet>egress-rate/ingress-rate**
- No support for ESM
- No support for LAG weight-threshold
- LAN and WAN port combinations in the same LAG:
 - support for 100GE LAN with 10GE WAN
 - no support for 100GE LAN with both 10GE LAN and 10GE WAN
 - no support for mixed 10GE LAN and 10GE WAN

The following ports do not support a customized LAG port **hash-weight** value other than **port-speed** and are not supported in a mixed-speed LAG:

- VSM ports
- 10/100 FE ports
- ESAT ports
- PXC ports

2.7.4.5 Per-link hashing

The hashing feature described in this section applies to traffic going over LAG and MC-LAG. Per-link hashing ensures all data traffic on a SAP or network interface uses a single LAG port on egress. Because all traffic for a specific SAP/network interface egresses over a single port, QoS SLA enforcement for that SAP, network interface is no longer impacted by the property of LAG (distributing traffic over multiple links). Internally-generated, unique IDs are used to distribute SAPs/network interface over all active LAG ports. As ports go UP and DOWN, each SAP and network interface is automatically rehashed so all active LAG ports are always used.

The feature is best suited for deployments when SAPs/network interfaces on a LAG have statistically similar BW requirements (because per SAP/network interface hash is used). If more control is required over which LAG ports SAPs/network interfaces egress on, a LAG link map profile feature described later in this guide may be used.

Per-link hashing, can be enabled on a LAG as long as the following conditions are met:

- LAG **port-type** must be standard.
- LAG **access adapt-qos** must be link or port-fair (for LAGs in **mode** access or hybrid).
- LAG mode is access/hybrid and the **access adapt-qos** mode is **distribute include-egr-hash-cfg**

2.7.4.5.1 Weighted per-link-hash

Weighted per-link-hash allows higher control in distribution of SAPs/interfaces/subscribers across LAG links when significant differences in SAPs/interfaces/subscribers bandwidth requirements could lead to an unbalanced distribution bandwidth utilization over LAG egress. The feature allows operators to configure for each SAPs/interfaces/subscribers on a LAG one of three unique classes and a weight value to be used to when hashing this service/subscriber across the LAG links. SAPs/interfaces/subscribers are hashed to LAG links, such that within each class the total weight of all SAPs/interfaces/subscribers on each LAG link is as close as possible to each other.

Multiple classes allow grouping of SAPs/interfaces/subscribers by similar bandwidth class/type. For example a class can represent: voice – negligible bandwidth, Broadband – 10 to 100 Mb/s, Extreme Broadband – 300 Mb/s and above types of service. If a class and weight are not specified for a specific service or subscriber, values of 1 and 1 are used respectively.

The following algorithm hashes SAPs, interfaces, and subscribers to LAG egress links:

- TPSDA subscribers are hashed to a LAG link when subscribers are active, MSE SAPs/interfaces are hashed to a LAG link when configured
- For a new SAP/interface/subscriber to be hashed to an egress LAG link, select the active link with the smallest current weight for the SAP/network/subscriber class.
- On a LAG link failure:

- Only SAPs/interfaces/subscribers on a failed link are rehashed over the remaining active links
- Processing order: per class from lowest numerical, within each class per weight from highest numerical value
- LAG link recovery/new link added to a LAG:
 - auto-rebalance disabled: existing SAPs/interfaces/subscribers remain on the currently active links, new SAPs/interfaces/subscribers naturally prefer the new link until balance reached.
 - auto-rebalance is enabled: when a new port is added to a LAG a non-configurable 5 second rebalance timer is started. Upon timer expiry, all existing SAPs/interfaces/subscribers are rebalanced across all active LAG links minimizing the number of SAPs/interfaces/subscribers moved to achieve rebalance. The rebalance timer is restarted if a new link is added while the timer is running. If a port bounces 5 times within a 5 second interval, the port is quarantined for 10 seconds. This behavior is not configurable.
 - On a LAG startup, the rebalance timer is always started irrespective of auto-rebalance configuration to avoid hashing SAPs/interfaces/subscribers to a LAG before ports have a chance to come UP.
- Weights for network interfaces are separated from weights for access SAPs/interfaces/subscribers.
- On a mixed-speed LAG, link selection is made with link speeds factoring into the overall weight for the same class of traffic. This means that higher-speed links are preferred over lower-speed links.

Optionally an operator can use a **tools perform lag load-balance** command to manually re-balance all weighted per-link-hashed SAPs/interfaces/subscribers on a LAG. The rebalance follows the algorithm as used on a link failure moving SAPs/interfaces/subscribers to different LAG links to minimize SAPs/interfaces/subscribers impacted.

Along with the restrictions for standard per-link hashing, the following restrictions exist:

- When weighted per-link-hash is deployed on a LAG, no other methods of hash for subscribers/SAPs/interfaces on that LAG (like service hash or LAG link map profile) should be deployed, because the weighted hash is not able to account for loads placed on LAG links by subscriber/SAPs/interfaces using the other hash methods.
- For the TPSDA model only the 1:1 (subscriber to SAP) model is supported.

This feature does not operate properly if the above conditions are not met.

2.7.4.6 Explicit per-link hash using LAG link mapping profiles

The hashing feature described in this section applies to traffic going over LAG and MC-LAG. LAG link mapping profile feature gives users full control of which links SAPs/network interface use on a LAG egress and how the traffic is rehashed on a LAG link failure. Some benefits that such functionality provides include:

- Ability to perform management level admission control onto LAG ports therefore increasing overall LAG BW utilization and controlling LAG behavior on a port failure.
- Ability to strictly enforce QoS contract on egress for a SAP or network interface or a group of SAPs or network interfaces by forcing egress over a single port and using **access adapt-qos** link or port-fair mode.

To enable LAG Link Mapping Profile Feature on a LAG, users configure one or more of the available LAG link mapping profiles on the LAG and then assign that profiles to all or a subset of SAPs and network interfaces as needed. Enabling per LAG link Mapping Profile is allowed on a LAG with services configured,

a small outage may take place as result of re-hashing SAP/network interface when a lag profile is assigned to it.

Each LAG link mapping profile allows users to configure:

- primary link** a port of the LAG to be used by a SAP or network interface when the port is UP. Note that a port cannot be removed from a LAG if it is part of any LAG link profile.
- secondary link** a port of the LAG to be used by a SAP or network interface as a backup when the primary link is not available (not configured or down) and the secondary link is UP

When neither primary, nor secondary links are available (not configured or down), use one of the following modes of operation:

- **discard**

Traffic for a specific SAP or network interface is dropped to protect other SAPs/network interfaces from being impacted by re-hashing these SAPs or network interfaces over remaining active LAG ports.



Note: SAP or network interface status is not affected when primary and secondary links are unavailable, unless an OAM mechanism that follows the datapath hashing on egress is used and causes a SAP or network interface to go down.

- **per-link-hash**

Traffic for a specific SAP or network interface is re-hashed over remaining active ports of a LAG links using per-link-hashing algorithm. This behavior ensures that SAPs or network interfaces using this profile are provided available resources of other active LAG ports even if it impacts other SAPs or network interfaces on the LAG. The system uses the QoS configuration to provide fairness and priority if congestion is caused by the default-hash recovery.

LAG link mapping profiles, can be enabled on a LAG as long as the following conditions are met:

- LAG **port-type** must be *standard*.
- LAG **access adapt-qos** must be *link* or *port-fair* (for LAGs in **mode** access or hybrid)
- All ports of a LAG on a router must belong to a single sub-group.
- Access adapt-qos mode is distribute include-egr-hash-cfg.

LAG link mapping profile can coexist with any-other hashing used over a specific LAG (for example, per-flow hashing or per-link-hashing). SAPs/network interfaces that have no link mapping profile configured are subject to LAG hashing, while SAPs/network interfaces that have configured LAG profile assigned are subject to LAG link mapping behavior, which is described above.

2.7.4.7 Consistent per-service hashing

The hashing feature described in this section applies to traffic going over LAG, Ethernet tunnels (eth-tunnel) in load-sharing mode, or CCAG load balancing for VSM redundancy. The feature does not apply to ECMP.

Per-service-hashing was introduced to ensure consistent forwarding of packets belonging to one service. The feature can be enabled using the **[no] per-service-hashing** configuration option under **config>service>epipe>load-balancing** and **config>service>vpls>load-balancing**, valid for Epipe, VPLS, PBB Epipe, IVPLS and BVPLS.

The following behavior applies to the usage of the [no] per-service-hashing option.

- The setting of the PBB Epipe/I-VPLS children dictates the hashing behavior of the traffic destined for or sourced from an Epipe/I-VPLS endpoint (PW/SAP).
- The setting of the B-VPLS parent dictates the hashing behavior only for transit traffic through the B-VPLS instance (not destined for or sourced from a local I-VPLS/Epipe children).

The following algorithm describes the hash-key used for hashing when the new option is enabled:

- If the packet is PBB encapsulated (contains an I-TAG Ethertype) at the ingress side and enters a B-VPLS service, use the ISID value from the I-TAG. For PBB encapsulated traffic entering other service types, use the related service ID.
- If the packet is not PBB encapsulated at the ingress side
 - For regular (non-PBB) VPLS and Epipe services, use the related service ID
 - If the packet is originated from an ingress IVPLS or PBB Epipe SAP
 - If there is an ISID configured use the related ISID value
 - If there is no ISID configured use the related service ID
 - For BVPLS transit traffic use the related flood list ID
 - Transit traffic is the traffic going between BVPLS endpoints
 - An example of non-PBB transit traffic in BVPLS is the OAM traffic
- The above rules apply to Unicast, BUM flooded without MMRP or with MMRP, IGMP snooped regardless of traffic type

Users may sometimes require the capability to query the system for the link in a LAG or Ethernet tunnel that is currently assigned to a specific service-id or ISID. This capability is provided using the **tools>dump>map-to-phy-port** {**ccag** *ccag-id* | **lag** *lag-id* | **eth-tunnel** *tunnel-index*} {**isid** *isid* [**end-isid** *isid*] | **service** *servid-id* | *svc-name* [**end-service** *service-id* | *svc-name*]} [**summary**] command.

An example usage is as follows:

```
A:Dut-B# tools dump map-to-phy-port lag 11 service 1
```

ServiceId	ServiceName	ServiceType	Hashing	Physical Link
1		i-vpls	per-service(if enabled)	3/2/8

```
A:Dut-B# tools dump map-to-phy-port lag 11 isid 1
```

ISID	Hashing	Physical Link
1	per-service(if enabled)	3/2/8

```
A:Dut-B# tools dump map-to-phy-port lag 11 isid 1 end-isid 4
```

ISID	Hashing	Physical Link
1	per-service(if enabled)	3/2/8
2	per-service(if enabled)	3/2/7
3	per-service(if enabled)	1/2/2
4	per-service(if enabled)	1/2/3

2.7.4.8 ESM

In ESM, egress traffic can be load balanced over LAG member ports based on the following entities:

- Per subscriber, in weighted and non-weighted mode
- Per Vport, on non HSQ cards in weighted and non-weighted
- Per secondary shaper on HSQ cards
- Per destination MAC address when ESM is configured in a VPLS (Bridged CO)

ESM over LAGs with configured PW ports require additional considerations:

- PW SAPs are not supported in VPLS services or on HSQ cards. This means that load balancing per secondary shaper or destination MAC are not supported on PW ports with a LAG configured under them.
- Load balancing on a PW port associated with a LAG with faceplate member ports (fixed PW ports) can be performed per subscriber or Vport.
- Load balancing on a FPE (or PXC)-based PW port is performed on two separate LAGs which can be thought of as two stages:
 - Load balancing on a PXC LAG where the subscribers are instantiated. In this first stage, the load balancing can be performed per subscriber or per Vport.
 - The second stage is the LAG over the network faceplate ports over which traffic exits the node. Load balancing is independent of ESM and must be examined in the context of Epipe or EVPN VPWS that is stitched to the PW port.

2.7.4.8.1 Load balancing per subscriber

Load balancing per subscriber supports two modes of operation.

The first mode is native non-weighted per-subscriber load balancing in which traffic is directly hashed per subscriber. Use this mode in SAP and subscriber (1:1) deployments and in SAP and service (N:1) deployments. Examples of services in SAP and services deployments are VoIP, video, or data.

In this mode of operation, the following configuration requirements must be met.

- Any form of **per-link-hash** command in a LAG under the **configure lag** hierarchy must be disabled. This is the default setting.
- If QoS schedulers or Vports are used on the LAG, their bandwidth must be distributed over LAG member ports in a port-fair operation.

Example: MD-CLI

```
configure lag "lag-100" access adapt-qos mode port-fair
```

Example: classic CLI

```
configure lag access adapt-qos port-fair
```

In this scenario, setting this parameter to in **adapt-qos** to mode **link** disables per-subscriber load balancing and enable per-Vport load balancing.

The second mode, the weighted per subscriber load balancing is supported only in SAP and subscriber (1:1) deployments, and it requires the following configurations.

Example: MD-CLI

```
configure lag "lag-100" per-link-hash weighted subscriber-hash-mode sap
```

Example: classic CLI

```
configure lag per-link-hash weighted subscriber-hash-mode sap
```

In this scenario where hashing is performed per SAP, as reflected in the CLI above, in terms of load balancing, per-SAP hashing produces the same results as per-subscriber hashing because SAPs and subscribers are in a 1:1 relationship. The end result is that the traffic is load balanced per-subscribers, regardless of this indirection between hashing and load-balancing.

With the **per-link-hash** option enabled, the SAPs (and with this, the subscribers) are dynamically distributed over the LAG member links. This dynamic behavior can be overridden by configuring the **lag-link-map-profiles** command under the static SAPs or under the **msap-policy**. This way, each static SAP, or a group of MSAPs sharing the same **msap-policy** are statically and deterministically assigned to a preordained member port in a LAG.

This mode allows classes and weights to be configured for a group of subscribers with a shared subscriber profile under the following hierarchy.

Example: MD-CLI

```
configure subscriber-mgmt sub-profile egress lag-per-link-hash class
configure subscriber-mgmt sub-profile egress lag-per-link-hash weight
```

Example: classic CLI

```
configure subscriber-mgmt sub-profile egress lag-per-link-hash class {1|2|3} weight <[1 to 1024]>
```

Default values for **class** and **weight** are 1. If all subscribers on a LAG are configured with the same values for class and weight, load balancing effectively becomes non-weighted.

If QoS schedulers and Vports are used on the LAG, their bandwidth should be distributed over LAG member ports in a port-fair operation.

Example: MD-CLI

```
configure lag "lag-100" access adapt-qos mode port-fair
```

Example: classic CLI

```
configure lag access adapt-qos port-fair
```

2.7.4.8.2 Load balancing per Vport

Load balancing per Vport applies to user bearing traffic, and not to the control traffic originated or terminated on the BNG, required to setup and maintain sessions, such as PPPoE and DHCP setup and control messages.

Per Vport load balancing supports two modes of operation.

In the first mode, non-weighted load balancing based on Vport hashing, the following LAG-related configuration is required.

The **per-link-hash** command must be disabled.

Example: MD-CLI

```
configure lag "lag-100" access adapt-qos mode link
```

Example: classic CLI

```
configure lag access adapt-qos link
```

If LAG member ports are distributed over multiple forwarding complexes, the following configuration is required.

```
configure subscriber-mgmt sub-profile vport-hashing
```

The second mode, weighted load balancing based on Vport hashing, supports **class** and **weight** parameters per Vport. To enable weighted traffic load balancing per Vport, the following configuration must be enabled.

```
configure lag per-link-hash weighted subscriber-hash-mode vport
```

The class and weight can be optionally configured under the Vport definition.

Example: MD-CLI

```
configure port ethernet access egress virtual-port lag-per-link-hash class
configure port ethernet access egress virtual-port lag-per-link-hash weight
```

Example: classic CLI

```
configure port ethernet access egress vport lag-per-link-hash class weight
```

2.7.4.8.3 Load balancing per secondary shaper

Load balancing based on a secondary shaper is supported only on HSQ cards and only in non-weighted mode. The following LAG-related configuration is required. The **per-link-hash** command first must be disabled.

Example: MD-CLI

```
configure lag "lag-100" access adapt-qos mode link
```

Example: classic CLI

```
configure lag access adapt-qos link
```

Use the following commands to disable **per-link-hash**.

Example: MD-CLI

```
configure lag delete per-link-hash
```

Example: classic CLI

```
configure lag no per-link-hash
```

2.7.4.8.4 Load balancing per destination MAC

This load balancing mode is supported only when ESM is enabled in VPLS in Bridged Central Office (CO) deployments. In this mode of operation, the following configuration is required. The **per-link-hash** command first must be disabled.

```
configure subscriber-mgmt msap-policy vpls-only-sap-parameters mac-da-hashing
configure service vpls sap 1/1/2:4 sub-sla-mgmt mac-da-hashing
```

2.7.4.9 IPv6 flow label load balancing

IPv6 flow label load balancing enables load balancing in ECMP and LAG based on the output of a hash performed on the triplet {SA, DA, Flow-Label} in the header of an IPv6 packet received on a IES, VPRN, R-VPLS, CsC, or network interface.

IPv6 flow label load balancing complies with the behavior described in RFC 6437. When the **flow-label-load-balancing** command is enabled on an interface, the router applies a hash on the triplet {SA, DA, Flow-Label} to IPv6 packets received with a non-zero value in the flow label.

If the flow label field value is zero, the router performs the hash on the packet header; using the existing behavior based on the global or interface-level commands.

When enabled, IPv6 flow label load balancing also applies hashing on the triplet {SA, DA, Flow-Label} of the outer IPv6 header of an SRv6 encapsulated packet that is received on a network interface of a SRv6 transit router.

At the ingress PE router, SRv6 supports inserting the output of the hash that is performed on the inner IPv4, IPv6, or Ethernet service packet header into the flow label field of the outer IPv6 header it pushes on the SRv6 encapsulated packet.

For more details of the hashing and spraying of packets in SRv6, see the *7750 SR and 7950 XRS Segment Routing and PCE User Guide*.

The flow label field in the outer header of a received IPv6 or SRv6 encapsulated packet is never modified in the datapath.

2.7.4.9.1 Interaction with other load balancing features

IPv6 flow label load balancing interacts with other load balancing features as follows:

- When the **flow-label-load-balancing** command is enabled on an interface and the global level **I4-load-balancing** command is also enabled, it applies to all IPv4 packets and to IPv6 packets with a flow label field of zero.

- The following global load-balancing commands apply independently to the corresponding non-IPv6 packet encapsulations:
 - **lsr-load-balancing**
 - **mc-enh-load-balancing**
 - **service-id-lag-hashing**
- When the **flow-label-load-balancing** command is enabled on an interface and the global load balancing **l2tp-load-balancing** command is enabled, it applies to the following situations:
 - packets received with L2TPv2 over UDP/IPv4 encapsulation
 - packets received with L2TPv3 over UDP/IPv4 encapsulation
 - packets received with L2TPv3 over UDP/IPv6 encapsulation if the flow label field is zero. Otherwise, flow label hashing applies.

Packets received with L2TPv3 directly over IPv6 are not hashed on the L2TPv3 session ID. Therefore, hashing of these packets is based on the other interface level hash commands if the flow label field is zero. If the flow label is not zero, flow label hashing applies.



Note: SR OS implementation of L2TPv3 supports UDP/IPv6 encapsulation only. However, third-party implementations may support L2TPv3 directly over IPv6 encapsulation.

- The global load-balancing command selects a different hashing algorithm and therefore applies all the time when enabled, including when the **flow-label-load-balancing** command is enabled on the interface: **system-ip-load-balancing**.
- When the **flow-label-load-balancing** command is enabled on an interface and the per-interface **spi-load-balancing** or **teid-load-balancing** commands are enabled, they apply to all IPv4 packets and to IPv6 packets with a flow label field of zero.
- The following per-interface load-balancing command applies independently to MPLS encapsulated packets: **lsr-load-balancing**
- The per-interface load balancing command, **egr-ip-load-balancing** and the **flow-label-load-balancing** command are mutually exclusive. The CLI enforces this exclusivity.
- The following per-LAG port packet spraying commands override the **flow-label-load-balancing** command. IPv6 packets, with a non-zero flow label value, are sprayed over LAG links according to the enabled LAG-spraying mode.
 - **per-link-hash**
 - **link-map-profile**

2.7.5 LAG hold-down timers

Users can configure multiple hold-down timers that allow control how quickly LAG responds to operational port state changes. The following timers are supported:

- **port-level hold-time up/down timer**

This optional timer allows user to control delay for adding/removing a port from LAG when the port comes UP/goes DOWN. Each LAG port runs the same value of the timer, configured on the primary LAG link. See the Port Link Dampening description in [Port features](#) for more details on this timer.

- **sub-group-level hold-time timer**

This optional timer allows user to control delay for a switch to a new candidate sub-group selected by LAG sub-group selection algorithm from the current, operationally UP sub-group. The timer can also be configured to never expire, which prevents a switch from operationally up sub-group to a new candidate sub-group (manual switchover is possible using tools perform force lag command). Note that, if the port link dampening is deployed, the port level timer must expire before the sub-group-selection takes place and this timer is started. Sub-group-level hold-down timer is supported with LAGs running LACP only.

- **LAG-level hold-time down timer**

This optional timer allows user to control delay for declaring a LAG operationally down when the available links fall below the required port/BW minimum. The timer is recommended for LAG connecting to MC-LAG systems. The timer prevents a LAG going down when MC-LAG switchover executes break-before-make switch. Note that, if the port link dampening is deployed, the port level timer must expire before the LAG operational status is processed and this timer is started.

2.7.6 BFD over LAG links

The router supports the application of BFD to monitor individual LAG link members to speed up the detection of link failures. When BFD is associated with an Ethernet LAG, BFD sessions are setup over each link member, and are referred to as micro-BFD sessions. A link is not operational in the associated LAG until the associated micro-BFD session is fully established. In addition, the link member is removed from the operational state in the LAG if the BFD session fails.

When configuring the local and remote IP address for the BFD over LAG link sessions, the **local-ip** parameter should always match an IP address associated with the IP interface to which this LAG is bound. In addition, the **remote-ip** parameter should match an IP address on the remote system and should also be in the same subnet as the **local-ip** address. If the LAG bundle is re-associated with a different IP interface, the **local-ip** and **remote-ip** parameters should be modified to match the new IP subnet. The **local-ip** and **remote-ip** values do not have to match in the case of hybrid mode, q-tag, or QinQ tagging.

2.7.7 Multi-chassis LAG

This section describes the Multi-Chassis LAG (MC-LAG) concept. MC-LAG is an extension of a LAG concept that provides node-level redundancy in addition to link-level redundancy provided by "regular LAG".

Typically, MC-LAG is deployed in a network-wide scenario providing redundant connection between different end points. The whole scenario is then built by combination of different mechanisms (for example, MC-LAG and redundant pseudowire to provide e2e redundant p2p connection or dual homing of DSLAMs in Layer 2/3 TPSDA).

2.7.7.1 Overview

Multi-chassis LAG is a method of providing redundant Layer 2/3 access connectivity that extends beyond link level protection by allowing two systems to share a common LAG end point.

The multi-service access node (MSAN) node is connected with multiple links toward a redundant pair of Layer 2/3 aggregation nodes such that both link and node level redundancy, are provided. By using a multi-chassis LAG protocol, the paired Layer 2/3 aggregation nodes (referred to as redundant-pair) appears to be a single node utilizing LACP toward the access node. The multi-chassis LAG protocol between a redundant-pair ensures a synchronized forwarding plane to and from the access node and synchronizes

the link state information between the redundant-pair nodes such that correct LACP messaging is provided to the access node from both redundant-pair nodes.

To ensure SLAs and deterministic forwarding characteristics between the access and the redundant-pair node, the multi-chassis LAG function provides an active/standby operation to and from the access node. LACP is used to manage the available LAG links into active and standby states such that only links from 1 aggregation node are active at a time to/from the access node.

Alternatively, when access nodes do not support LACP, the **power-off** option can be used to enforce the active/standby operation. In this case, the standby ports are **trx_disabled** (power off transmitter) to prevent usage of the LAG member by the access-node. Characteristics related to MC are:

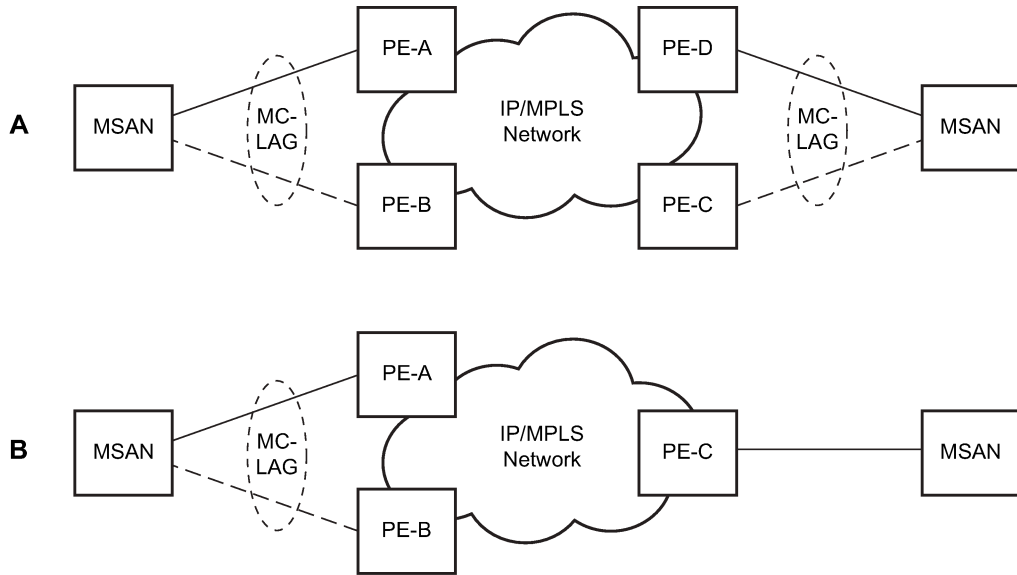
- Selection of the common system ID, system-priority and administrative-key are used in LACP messages so partner systems consider all links as the part of the same LAG.
- Extension of selection algorithm to allow selection of active sub-group.
 - The sub-group definition in LAG context is still local to the single box, meaning that even if sub-groups configured on two different systems have the same sub-group-id they are still considered as two separate subgroups within a specified LAG.
 - Multiple sub-groups per PE in an MC-LAG is supported.
 - In case there is a tie in the selection algorithm, for example, two sub-groups with identical aggregate weight (or number of active links) the group which is local to the system with lower system LACP priority and LAG system ID is taken.
- Providing inter-chassis communication channel allows inter-chassis communication to support LACP on both system. This communication channel enables the following:
 - Supports connections at the IP level which do not require a direct link between two nodes. The IP address configured at the neighbor system is one of the addresses of the system (interface or loop-back IP address).
 - The communication protocol provides heartbeat mechanism to enhance robustness of the MC-LAG operation and detecting node failures.
 - Support for operator actions on any node that force an operational change.
 - The LAG group-ids do not have to match between neighbor systems. At the same time, there can be multiple LAG groups between the same pair of neighbors.
 - Verification that the physical characteristics, such as speed and auto-negotiation is configured and initiates operator notifications (traps) if errors exist. Consistency of MC-LAG configuration (system-id, administrative-key and system-priority) is provided. Similarly, load-balancing mode of operation must be consistently configured on both nodes.
 - Traffic over the signaling link is encrypted using a user configurable message digest key.
- MC-LAG function provides active/stand-by status to other software applications to build a reliable solution.

[Figure 36: MC-LAG Layer 2 dual homing to remote PE pairs](#) and [Figure 37: MC-LAG Layer 2 dual homing to local PE pairs](#) show the different combinations of MC-LAG attachments that are supported. The supported configurations can be sub-divided into following sub-groups:

- Dual-homing to remote PE pairs
 - both end-points attached with MC-LAG
 - one end-point attached
- Dual-homing to local PE pair

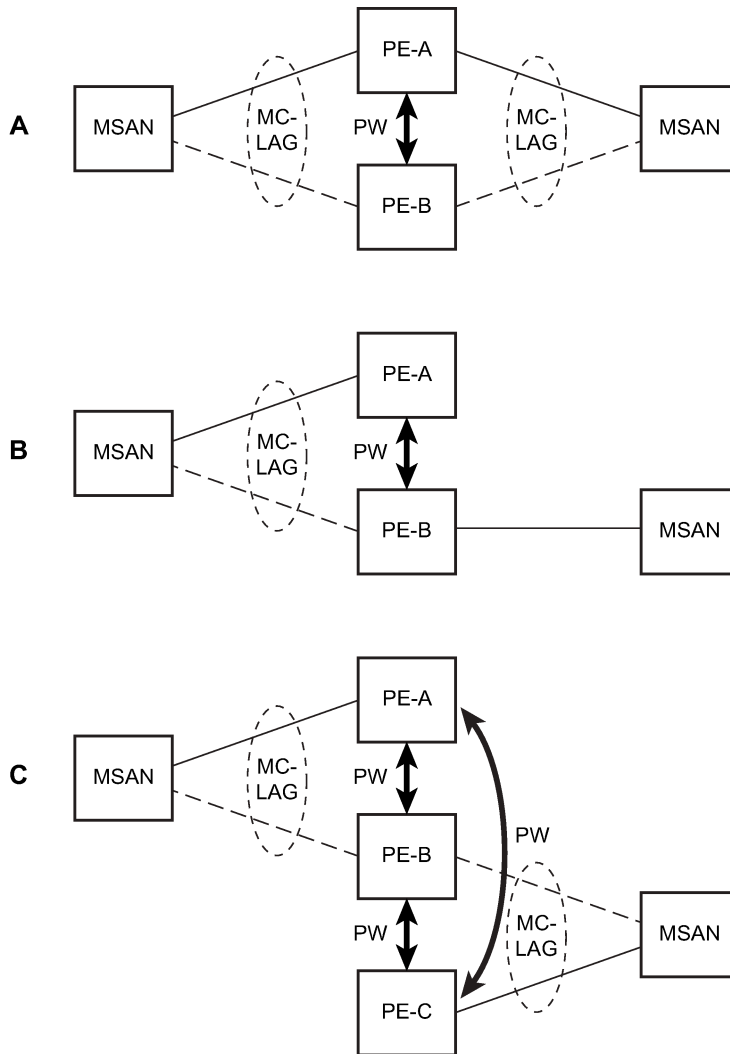
- both end-points attached with MC-LAG
- one end-point attached with MC-LAG
- both end-points attached with MC-LAG to two overlapping pairs

Figure 36: MC-LAG Layer 2 dual homing to remote PE pairs



Fig_6

Figure 37: MC-LAG Layer 2 dual homing to local PE pairs



Fig_7

The forwarding behavior of the nodes abide by the following principles. Note that logical destination (actual forwarding decision) is primarily determined by the service (VPLS or VLL) and the principle below applies only if destination or source is based on MC-LAG:

- Packets received from the network are forwarded to all local active links of the specific destination-sap based on conversation hashing. In case there are no local active links, the packets are cross-connected to inter-chassis pseudowire.
- Packets received from the MC-LAG sap are forwarded to active destination pseudowire or active local links of destination-sap. In case there are no such objects available at the local node, the packets are cross-connected to inter-chassis pseudowire.

2.7.7.2 MC-LAG and SRRP

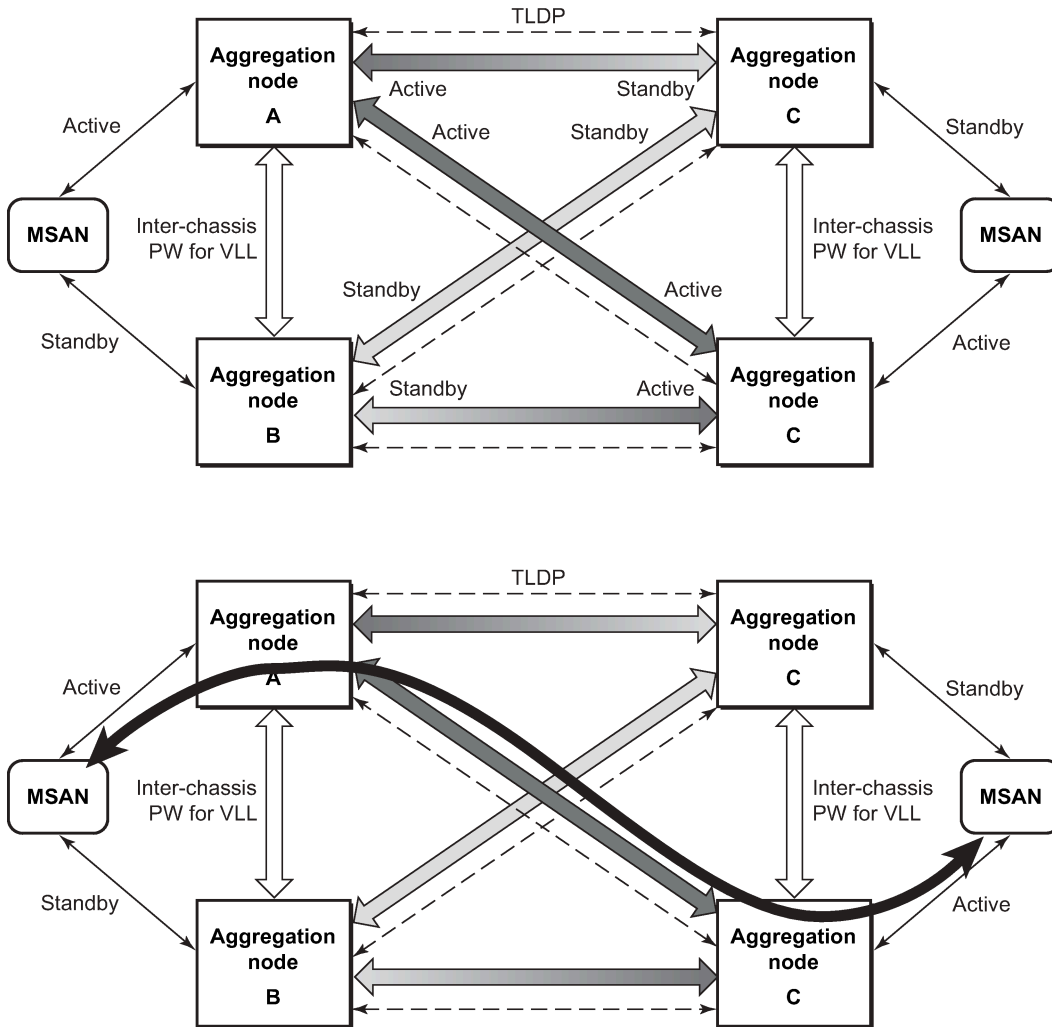
MC-LAG and Subscriber Routed Redundancy Protocol (SRRP) enable dual-homed links from any IEEE 802.1ax (formerly 802.3ad) standards-based access device (for example, a IP DSLAM, Ethernet switch or a Video on Demand server) to multiple Layer 2/3 or Layer 3 aggregation nodes. In contrast with slow recovery mechanisms such as Spanning Tree, multi-chassis LAG provides synchronized and stateful redundancy for VPN services or triple play subscribers in the event of the access link or aggregation node failing, with zero impact to end users and their services.

See the *7450 ESS, 7750 SR, and VSR Triple Play Service Delivery Architecture Guide* for information about SRRP.

2.7.7.3 P2P redundant connection across Layer 2/3 VPN network

[Figure 38: Point-to-Point \(P2P\) redundant connection through a Layer 2 VPN network](#) shows the connection between two multi-service access nodes (MSANs) across a network based on Layer 2/3 VPN pseudowires. The connection between MSAN and a pair of PE routers is realized by MC-LAG. From an MSAN perspective, a redundant pair of PE routers acts as a single partner in LACP negotiation. At any time, only one of the routers has an active link in a specified LAG. The status of LAG links is reflected in status signaling of pseudowires set between all participating PEs. The combination of active and stand-by states across LAG links as well as pseudowires gives only one unique path between a pair of MSANs.

Figure 38: Point-to-Point (P2P) redundant connection through a Layer 2 VPN network



OSSG116

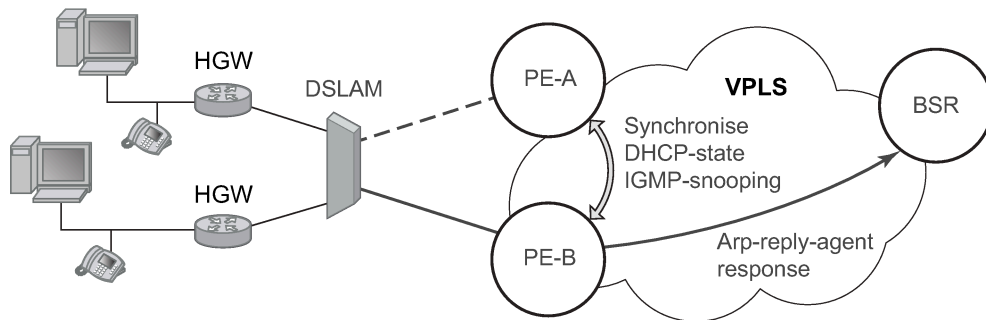
Note that the configuration in [Figure 38: Point-to-Point \(P2P\) redundant connection through a Layer 2 VPN network](#) shows one particular configuration of VLL connections based on MC-LAG, particularly the VLL connection where two ends (SAPs) are on two different redundant-pairs. In addition to this, other configurations are possible, such as:

- Both ends of the same VLL connections are local to the same redundant-pair.
- One end VLL endpoint is on a redundant-pair the other on single (local or remote) node.

2.7.7.4 DSLAM dual homing in Layer 2/3 TPSDA model

[Figure 39: DSLAM dual-homing using MC-LAG](#) shows a network configuration where DSLAM is dual homed to pair of redundant PEs by using MC-LAG. Inside the aggregation network redundant-pair of PEs is connecting to VPLS service which provides reliable connection to single or pair of Broadband Service Routers (BSRs).

Figure 39: DSLAM dual-homing using MC-LAG



OSSG110

MC-LAG and pseudowire connectivity, PE-A and PE-B implement enhanced subscriber management features based on DHCP-snooping and creating dynamic states for every subscriber-host. As in any point of time there is only one PE active, it is necessary to provide the mechanism for synchronizing subscriber-host state-information between active PE (where the state is learned) and stand-by PE. In addition, VPLS core must be aware of active PE to forward all subscriber traffic to a PE with an active LAG link. The mechanism for this synchronization is outside of the scope of this document.

2.7.8 LAG IGP cost

When using a LAG, it is possible to take an operational link degradation into consideration by setting a configurable degradation threshold. The following alternative settings are available through configuration:

- port-threshold value
- weight-threshold value
- hash-weight-threshold value

When the LAG operates under normal circumstances and is included in an IS-IS or OSPF routing instance, the LAG must be associated with an IGP link cost. This LAG cost can either be statically configured in the IGP context or set dynamically by the LAG based upon the combination of the interface speed and reference bandwidth.

Under operational LAG degradation however, it is possible for the LAG to set a new updated dynamic or static threshold cost taking the gravity of the degradation into consideration.

As a consequence, there are some IGP link cost alternatives available, for which the most appropriate must be selected. The IGP uses the following priority rules to select the most appropriate IGP link cost:

1. Static LAG cost (from the LAG threshold action during degradation)
2. Explicit configured IGP cost (from the configuration under the IGP routing protocol context)
3. Dynamic link cost (from the LAG threshold action during degradation)
4. Default metric (no cost is set anywhere)

For example:

- Static LAG cost overrules the configured metric.
- Dynamic cost does not overrule configured metric or static LAG cost.

2.8 G.8031 protected Ethernet tunnels

The Nokia PBB implementation offers the capability to use core Ethernet tunnels compliant with ITU-T G.8031 specification to achieve 50 ms resiliency for failures in a native Ethernet backbone. For further information about Ethernet tunnels, see "G.8031 Protected Ethernet Tunnels" in the *7450 ESS*, *7750 SR*, *7950 XRS*, and *VSR Services Overview Guide*.

2.9 G.8032 protected Ethernet rings

Ethernet ring protection switching offers ITU-T G.8032 specification compliance to achieve resiliency for Ethernet Layer 2 networks. Similar to G.8031 linear protection (also called Automatic Protection Switching (APS)), G.8032 (Eth-ring) is also built on Ethernet OAM and often referred to as Ring Automatic Protection Switching (R-APS).

For further information about Ethernet rings, see "G.8032 Protected Ethernet Rings" in the *7450 ESS*, *7750 SR*, *7950 XRS*, and *VSR Services Overview Guide*.

2.10 Ethernet port monitoring

Ethernet ports can record and recognize various medium statistics and errors. There are two main types of errors:

- **frame based**

Frame based errors are counted when the arriving frame has an error that means the frame is invalid. These types of errors are only detectable when frames are presents on the wire.

- **symbol based**

Symbol errors are invalidly encoded symbols on the physical medium. Symbols are always present on an active Ethernet port regardless of the presence of frames.

CRC-Monitor and Symbol-Monitor allows the user to monitor ingress error conditions on the Ethernet medium and compare these error counts to the thresholds. CRC-Monitor monitors CRC errors. Symbol-Monitor monitors symbol errors. Symbol Error is not supported on all Ethernet ports. Crossing a signal degrade (SD) threshold causes a log event to be raised. Crossing the configured signal failure (SF) threshold causes the port to enter an operation state of down. The user may consider the configuration of other protocols to convey the failure, through timeout conditions.

The error rates are in the form of $M \cdot 10^E - N$. The user has the ability to configure both the threshold (N) and a multiplier (M). By default if the multiplier is not configured the multiplier is 1. As an example, sd-threshold 3 would result in a signal degrade error rate of $1 \cdot 10^E - 3$ (one error per 1000). Changing the configuration to would sd-threshold 3 multiplier 5 result in a signal degrade rate of $5 \cdot 10^E - 3$ (5 errors per 1000). The signal degrade value must be a lower error rate than the signal failure threshold. This threshold can be used to provide notification that the port is operating in a degraded but not failed condition. These do not equate to a bit error rate (BER). CRC-Monitor provides a CRC error rate. Symbol-Monitor provides a symbol error rate.

The configured error thresholds are compared to the user specified sliding window to determine if one or both of the thresholds have been crossed. Statistics are gathered every second. This means that every

second the oldest statistics are dropped from the calculation. The default 10 second sliding window means that at the 11th second, the oldest 1-second statistical data is dropped and the 11th second is included.

Symbol error crossing differs slightly from CRC-based error crossing. The error threshold crossing is calculated based on the window size and the fixed number of symbols that arrive (ingress) on that port during that window. The following configuration demonstrates this concept.

```

config>port>ethernet# info detail
-----
      symbol-monitor
      sd-threshold 5 multiplier 5
      sf-threshold 3 multiplier 5
      no shutdown
    exit

show port 2/1/2 ethernet
=====
Ethernet Interface
=====
Description          : 2/1/2
Interface            : 2/1/2
Link-level           : Ethernet
Admin State          : down
Oper State            : down
Physical Link         : No
Single Fiber Mode    : No
IfIndex              : 69271552
Last State Change    : 06/29/2014 05:04:12
Last Cleared Time    : N/A
Phys State Chng Cnt : 0

Oper Speed           : N/A
Config Speed         : 1 Gbps
Oper Duplex          : N/A
Config Duplex        : full
MTU                  : 9212
Min Frame Length     : 64 Bytes
Hold time up         : 0 seconds
Hold time down       : 0 seconds
DDM Events           : Enabled

Configured Mode      : network
Dot1Q Ethertype      : 0x8100
PBB Ethertype        : 0x88e7
Ing. Pool % Rate     : 100
Ing. Pool Policy     : n/a
Egr. Pool Policy     : n/a
Net. Egr. Queue Pol : default
Egr. Sched. Pol      : n/a
Auto-negotiate       : true
Oper Phy-tx-clock    : not-applicable
Accounting Policy    : None
Acct Plcy Eth Phys   : None
Egress Rate          : Default
Load-balance-algo    : Default

Encap Type           : null
QinQ Ethertype       : 0x8100
Egr. Pool % Rate     : 100

MDI/MDX              : unknown

Collect-stats        : Disabled
Collect Eth Phys     : Disabled
Ingress Rate         : Default
LACP Tunnel          : Disabled

Down-when-looped    : Disabled
Loop Detected        : False
Use Broadcast Addr   : False

Keep-alive           : 10
Retry                 : 120

Sync. Status Msg.   : Disabled
Tx DUS/DNU           : Disabled
SSM Code Type        : sdh

Rx Quality Level     : N/A
Tx Quality Level     : N/A

Down On Int. Error  : Disabled

CRC Mon SD Thresh    : Disabled
CRC Mon SF Thresh    : Disabled

CRC Mon Window       : 10 seconds

Sym Mon SD Thresh    : 5*10E-5
Sym Mon SF Thresh    : 5*10E-3

Sym Mon Window       : 10 seconds
Tot Sym Mon Errs     : 0

EFM OAM              : Disabled
EFM OAM Link Mon     : Disabled

```

```

Configured Address : 8c:90:d3:a0:c7:42
Hardware Address   : 8c:90:d3:a0:c7:42

Transceiver Data

Transceiver Status : not-equipped
=====
Traffic Statistics
=====
                                     Input          Output
-----
Octets                    0              0
Packets                   0              0
Errors                    0              0
Utilization (300 seconds) 0.00%          0.00%
=====

Port Statistics
=====
                                     Input          Output
-----
Unicast Packets           0              0
Multicast Packets         0              0
Broadcast Packets         0              0
Discards                  0              0
Unknown Proto Discards    0
=====

Ethernet-like Medium Statistics
=====
Alignment Errors : 0  Sngl Collisions : 0
FCS Errors       : 0  Mult Collisions : 0
SQE Test Errors  : 0  Late Collisions : 0
CSE              : 0  Excess Collisns : 0
Too long Frames  : 0  Int MAC Tx Errs : 0
Symbol Errors    : 0  Int MAC Rx Errs : 0
In Pause Frames  : 0  Out Pause Frames : 0
=====

```

The above configuration results in an SD threshold of $5 \cdot 10^{-5}$ (0.00005) and an SF threshold of $5 \cdot 10^{-3}$ (0.005) over the default 10-second window. If this port is a 1GbE port supporting symbol monitoring then the error rate is compared against 1,250,000,000 symbols (10 seconds worth of symbols on a 1GbE port 125,000,000). If the error count in the current 10 second sliding window is less than 62,500 then the error rate is below the signal degrade threshold and no action is taken. If the error count is between 62,501 and 6,250,000 then the error rate is above signal degrade but has not breached the signal failure signal threshold and a log event is raised. If the error count is above 6,250,000 the signal failure threshold is crossed and the port enters an operation state of down. Consider that this is a very simple example meant to demonstrate the function and not meant to be used as a guide for configuring the various thresholds and window times.

A port is not returned to service automatically when a port enters the failed condition as a result of crossing a signal failure threshold for both CRC-Monitor and Symbol-Monitor. Because the port is operationally down without a physical link error monitoring stops. The user may enable the port using the **shutdown** and **no shutdown port** commands. Other port transition functions like clearing the MDA or slot, removing the cable, and other physical link transition functions.

2.11 802.3ah OAM

802.3ah Clause 57 (**efm-oam**) defines the Operations, Administration, and Maintenance (OAM) sublayer, which provides mechanisms useful for monitoring link operation such as remote fault indication and remote loopback control. In general, OAM provides network users the ability to monitor the health of the network and quickly determine the location of failing links or fault conditions. **efm-oam** described in this clause provides data link layer mechanisms that complement applications that may reside in higher layers.

OAM information is conveyed in slow protocol frames called OAM protocol data units (OAMPDUs). OAMPDUs contain the appropriate control and status information used to monitor, test and troubleshoot OAM-enabled links. OAMPDUs traverse a single link, being passed between peer OAM entities, and therefore, are not forwarded by MAC clients (like bridges or switches).

The following **efm-oam** functions are supported:

- **efm-oam** capability discovery
- active and passive modes
- remote failure indication (the handling of critical link events such as link fault, dying gasp, and so on)
- loopback; a mechanism is provided to support a data link layer frame-level loopback mode. Both remote and local loopback modes are supported.
- **efm-oam** PDU tunneling
- high resolution timer for **efm-oam** in 100 ms interval (minimum)
- **efm-oam** link monitoring
- non-zero Vendor Specific Information Field, the 32-bit field is encoded using the format 00:PP:CC:CC and references TIMETRA-CHASSIS-MIB.

– **00**

This must be zeros.

– **PP**

This represents the platform type based on the installed IOM from `tmnxHwEquippedPlatform`. 7450 ESS deployments may yield different platform values in the same chassis. Because this is IOM-specific, the IOM's unique hardware ID (`tmnxCardHwIndex`) must be included to retrieve the correct value.

– **CC:CC**

This represents the chassis type index value from `tmnxChassisType` which is indexed in `tmnxChassisTypeTable`. The table identifies the specific chassis backplane.

The value 00:00:00:00 is sent for all releases that do not support the non-zero value or are unable to identify the required elements. There is no decoding of the peer or local vendor information fields on the network element. The hexadecimal value is included in the **show port *port-id* ethernet efm-oam** output.

When the **efm-oam** protocol fails to negotiate a peer session or encounters a protocol failure following an established session the *Port State* enters the *Link Up* condition. This port state is used by many protocols to indicate the port is administratively UP and there is physical connectivity but a protocol, such as **efm-oam**, has caused the ports operational state to enter a DOWN state. A reason code has been added to help discern if the **efm-oam** protocol is the underlying reason for the Link Up condition.

```
show port
=====
```

```

Ports on Slot 1
=====
Port      Admin Link Port  Cfg  Oper LAG/  Port Port Port  C/Q/S/XFP/
Id        State State State MTU  MTU  Bndl Mode Encp Type  MDIMDX
-----
1/1/1     Down  No   Down   1578 1578  - netw null xcme
1/1/2     Down  No   Down   1578 1578  - netw null xcme
1/1/3     Up    Yes  Link Up 1522 1522  - accs qinq xcme
1/1/4     Down  No   Down   1578 1578  - netw null xcme
1/1/5     Down  No   Down   1578 1578  - netw null xcme
1/1/6     Down  No   Down   1578 1578  - netw null xcme

# show port 1/1/3
=====
Ethernet Interface
=====
Description      : 10/100/Gig Ethernet SFP
Interface        : 1/1/3
Link-level       : Ethernet
Admin State      : up
Oper State       : down
Reason Down     : efm0amDown
Physical Link    : Yes
Single Fiber Mode : No
IfIndex          : 35749888
Last State Change : 12/18/2012 15:58:29
Last Cleared Time : N/A
Phys State Chng Cnt: 1

Oper Speed       : N/A
Config Speed    : 1 Gbps
Oper Duplex     : N/A
Config Duplex   : full

MTU              : 1522
Min Frame Length : 64 Bytes
Hold time up    : 0 seconds
Hold time down  : 0 seconds
DDM Events      : Enabled

Configured Mode  : access
Dot1Q Ethertype : 0x8100
PBB Ethertype   : 0x88e7
Ing. Pool % Rate : 100
Ing. Pool Policy : n/a
Egr. Pool Policy : n/a
Net. Egr. Queue Pol: default
Egr. Sched. Pol : n/a
Auto-negotiate  : true
Oper Phy-tx-clock : not-applicable
Accounting Policy : None
Acct Plcy Eth Phys : None
Egress Rate     : Default
Load-balance-algo : Default

Encap Type      : QinQ
QinQ Ethertype  : 0x8100
Egr. Pool % Rate : 100

MDI/MDX         : unknown

Collect-stats   : Disabled
Collect Eth Phys : Disabled
Ingress Rate    : Default
LACP Tunnel     : Disabled

Down-when-looped : Disabled
Loop Detected    : False
Use Broadcast Addr : False

Keep-alive      : 10
Retry           : 120

Sync. Status Msg. : Disabled
Tx DUS/DNU       : Disabled
SSM Code Type    : sdh

Rx Quality Level : N/A
Tx Quality Level : N/A
ESMC Tunnel      : Disabled

Down On Int. Error : Disabled

CRC Mon SD Thresh : Disabled
CRC Mon SF Thresh : Disabled
CRC Mon Window    : 10 seconds

Configured Address : d8:ef:01:01:00:03
Hardware Address   : d8:ef:01:01:00:03

```

The user also has the opportunity to decouple the **efm-oam** protocol from the port state and operational state. In cases where a user wants to remove the protocol, monitor the protocol only, migrate, or make

changes the **ignore-efm-state** can be configured in the **port>ethernet>efm-oam** context. When the **ignore-efm-state** command is configured on a port the protocol continues as normal. However, any failure in the protocol state machine (discovery, configuration, time-out, loops, and so on) does not impact the port on which the protocol is active and the optional ignore command is configured. There is only a protocol warning message if there are issues with the protocol. The default behavior when this optional command is not configured means the port state is affected by any **efm-oam** protocol fault or clear conditions. Adding and removing this optional ignore command immediately represents the *Port State* and *Oper State* based on the active configuration. For example, if the **ignore-efm-state** is configured on a port that is exhibiting a protocol error that protocol error does not affect the port state or operational state and there is no *Reason Down* code. If the **ignore-efm-state** is removed from a port with an existing **efm-oam** protocol error, the port transitions to *Link UP*, *Oper Down* with the reason code *efmOamDown*.

2.11.1 OAM events

The Information OAMPDU is transmitted by each peer at the configured intervals. This OAMPDU performs keepalive and critical notification functions. Various local conditions are conveyed through the setting of the Flags field. The following Critical Link Event defined in IEEE 802.3 Section 57.2.10.1 are supported:

- **link fault**

The PHY has determined a fault has occurred in the receive direction of the local DTE.

- **dying gasp**

An unrecoverable local failure condition has occurred.

- **critical event**

An unspecified critical event has occurred.

The local node can set or unset the various Flag fields based on the operational state of the port, shutdown or activation of the efm-oam protocol or locally raised events. These Flag fields maintain the setting for the continuance of a particular event. Changing port conditions, protocol state or operator intervention may impact the setting of these fields in the Information OAMPDU.

A peer processing the Information OAMPDU can take a configured action when one or more of these Flag fields are set. By default, receiving a set value for any of the Flag fields causes the local port to enter the previously mentioned Link Up port state and an event is logged. If this default behavior is not wanted, the operator may choose to log the event without affecting the local port. This is configurable per Flag field using the options under **config>port>ethernet>efm-oam>peer-rdi-rx**.

2.11.1.1 Link monitoring

The EFM-OAM protocol provides the ability to monitor the link for error conditions that may indicate the link is starting to degrade or has reached an error rate that exceeds an acceptable threshold.

Link monitoring can be enabled for three types of frame errors; **errored-frame**, **errored-frame-period** and **errored-frame-seconds**. The **errored-frame** monitor is the number of frame errors compared to the threshold over a window of time. The **errored-frame-period** monitor is the number of frame errors compared to the threshold over a window of number of received packets. This window is checked once per second to see if the window parameter has been reached. The **errored-frame-seconds** monitor is the number of errored seconds compared to the threshold over a window of time. An errored second is any second with a single frame error.

An errored frame is counted when any frame is in error as determined by the Ethernet physical layer, including jabbers, fragments, FCS or CRC and runts. This excludes jumbo frames with a byte count higher than 9212, or any frame that is dropped by the physical layer before reaching the monitoring function.

Each frame error monitor functions independently of other monitors. Each of monitor configuration includes an optional signal degrade threshold **sd-threshold**, a signal failure threshold **sf-threshold**, a **window** and the ability to communicate failure events to the peer by setting a Flag field in the Information OAMPDU or the generation of the Event Notification OAMPDU, **event-notification**. The parameters are uniquely configurable for each monitor.

A degraded condition is raised when the configured signal degrade **sd-threshold** is reached. This provides a first level log only action indicating a link could become unstable. This event does not affect the port state. The critical failure condition is raised when the configured **sf-threshold** is reached. By default, reaching the signal failure threshold causes the port to enter the *Link Up* condition unless the local signal failure **local-sf-action** has been modified to a **log-only** action. Signal degrade conditions for a monitor in signal failed state is suppressed until the signal failure has been cleared.

The initial configuration or the modification of either of the threshold values take effect in the current window. When a threshold value for a monitor is modified, all active local events for that specific monitor are cleared. The modification of the threshold acts the same as the **clear** command described later in this section.

Notification to the peer is required to ensure the action taken by the local port detecting the error and its peer are synchronized. If peers do not take the same action then one port may remain fully operational while the other enters a non-operational state. These threshold crossing events do not shutdown the physical link or cause the protocol to enter a non-operational state. The protocol and network element configuration is required to ensure these asymmetrical states do not occur. There are two options for exchanging link and event information between peers; Information OAMPDU and the Event Notification OAMPDU.

As discussed earlier, the Information OAMPDU conveys link information using the Flags field; dying gasp, critical link and link fault. This method of communication has a number of significant advantages over the Event Notification OAMPDU. The Information OAMPDU is sent at every configured **transmit-interval**. This allows the most recent information to be sent between peers, a critical requirement to avoid asymmetrical forwarding conditions. A second major advantage is interoperability with devices that do not support Link Monitoring and vendor interoperability. This is the lowest common denominator that offers a robust communication to convey link event information. Since the Information OAMPDU is already being sent to maintain the peering relationship this method of communication adds no additional overhead. The **local-sf-action** options allow the dying gasp and critical event flags to be set in the Information OAMPDU when a signal failure threshold is reached. It is suggested that this be used in place of or in conjunction with Event Notification OAMPDU.

Event Notification OAMPDU provides a method to convey very specific information to a peer about various Link Events using Link Event TLVs. A unique Event Notification OAMPDU is generated for each unique frame error event. The intention is to provide the peer with the Sequence Number, Event Type, Timestamp, and the local information that caused the generation of the OAMPDU; window, threshold, errors and error running total and event running total specific to the port.

- **sequence number**

The unique identification indicating a new event.

- **window**

The size of the unique measurement period for the error type. The window is only checked at the end. There is no mid-window checking.

- **threshold**

The value of the configured sf-threshold.

- **errors**

The errors counted in that specific window.

- **error running total**

The number of errors accumulated for that event type since monitoring started and the protocol and port have been operational or a reset function has occurred.

- **event running total**

The number of events accumulated for that event type since the monitoring started and the protocol and port have been operational.

By default, the Event Notification OAMPDU is generated by the network element detecting the signal failure event. The Event Notification OAMPDU is sent only when the initial frame event occurs. No Event Notification OAMPDU is sent when the condition clears. A port that has been operationally affected as a result of a Link Monitoring frame error event must be recovered manually. The typical recovery method is to shutdown the port and no shutdown the port. This clears all events on the port. Any function that affects the port state, physical fiber pull, soft or hard reset functions, protocol restarts, and so on, also clears all local and remote events on the affected node experiencing the operation. None of these frame errors recovery actions cause the generation of the Event Notification OAMPDU. If the chosen recovery action is not otherwise recognized by the peer and the Information OAMPDU Flag fields have not been configured to maintain the current event state, there is a high probability that the ports have different forwarding states, notwithstanding any higher level protocol verification that may be in place.

A burst of between one and five Event Notification OAMPDU packets may be sent. By default, only a single Event Notification OAMPDU is generated, but this value can be changed under the **local-sf-action** context. An Event Notification OAMPDU is only processed if the peer had previously advertised the EV capability. The EV capability is an indication the remote peer supports link monitoring and may send the Event Notification OAMPDU.

The network element receiving the Event Notification OAMPDU uses the values contained in the Link event TLVs to determine if the remote node has exceeded the failure threshold. The locally configured action determines how and if the local port is affected. By default, processing of the Event Notification OAMPDU is log only and does not affect the port state. By default, processing of the Information OAMPDU Flag fields is port affecting. When Event Notification OAMPDU has been configured as port affecting on the receiving node, action is only taken when errors are equal to or above the threshold and the threshold value is not zero. No action is taken when the errors value is less than the threshold or the threshold is zero.

Symbol error, **errored-symbols**, monitoring is also supported but requires specific hardware revisions and the appropriate code release. The symbol monitor differs from the frame error monitors. Symbols represent a constant load on the Ethernet wire whether service frames are present or not. This means the optional signal degrade threshold **sd-threshold** has an additional purpose when configured as part of the symbol error monitor. When the signal degrade threshold is not configured, the symbol monitor acts similar to the frame error monitors, requiring manual intervention to clear a port that has been operationally affected by the monitor. When the optional signal degrade threshold is configured, it again represents the first level warning. However, it has an additional function as part of the symbol monitor. If a signal failure event has been raised, the configured signal degrade threshold becomes the equivalent to a lowering threshold. If a subsequent window does not reach the configured signal degrade threshold then the previous event is cleared and the previously affected port is returned to service without operator intervention. This return to service automatically clears any previously set Information OAMPDU Flags fields set as a result of the signal failure threshold. The Event Notification OAMPDU is generated with the symbol error Link TLV that

contains an error count less than the threshold. This indicates to the peer that initial problem has been resolved and the port should be returned to service.

The **errored-symbol** window is a measure of time that is automatically converted into the number of symbols for that specific medium for that period of time. The standard MIB entries "dot3OamErrSymPeriodWindowHi" and "dot3OamErrSymPeriodWindowLo" are marked as read-only instead of read-write. These values cannot be configured directly. The configuration of the **window** converts the time and programs the two MIB values in an appropriate manner. Both the configured **window** and the number of symbols are displayed under the **show port port-id ethernet efm-oam** command.

```
show port 1/1/1 ethernet efm-oam
=====
Ethernet Oam (802.3ah)
=====
Admin State          : up
Oper State           : operational
Mode                 : active
Pdu Size             : 1518
Config Revision      : 0
Function Support     : LB
Transmit Interval    : 1000 ms
Multiplier           : 5
Hold Time            : 0
Tunneling            : false
Loop Detected        : false
Grace Tx Enable      : true (inactive)
Grace Vendor OUI     : 00:16:4d
Dying Gasp on Reset : true (inactive)
Soft Reset Tx Act    : none
Trigger Fault        : none
Vendor OUI           : 00:16:4d (alu)
Vendor Info          : 00:01:00:02
Peer Mac Address     : d8:1c:01:02:00:01
Peer Vendor OUI      : 00:16:4d (alu)
Peer Vendor Info     : 00:01:00:02
Peer Mode            : active
Peer Pdu Size        : 1518
Peer Cfg Revision    : 0
Peer Support         : LB
Peer Grace Rx        : false
Loopback State       : None
Loopback Ignore Rx   : Ignore
Ignore Efm State     : false
Link Monitoring      : disabled
Peer RDI Rx
  Critical Event      : out-of-service
  Dying Gasp          : out-of-service
  Link Fault          : out-of-service
  Event Notify        : log-only
Local SF Action
  Event Burst         : 1
  Port Action         : out-of-service
  Dying Gasp          : disabled
  Critical Event      : disabled
Discovery
  Ad Link Mon Cap    : yes
Errored Frame
  Enabled             : no
  Event Notify        : enabled
  SF Threshold        : 1
  SD Threshold        : disabled (0)
  Window              : 10 ds
Errored Symbol Period
  Enabled             : no
Errored Frame Period
  Enabled             : no
  Event Notify        : enabled
  SF Threshold        : 1
  SD Threshold        : disabled (0)
  Window              : 1488095 frames
Errored Frame Seconds Summary
  Enabled             : no
```

```

Event Notify      : enabled          Event Notify      : enabled
SF Threshold      : 1                SF Threshold      : 1
SD Threshold      : disabled (0)     SD Threshold      : disabled (0)
Window (time)     : 10 ds            Window            : 600 ds
Window (symbols)  : 125000000
=====
Active Failure Ethernet OAM Event Logs
=====
Number of Logs : 0
=====
Ethernet Oam Statistics
=====
                                     Input          Output
-----
Information                238522          238522
Loopback Control              0                0
Unique Event Notify          0                0
Duplicate Event Notify        0                0
Unsupported Codes            0                0
Frames Lost                   0                0
=====

```

A **clear** command "**clear port *port-id* ethernet efm-oam events [local | remote]**" has been added to clear port affecting events on the local node on which the command is issued. When the optional [**local | remote**] options are omitted, both local and remote events are cleared for the specified port. This command is not specific to the link monitors as it clears all active events. When local events are cleared, all previously set Information OAMPDU Flag fields are cleared regardless of the cause of the event that set the Flag field.

In the case of symbol errors only, if Event Notification OAMPDU is enabled for symbol errors and a local symbol error signal failure event exists at the time of the clear, the Event Notification OAMPDU generates with an error count of zero and a threshold value reflecting the local signal failure threshold. An error value lower than the threshold value indicates the local node is not in a signal failed state. The Event Notification OAMPDU is not generated in the case where the **clear** command clears local frame error events. This is because frame error event monitors only acts on an Event Notification OAMPDU when the error value is higher than the threshold value, a lower value is ignored. As stated previously, there is no automatic return to service for frame errors.

If the clear command clears remote events, events conveyed to the local node by the peer, no notification is generated to the peer to indicate a clear function has been performed. Since the Event Notification OAMPDU is only sent when the initial event was raised, there is no further Event Notification and blackholes can result. If the Information OAMPDU Flag fields are used to ensure a constant refresh of information, the remote error is reinstated as soon as the next Information OAMPDU arrives with the appropriate Flag field set.

Local and remote EFM-OAM port events are stored in the EFM-OAM event logs. These logs maintain and display active and cleared signal failure degrade events. These events are interacting with the EFM-OAM protocol. This logging is different than the time stamped events for information logging purposes included with the system log. To view these events, the **event-log** option has been added to the **show port port-id ethernet efm-oam** command. This includes the location, the event type, the counter information or the decoded Network Event TLV information, and if the port has been affected by this active event. A maximum of 12 port events are retained. The first three indexes are reserved for the three Information Flag fields, dying gasp, critical link, and link fault. The other nine indexes maintain the current state for the various error monitors in a most recent behavior and events can wrap the indexes, dropping the oldest event.

In mixed environments where Link Monitoring is supported on one peer but not the other the following behavior is normal, assuming the Information OAMPDU has been enabled to convey the monitor fault

event. The arriving Flag field fault triggers the EFM-OAM protocol on the receiving unsupportive node to move from operational to "send local and remote". The protocol on the supportive node that set the Flag field to convey the fault enters the "send local and remote ok" state. The supportive node maintains the Flag field setting until the condition has cleared. The protocol recovers to the operational state once the original event has cleared; assuming no other fault on the port is preventing the negotiation from progressing. If both nodes were supportive of the Link Monitoring process, the protocol would remain operational.

In summary, Link monitors can be configured for frame and symbol monitors (specific hardware only). By default, Link Monitoring and all monitors are shutdown. When the Link Monitoring function is enabled, the capability (EV) is advertised. When a monitor is enabled, a default window size and a default signal failure threshold are activated. The local action for a signal failure threshold event is to shutdown the local port. Notification is sent to the peer using the Event Notification OAMPDU. By default, the remote peer does not take any port action for the Event Notification OAMPDU. The reception is only logged. It is suggested the operator evaluate the various defaults and configure the **local-sf-action** to set one of the Flag fields in the Information OAMPDU using the **info-notifications** command options when fault notification to a peer is required. Non-Nokia vendor specific information is not processed.

2.11.1.1 Capability advertising

A supported capability, sometimes requiring activation, is advertised to the peer. The EV capability is advertisement when Link Monitoring is active on the port. This can be disabled using the optional command **no link-monitoring** under the **config>port>ethernet>efm-oam>discovery>advertise-capabilities** context.

2.11.2 Remote loopback

EFM OAM provides a link-layer frame loopback mode that can be remotely controlled.

To initiate remote loopback, the local EFM OAM client sends a loopback control OAM PDU by enabling the OAM remote-loopback command. After receiving the loopback control OAM PDU, the remote OAM client puts the remote port into local loopback mode.

To exit remote loopback, the local EFM OAM client sends a loopback control OAM PDU by disabling the OAM remote-loopback command. After receiving the loopback control OAM PDU, the remote OAM client puts the port back into normal forwarding mode.

During remote loopback test operation, all frames except EFM OAM PDUs are dropped at the local port for the receive direction, where remote loopback is enabled. If local loopback is enabled, then all frames except EFM OAM PDUs are dropped at the local port for both the receive and transmit directions. This behavior may result in many protocols (such as STP or LAG) resetting their state machines.

When a port is in loopback mode, service mirroring does not work if the port is a mirror-source or a mirror-destination.

2.11.3 802.3ah OAM PDU tunneling for Epipe service

Nokia routers support 802.3ah. Customers who subscribe to Epipe service treat the Epipe as a wire, so they demand the ability to run 802.3ah between their devices which are located at each end of the Epipe.

This feature only applies to port-based Epipe SAPs because 802.3ah runs at port level not VLAN level. Hence, such ports must be configured as null encapsulated SAPs.

When OAM PDU tunneling is enabled, 802.3ah OAM PDUs received at one end of an Epipe are forwarded through the Epipe. 802.3ah can run between devices that are located at each end of the Epipe. When OAM PDU tunneling is disabled (by default), OAM PDUs are dropped or processed locally according to the **efm-oam** configuration (**shutdown** or **no shutdown**).

Enabling 802.3ah for a specific port and enabling OAM PDU tunneling for the same port are mutually exclusive. Enforcement is performed at the CLI level.

2.11.3.1 802.3ah grace announcement

The SR OS implementation of the EFM-OAM protocol supports vendor-specific soft reset graceful recovery. This feature is not enabled by default. It is configured using the **grace-tx-enable** command under the **config>system>ethernet>efm-oam** and the **config>port>ethernet>efm-oam** contexts. When this feature is enabled, the EFM-OAM protocol does not enter a non-operational state when both nodes acknowledge the grace function. The ports associated with the hardware that has successfully executed the soft reset clears all local and remote events. The peer that acknowledges the graceful restart procedure for EFM-OAM clears all remote events that it has received from the peer that performed the soft reset. The local events are not cleared on the peer that has not undergone soft reset. The Information OAM PDU Flag fields are critical in propagating the local event to the peer. The Event Notification OAM PDU, which is only sent when the event is initially raised, is not sent.

A vendor-specific Grace TLV is included in the Information PDU generated as part of the 802.3ah OAM protocol when a network element undergoes an ISSU function. Nodes that support the Soft Reset messaging functions allow the local node to generate the Grace TLV.

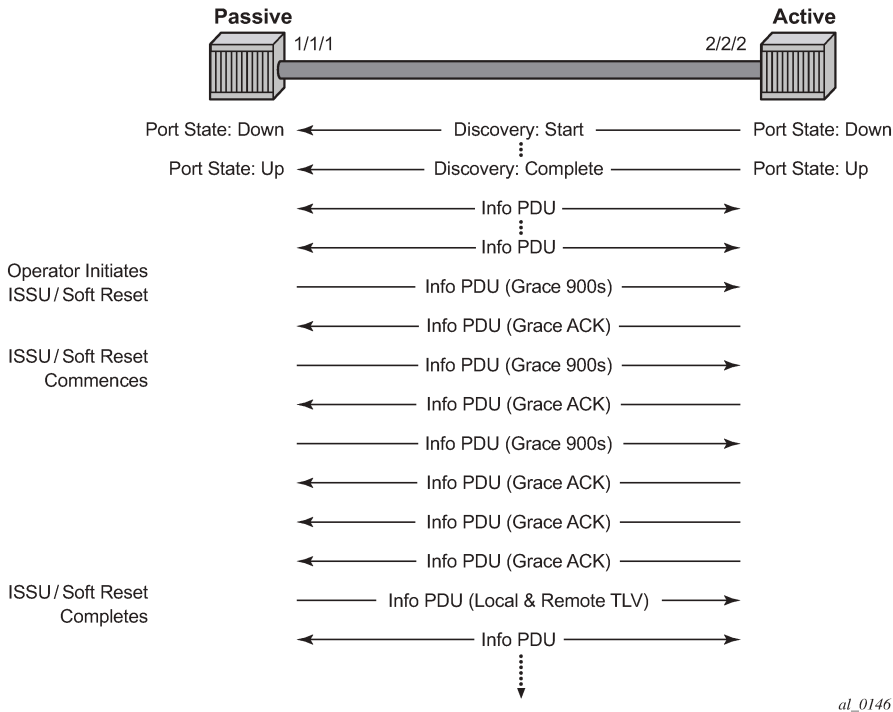
The Grace TLV informs a remote peer that the negotiated interval and multiplier should be ignored and the new 900s timeout interval should be used to timeout the session. The peer receiving the Grace TLV must be able to parse and process the vendor-specific messaging.

Use the **grace-tx-enable** command to enable this feature. This command exists at two levels of the hierarchy: system level and port level. By default, this feature is enabled on the port. At the system level, this command defaults to disabled. To enable this feature, both the port and the system commands must be enabled. If either is not enabled, the combination does not allow those ports to generate the vendor specific Grace TLV. This feature must be enabled at both the system and port level before the ISSU or soft reset function. If this feature is enabled during a soft reset or after the ISSU function is already in progress, it has no affect during that window. Both Passive and Active 802.3ah OAM peers can generate the Grace TLV as part of the informational PDU.

There is no CLI command to enable this feature on the receiving node. As long as the receiver understands and can parse the Grace TLV, it enters the grace mode of operation. The Nokia 7750 SR minimum release required to support the reception and processing of the Grace TLV is Release 11.0.R.4.

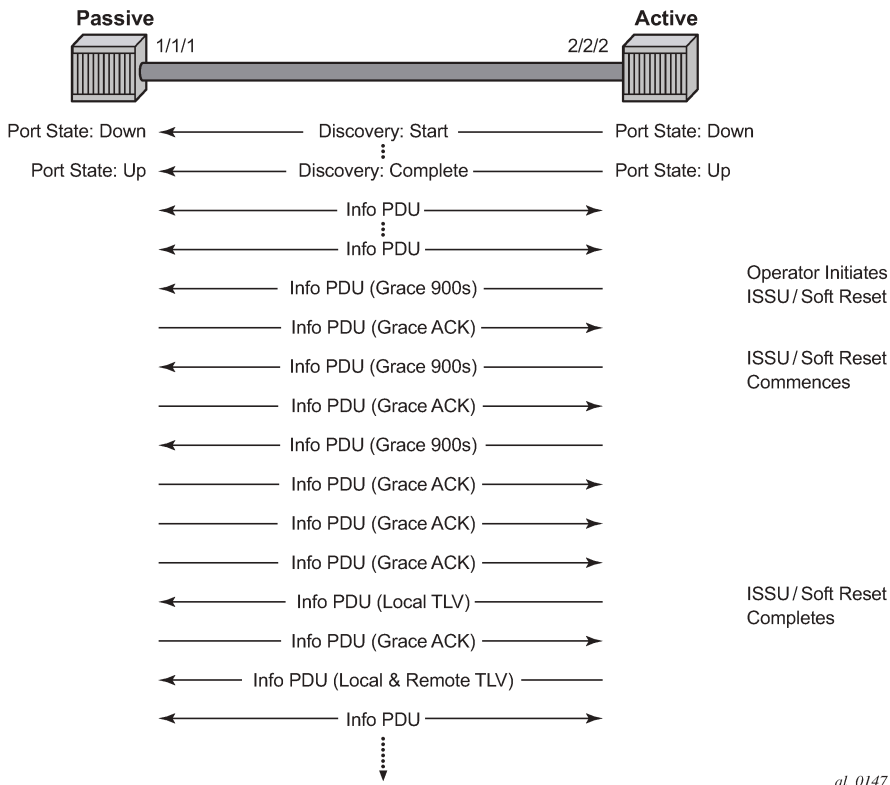
The following basic protocol flows demonstrate the interaction between passive-active and active-active peer combinations supporting the Grace TLV. In [Figure 40: Grace TLV passive node with soft reset](#), the passive node is entering an ISSU on a node that supports soft reset capabilities.

Figure 40: Grace TLV passive node with soft reset



In [Figure 41: Grace TLV active node with soft reset](#), the Active node is experiencing the ISSU function on a node that supports soft reset capabilities.

Figure 41: Grace TLV active node with soft reset



The difference between [Figure 40: Grace TLV passive node with soft reset](#) and [Figure 41: Grace TLV active node with soft reset](#) is subtle but important. When an active node performs this function it generates an Informational TLV with the Local TLV following the successful soft reset. When the active node receives the Information PDU with the Grace Ack, it sends its own Information PDU with both Local and Remote TLV completed, which completes the protocol restart. When a passive node is reset the passive port waits to receive the 802.3ah OAM protocol before sending its own Information PDU with both the Local and Remote TLV, therefore completing the protocol restart.

The renegotiation process allows the node, which experienced the soft reset, to rebuild the session without restarting the session from the discovery phase. This significantly reduces the native protocol impact on data forwarding.

Any situation that could cause the renegotiation to fail forces the protocol to revert to the discovery phase and fail the graceful restart. During a Major ISSU when the EFM-OAM session is held operational by the Grace function, if the peer MAC address of the session changes, there is no log event raised for the MAC address change.

The vendor-specific grace function benefits are realized when both peers support the transmitting, receiving, and processing of the vendor-specific Grace TLV. In the case of mixed code versions, products, or vendor environments, a standard EFM-OAM message to the peer can be used to instruct the peer to treat the session as failed. When the command **dying-gasp-tx-on-reset** is active on a port, the soft reset function triggers ETH-OAM to set the dying gasp flag or critical event flag in the Information OAMPDU. An initial burst of three Informational OAM PDUs is sent using a 1-second spacing, regardless of the protocol interval. The peer may process these flags to affect its port state and take the appropriate action. The control of the local port state where the soft reset is occurring is left to the soft reset function. This EFM-

OAM function does not affect local port state. If the peer has acted on the exception flags and affected its port state, the local node must take an action to inform the upstream nodes that a condition has occurred and forwarding is no longer possible. Routing protocols like ISIS and OSPF overload bits are typically used in routed environments to accomplish this notification.

This feature is similar to **grace-tx-enable**. Intercepting system messaging, when the feature is active on a port (enabled both at the port and at the system level) and when the messaging occurs, is a similar concept. However, because the **dying-gasp-tx-on-reset** command is not a graceful function it is interruptive and service affecting. Using **dying-gasp-tx-on-reset** requires peers to reestablish the peering session from an initial state, not rebuild the state from previous protocol information. The transmission of the dying gasp or the critical event commences when the soft reset occurs and continues for the duration of the soft reset.

If both functions are active on the same port, the **grace-tx-enable** function is preferred if the peer is setting and sending the Vendor OUI to 00:16:4d (ALU) in the Information OAM PDU. In this situation, the dying gasp function is not invoked. A secondary Vendor OUI can be configured using the **grace-vendor-oui oui** command, should an additional Vendor OUI prefer to support the reception, parsing, and processing of the vendor-specific grace message instead of the dying gasp. If only one of those functions is active on the port then that specific function is called. The grace function should not be enabled if the peer Vendor OUI is equal to 00:16:4d (ALU) and the peer does not support the grace function.

ETH-OAM allows the use of the **trigger-fault {dying-gasp | critical-event}** command to generate a fault condition. This sets the appropriate flag fields in the Information OAM PDU and transitions a previously operational local port to Link Up. Removing this command from the configuration stops the flags from being set and allows the port to return to service, assuming no other faults would prevent this resumption of service. In cases where a port must be administratively shut down, this command can be used to signal a peer using the EFM-OAM protocol, and the session should be considered failed.

These features do not support clearing of an IOM that does not trigger a soft reset. IOM clearing is a forceful event that does not trigger graceful protocol renegotiation.

The **show** commands are enhanced to help users determine the state of the 802.3ah OAM Grace function and whether the peer is generating or receiving the Grace TLV.

System level information can be viewed using the **show system info** command, as shown in the following example output:

```
show system information
=====
System Information
=====
System Name       : system-name
System Type       : 7750 SR-12
System Version    : 11.0r4
System Contact    :
System Location   :
System Coordinates :
System Active Slot : A
System Up Time    : 62 days, 20:29:48.96 (hr:min:sec)

...snip...

EFM OAM Grace Tx Enable: False
=====
```

The system-level EFM OAM Grace Tx Enable field displays one of the following two states.

- **false**

The system-level functionality is not enabled. Grace is not generated on any ports regardless of the state of the option on the individual ports.

- **true**

The system-level functionality is enabled and the determination of whether to send grace is based on the state of the option configured at the port level.

Individual ports also contain information about the current port configuration and whether the Grace TLV is being sent or received.

The port-level Grace Tx Enable field has two enable states with the current state in brackets to the right.

- **false**

The port-level functionality is not enabled. Grace is not generated on the port regardless of the state of the option at the system level.

- **true**

The port-level functionality is enabled and the determination of whether to send grace is based on the state of the option configured at the system level. The following applies:

- (inactive) Not currently sending Grace TLV
- (active) Currently sending the Grace TLV as part of the Information PDU

The port-level Peer Grace Rx field displays one of the following two states.

- **false**

The port is not receiving Grace TLV from the peer.

- **true**

The port is receiving Grace TLV from the peer.

2.12 MTU configuration guidelines

Observe the following general rules when planning your service and physical MTU configurations:

- The router must contend with MTU limitations at many service points. The physical (access and network) port, service, and SDP MTU values must be individually defined.
- Identify the ports that are designated as network ports intended to carry service traffic.
- MTU values should not be modified frequently.
- MTU values must conform to both of the following conditions:
 - The service MTU must be less than or equal to the SDP path MTU.
 - The service MTU must be less than or equal to the access port (SAP) MTU.
- When the network group encryption (NGE) feature is enabled, additional bytes because of NGE packet overhead must be considered. See the “NGE Packet Overhead and MTU Considerations” section in the *7450 ESS, 7750 SR, 7950 XRS, and VSR Services Overview Guide* for more information.

2.12.1 Default MTU values

[Table 26: MTU default values](#) shows the default MTU values which are dependent upon the (sub-) port type, mode, and encapsulation.

Table 26: MTU default values

Port type	Mode	Encap type	Default (bytes)
Ethernet	access	null	1514
Ethernet	access	dot1q	1518
Fast Ethernet	network	—	1514
Other Ethernet	network	—	9212 ⁵

2.12.2 Modifying MTU defaults

MTU parameters must be modified on the service level as well as the port level.

- The service-level MTU parameters configure the service payload (Maximum Transmission Unit – MTU) in bytes for the service ID overriding the service-type default MTU.
- The port-level MTU parameters configure the maximum payload MTU size for an Ethernet port or SONET/SDH SONET path (sub-port) or TDM port/channel, or a channel that is part of a multilink bundle or LAG.

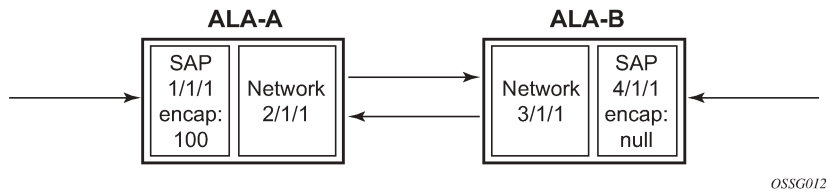
The default MTU values must be modified to ensure that packets are not dropped because of frame size limitations. The service MTU must be less than or equal to both the SAP port MTU and the SDP path MTU values. When an SDP is configured on a network port using default port MTU values, the operational path MTU can be less than the service MTU. In this case, enter the `show service sdp` command to check the operational state. If the operational state is down, then modify the MTU value accordingly.

2.12.3 Configuration example

In order for the maximum length service frame to successfully travel from a local ingress SAP to a remote egress SAP, the MTU values configured on the local ingress SAP, the SDP (GRE or MPLS), and the egress SAP must be coordinated to accept the maximum frame size the service can forward. For example, the targeted MTU values to configure for a distributed Epipe service (ALA-A and ALA-B) are shown in [Figure 42: MTU configuration example](#).

⁵ The default MTU for Ethernet ports other than Fast Ethernet is actually the lesser of 9212 and any MTU limitations imposed by hardware which is typically 16K.

Figure 42: MTU configuration example



Because ALA-A uses Dot1q encapsulation, the SAP MTU must be set to 1518 to be able to accept a 1514 byte service frame (see [Table 26: MTU default values](#) for MTU default values). Each SDP MTU must be set to at least 1514 as well. If ALA-A's network port (2/1/1) is configured as an Ethernet port with a GRE SDP encapsulation type, then the MTU value of network ports 2/1/1 and 3/1/1 must each be at least 1556 bytes (1514 MTU + 28 GRE/Martini + 14 Ethernet). Finally, the MTU of ALA-B's SAP (access port 4/1/1) must be at least 1514, as it uses null encapsulation.

[Table 27: MTU configuration example values](#) shows example MTU configuration values.

Table 27: MTU configuration example values

	ALA-A		ALA-B	
	Access (SAP)	Network	Network	Access (SAP)
Port (slot/MDA/port)	1/1/1	2/1/12	3/1/1	4/1/1
Mode or ECAP-type	dot1q	network	network	null
MTU	1518	1556	1556	1514

2.13 Deploying preprovisioned components

When a card, MDA, XCM or XMA is installed in a preprovisioned slot, the device detects discrepancies between the preprovisioned card type configurations and the types actually installed. Similarly for cards or XMAs that have license levels, the device shall also detect discrepancies between the provisioned level and the level of the installed card or XMA. Error messages display if there are inconsistencies and the card does not initialize.

When the correct preprovisioned cards are installed into the appropriate chassis slot, alarm, status, and performance details display.

2.14 Setting fabric speed

The **set-fabric-speed** command under the **tools>perform>system** context sets the fabric speed for the 7750 SR-7/12/12e, 7450 ESS-7/12, and 7950 XRS-20/20e and is associated with the FP3 or newer generation of switch fabric.

2.14.1 7750 SR-7/12/12e and 7450 ESS-7/12

The **fabric-speed-a** parameter enables the chassis to operate at the following speeds using N+1 switch fabric redundancy:

- up to 100 Gb/s per slot for the 7450 ESS-7/12 and the 7750 SR-7/12
- up to 200 Gb/s per slot for the 7750 SR-12e

This parameter is compatible with SFM5, which allows a mixture of FP2- and FP3-based cards to coexist. This fabric speed is displayed as "6 Gig" in the **show>chassis** output.

The **fabric-speed-b** parameter enables the chassis to operate at the following speeds using N+1 switch fabric redundancy:

- up to 200 Gb/s per slot for the 7450 ESS-7/12 and the 7750 SR-7/12
- up to 400 Gb/s per slot for the 7750 SR-12e

This parameter is compatible with SFM5. All cards in the system must be FP3-based. The system does not support any FP2-based cards when the chassis is set to **fabric-speed-b**. This fabric speed is displayed as "10 Gig" in the **show>chassis** output.



Note: To set **fabric-speed-b** for the 7750 SR-7/12 and 7450 ESS-7/12, the chassis must support 200 Gb/s per slot capability. To check if the chassis supports **fabric-speed-b**, execute the **show>system>switch-fabric** command and verify that the "chassis is 200G/slot capable" message is displayed.

The **fabric-speed-c** parameter enables the use of both FP3- and FP4-based cards and is compatible with SFM6. This speed is mandatory if FP4 cards are used. The performance of FP3 cards is the same as **fabric-speed-b**. This fabric speed is displayed as "S4" in the **show>chassis** output.



Note: To set **fabric-speed-c** for the 7750 SR-7/12/12e and 7450 ESS-7/12, the chassis must support S4 fabric capability. To check if the chassis supports **fabric-speed-c**, execute the **show>system>switch-fabric** command and verify that the "chassis is s4 fabric capable" message is displayed.

If no fabric speed has previously been set on the chassis, either **fabric-speed-a** or **fabric-speed-b** is automatically selected and set based on the physically equipped cards. If no FP2-based cards are physically equipped and the chassis meets the required capabilities, **fabric-speed-b** is selected; if not, **fabric-speed-a** is selected.

2.14.2 7950 XRS-20/20e

The **none** parameter enables the 7950 XRS-20/20e to use only FP3 cards. This fabric speed is compatible with the following SFMs: **sfm-x20**, **sfm-x20-b**, and **sfm-x20s-b**. When the **none** parameter is used, there is no fabric speed configured or displayed in the **show>chassis** output.

The **fabric-speed-c** parameter enables the use of both FP3- and FP4-based cards and is compatible with **sfm2-x20s**. For the 7950 XRS-20/20e, the performance of FP3 cards is the same as the **none** parameter. This fabric speed is displayed as "S4" in the **show>chassis** output.

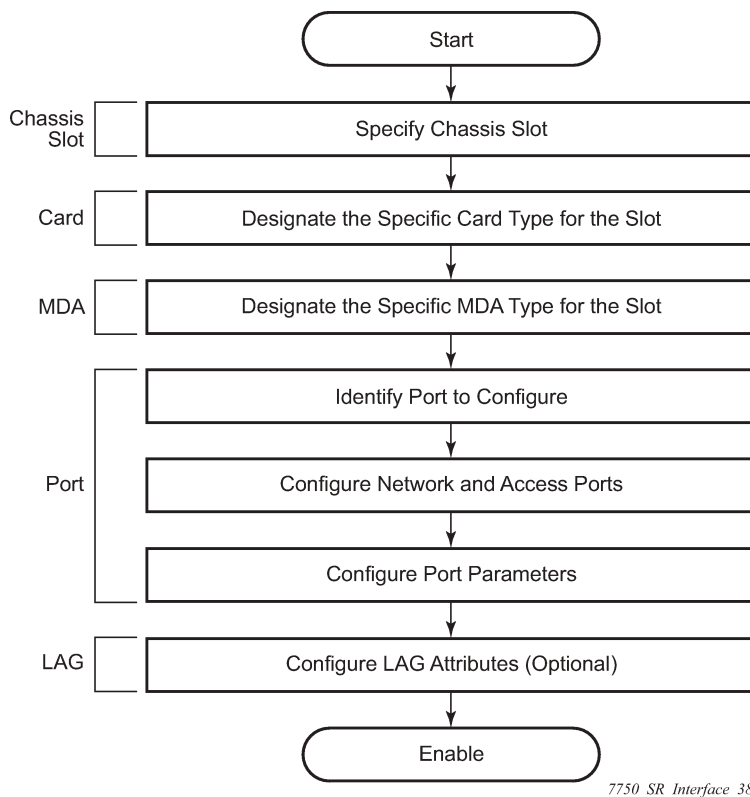
If no fabric speed has previously been set on the chassis, the **none** parameter is set as the default.

The **tools>perform>system>set-fabric-speed** command is not used in 7950 XRS-40 systems.

2.15 Configuration process overview

Figure 43: Slot, card, MDA, and port configuration and implementation flow displays the process to provision chassis slots, cards, MDAs, and ports.

Figure 43: Slot, card, MDA, and port configuration and implementation flow



2.16 Configuration notes

The following information describes provisioning restrictions:

- If a card or MDA type is installed in a slot provisioned for a different type, the card does not initialize.
- A card or MDA installed in an unprovisioned slot remains administratively and operationally down until the card type and MDA is specified.
- Ports cannot be provisioned until the slot, card and MDA type are specified.
- cHDLC does not support HDLC windowing features, nor other HDLC frame types such as S-frames.
- cHDLC operates in the HDLC Asynchronous Balanced Mode (ABM) of operation.
- APS configuration rules:
 - A physical port (either working or protection) must be shut down before it can be removed from an APS group port.

- For a single-chassis APS group, a working port must be added first. Then a protection port can be added or removed at any time.
- A protection port must be shut down before being removed from an APS group.
- A path cannot be configured on a port before the port is added to an APS group.
- A working port cannot be removed from an APS group until the APS port path is removed.
- When ports are added to an APS group, all path-level configurations are available only on the APS port level and configuration on the physical member ports are blocked.
- For APS-protected bundles, all members of a working bundle must reside on the working port of an APS group. Similarly all members of a protecting bundle must reside on the protecting circuit of that APS group.

2.17 Configuring physical ports with CLI

This section provides information to configure cards, MDAs, and ports.

2.17.1 Preprovisioning guidelines

SR OSs have a console port, either located on the CPM or CCM, or integrated into the chassis (on the 7750 SR-c4 models), to connect terminals to the router.

Configure parameters from a system console connected to a router console port, using Telnet to access a router remotely or SSH to open a secure shell connection.

2.17.1.1 Predefining entities

To initialize a card, the chassis slot, line card type, and MDA type must match the preprovisioned parameters. In this context, preprovisioning means to configure the entity type (such as the card type, MDA type, port, and interface) that is planned for a chassis slot, card, or MDA. Preprovisioned entities can be installed but not enabled or the slots can be configured but remain empty until populated. Provisioning means that the preprovisioned entity is installed and enabled.

You can:

- Preprovision ports and interfaces after the line card and MDA types are specified.
- Install line cards in slots with no pre-configuration parameters specified. After the card is installed, the card and MDA types must be specified.
- Install a line card in a slot provisioned for a different card type (the card does not initialize). The existing card and MDA configuration must be deleted and replaced with the current information.

2.17.1.2 Preprovisioning a port

Before a port can be configured, the slot must be preprovisioned with an allowed card type and the MDA must be preprovisioned with an allowed MDA type. Some recommendations to configure a port include:

- **Ethernet**

- Configure an access port for customer facing traffic on which services are configured.
- An encapsulation type may be specified to distinguish services on the port or channel. Encapsulation types are not required for network ports.
- To configure an Ethernet access port, see [Ethernet access port](#).
- **SONET/SDH**
 - SONET/SDH can be used only when configuring OC-3 and OC-12 SONET paths.
 - To configure a SONET path, see [Configuring SONET/SDH port parameters](#).
 - Configure a network port or channel to participate in the service provider transport or infrastructure network.
 - Accounting policies can only be associated with network ports/channels and Service Access Ports (SAPs). Accounting policies are configured in the **config>log> accounting-policy** context.
 - To configure an Ethernet network port, see [Ethernet network port](#).
- **channelized**

Channelized ports can only be configured on TDM satellites such as the channelized OC-3-SFP or channelized OC-12-SFP.

2.17.1.3 Maximizing bandwidth use

After ports are preprovisioned, Link Aggregation Groups (LAGs) can be configured to increase the bandwidth available between two nodes.

All physical links or channels in a LAG/bundle combine to form one logical connection. A LAG/bundle also provides redundancy in case one or more links that participate in the LAG/bundle fail. To configure channelized port for Sonet/SDH high speed channels (ASAP MDAs only), see [Configuring SONET/SDH port parameters](#). For command syntax for LAG, see [Configuring LAG parameters](#).

2.17.2 Basic configuration

The most basic configuration must specify the following:

- chassis slot
- line card type (must be an allowed card type)
- MDA slot
- MDA (must be an allowed MDA type)
- specific port to configure

The following is an example of card configuration for the 7750 SR:

- **classic CLI configuration**

```
A:Dut-B# admin display-config | match "Card Configuration" post-lines 22
echo "Card Configuration"
#-----
card 6
    card-type iom4-e
    no shutdown
exit
```



```

card 7
  card-type iom4-e
  mda 1
    mda-type me10-10gb-sfp+
    no shutdown
  exit
  mda 2
    mda-type me1-100gb-cfp2
    no shutdown
  exit
  no shutdown
exit
card 8
  card-type iom4-e
  no shutdown
exit
#-----

```

- **MD-CLI configuration**

```

A:admin@Dut-B# info
configure {
  card 6 {
    card-type iom4-e
    mda 1 {
      mda-type me1-100gb-cfp2
    }
    fp 1 {
    }
  }
  card 7 {
    card-type iom4-e
    mda 1 {
      mda-type me10-10gb-sfp+
    }
    mda 2 {
      mda-type me1-100gb-cfp2
    }
    fp 1 {
    }
  }
  card 8 {
    card-type iom4-e
    mda 1 {
      mda-type me10-10gb-sfp+
    }
  }
}

```

The following is an example of card configurations for the 7950 XRS:

```

A:7950 XRS-20# configure card 1
A:7950 XRS-20>config>card# info
-----
card-type xcm-x20
mda 1
  mda-type cx20-10g-sfp
  no shutdown
exit
mda 2
  mda-type cx2-100g-cfp
  no shutdown
exit
no shutdown

```

2.17.3 Common configuration tasks

The following sections are basic system tasks that must be performed.

2.17.3.1 Configuring cards and MDAs

Card configurations include a chassis slot designation. A slot must be preconfigured with the type of cards and MDAs which are allowed to be provisioned.

The following example shows card and MDA configurations for the 7750 SR or 7450 ESS:

- **classic CLI commands**

```
echo "Card Configuration"
#-----
card 8
  card-type iom4-e
  mda 1
    mda-type me10-10gb-sfp+
    no shutdown
  exit
  mda 2
    mda-type me1-100gb-cfp2
    no shutdown
  exit
  no shutdown
  exit
#-----
```

- **MD-CLI commands**

```
(gl)[/]
A:admin@Dut-E# configure card 8

(gl)[/configure card 8]
A:admin@Dut-E# info
  card-type iom4-e
  mda 1 {
    mda-type me10-10gb-sfp+
  }
  mda 2 {
    mda-type me1-100gb-cfp2
  }
  fp 1 {
  }
}
```

The following example shows card configurations for the 7950 XRS:

```
A:7950 XRS-20# configure card 1
A:7950 XRS-20>config>card# info
-----
card-type xcm-x20
mda 1
  mda-type cx20-10g-sfp
  no shutdown
exit
```

```

mda 2
  mda-type cx2-100g-cfp
  no shutdown
exit
no shutdown
-----

```

2.17.3.1.1 Configuring FP network pool parameters

FP-level pools are used by ingress network queues. Network policies can be applied (optional) to create and edit QoS pool resources for ingress network queues. Network-queue and slope policies are configured in the **config>qos** context.

The following example shows an FP pool configuration for 7750 SR or 7450 ESS:

```

*A:PE>config>card>fp# info
-----
  ingress
    network
      pool "default"
      resv-cbs 50
      slope-policy "slope1"
    exit
    queue-policy "10"
  exit
exit
-----
*A:PE>config>card>fp#

```

2.17.3.2 Configuring ports

This section provides the CLI syntax and examples to configure port parameters.

2.17.3.2.1 Configuring port pool parameters

The buffer space is portioned out on a per port basis. Each port gets an amount of buffering which is its fair-share based on the port's bandwidth compared to the overall active bandwidth.

This mechanism takes the buffer space available and divides it into a portion for each port based on the port's active bandwidth relative to the amount of active bandwidth for all ports associated with the buffer space. The number of ports sharing the same buffer space depends on the type of MDAs populated on the IOM. An active port is considered to be any port that has an active queue associated. After a queue is created for the port, the system allocates the appropriate amount of buffer space to the port. This process is independently performed for both ingress and egress.

Normally, the amount of active bandwidth is considered as opposed to total potential bandwidth for the port when determining the port's fair share. If a port is channelized and not all bandwidth is allocated, only the bandwidth represented by the configured channels with queues configured is counted toward the bandwidth represented by the port. Also, if a port may operate at variable speeds (as in some Ethernet ports), only the current speed is considered. Based on the above, the number of buffers managed by a port may change because of queue creation and deletion, channel creation and deletion and port speed variance on the local port or other ports sharing the same buffer space.

After the active bandwidth is calculated for the port, the result may be modified through the use of the **ing-percentage-of-rate** and **egr-percent-of-rate** commands. The default value of each is 100% which allows the system to use all of the ports active bandwidth when deciding the relative amount of buffer space to allocate to the port. When the value is explicitly modified, the active bandwidth on the port is changed according to the specified percentage. If a value of 50% is given, the ports active bandwidth is multiplied by 5, if a value of 150% is given, the active bandwidth is multiplied by 1.5. The ports rate percentage parameters may be modified at any time.

Examples:

- To modify (in this example, to double) the size of buffer allocated on ingress for a port:

```
B:SR7-10# configure port 1/2/1 modify-buffer-allocation-rate ing-percentage-of-rate 200
```

- To modify (in this example, to double) the size of buffer allocated on egress for a port:

```
B:SR7-10# configure port 1/2/1 modify-buffer-allocation-rate egr-percentage-of-rate 200
```

The default buffer allocation has the following characteristics:

- Each port manages a buffer according to its active bandwidth (ports with equal active bandwidth get the same buffer size).
- An access port has 2 default pools created: access-ingress and access-egress.
- A network port has 2 default pools created: ingress-FP (common pool for all ingress network ports) and network-egress.
- All queues defined for a port receive buffers from the same buffer pool.

The following example shows port pool configurations:

```
A:ALA-B>config>port# info
-----
    access
      egress
        pool
          slope-policy "slopePolicy1"
        exit
      exit
    exit
  network
    egress
      pool
        slope-policy "slopePolicy2"
      exit
    exit
  exit
no shutdown
-----
```

The following shows a CBS configuration over subscription example:

```
*A:Dut-T>config>port# info
-----
    access
      ingress
        pool
          amber-alarm-threshold 10
          resv-cbs 10 amber-alarm-action step 1 max 30
        exit
      exit
    exit
  exit
no shutdown
-----
```

```

        exit
    exit
    ethernet
        mode access
        encap-type dot1q
    exit
    no shutdown

```

2.17.3.2.2 Changing hybrid-buffer-allocation

The following example shows a hybrid-buffer-allocation value change (from default) for ingress. In this example, the network-egress buffer pool is two times the size of the access-egress.

```

A:SR>config>port>hybrid-buffer-allocation# info
-----
egr-weight access 20 network 40

```

2.17.3.2.3 Changing APS parameters



Note: Nokia recommends grouping working lines and protect lines on separate IOMs.

APS configuration rules:

- A working port must be added first. Then a protection port can be added or removed at any time.
- A protection port must be shutdown before being removed from an APS group.
- A path cannot be configured on a port before the port is added to an APS group.
- A working port cannot be removed from an APS group until the APS port path is removed.
- When ports are added to an APS group, all path-level configurations are available only on the APS port level and configuration on the physical member ports are blocked.
- For a multi-chassis APS group, only one member circuit (either working or protect) can be added. Note that the neighbor IP address of an APS group must be configured before adding a member circuit in it. The configuration of a non-zero neighbor IP address indicates the APS group as multi-chassis. Thus, the member circuit and services must be removed before adding or removing the neighbor IP address (for example, before converting an APS group from multi-chassis to single-chassis or single-chassis to multi-chassis).

The following shows an example configuration for a TDM SC-APS group using SONET/SDH channelization, and a Cpipe service using an APS SAP and a physical TDM SAP:

```

A:ALA-274>config# port (1/1/1)
-----
    sonet-sdh
        speed oc3
    exit
    no-shutdown
-----
A:ALA-274>config>port# aps-1
-----
    aps
        working-circuit tsat-1/1/2
        protect-circuit tsat-20/1/4

```

```

exit
sonet-sdh
    path sts1-1
    exit
    path sts1-3
    exit
    path vt15-1.1.1
    exit
    path vt15-3.7.4
    exit
exit
tdm
    ds1 1.1.1
        framing ds1-unframed
        clock-source node-timed
        channel-group 1
    exit
    exit
    ds1 3.7.4
        framing ds1-unframed
        clock-source node-timed
        channel-group 1
    exit
    exit
exit
no-shutdown
exit
-----
A:ALA-274>cpipe 221 customer 1 vc-type satop-t1 create
    sap aps-100.1.1.1.1 create
    exit
    sap tsat-1/1/1.1.1.1.1 create
    exit
exit

```

The following shows an example of the configuration for the working circuit/node of an MC-APS group:

```

A:ALA-274>config>port (10/1/3)# info
-----
description "APS Group"
aps
    neighbor 192.0.2.2
    working-circuit tsat-10/1/3
exit
no shutdown
-----
A:ALA-274>config>port#

```

The following shows an example of the configuration for the protect circuit/node of an MC-APS group:

```

A:ALA-274>config>port (20/1/4)# info
-----
description "APS Group"
aps
    neighbor 192.0.2.1
    protect-circuit tsat-20/1/4
exit
no shutdown
-----
A:ALA-274>config>port#

```

2.17.3.2.4 Configuring Ethernet port parameters

2.17.3.2.4.1 Ethernet network port

A network port is network facing and participates in the service provider transport or infrastructure network processes.

The following example shows a network port configuration:

```
A:ALA-B>config>port# info
-----
description "Ethernet network port"
 ethernet
  exit
  no shutdown
-----
A:ALA-B>config>port#
```

2.17.3.2.4.2 Ethernet access port

Services are configured on access ports used for customer-facing traffic. If a Service Access Port (SAP) is to be configured on a port, it must be configured as access mode. When a port is configured for access mode, the appropriate encapsulation type can be specified to distinguish the services on the port. After a port has been configured for access mode, multiple services may be configured on the port.

The following example shows an Ethernet access port configuration:

```
A:ALA-A>config>port# info
-----
description "Ethernet access port"
 access
  egress
    pool
      slope-policy "slopePolicy1"
    exit
  exit
exit
network
  egress
    pool
      slope-policy "slopePolicy2"
    exit
  exit
exit
 ethernet
  mode access
  encap-type dot1q
  exit
  no shutdown
-----
A:ALA-A>config>port#
```

2.17.3.2.4.3 Configuring 802.1x authentication port parameters

The following example shows an 802.1x port configuration:

```
A:ALA-A>config>port>ethernet>dot1x# info detail
-----
      port-control auto
      radius-plcy dot1xpolicy
      re-authentication
      re-auth-period 3600
      max-auth-req 2
      transmit-period 30
      quiet-period 60
      supplicant-timeout 30
      server-timeout 30
      no tunneling
-----
```

2.17.3.2.5 Configuring SONET/SDH port parameters

Only the TDM satellite has native SONET/SDH ports.

When an Ethernet port is configured in WAN mode (xgig wan), you can change specific SONET/SDH parameters to reflect the SONET/SDH requirements for this port.

The following CLI output shows an example of a SONET/SDH configuration for a WAN PHY Ethernet port:

```
*A:7950 XRS-20>config>port# info
-----
      shutdown
      ethernet
      xgig wan
      exit
      sonet-sdh
      tx-dus
      suppress-lo-alarm
      threshold ber-sd rate 4
      section-trace increment-z0
      path
      trace-string "hello"
      report-alarm pais
      signal-label 0x20
      exit
      exit
-----
```

2.17.3.2.6 Configuring channelized ports

When configuring channelized ports, the port ID is specified in different ways depending on the level of channelization. Ethernet ports cannot be channelized. [Table 28: Channelization options available on the 7750 SR channelized MDAs](#) lists the channelization options and port syntax available on the 7750 SR channelized MDAs.

Table 28: Channelization options available on the 7750 SR channelized MDAs

Framing	Channelization/mapping option	Channelized MDAs supporting services on the port/channel
Up to 2,048 kb/s (n*DS0 within E1 up to E1)		
SDH	STM4>AUG4>AUG1>VC4>TUG3>TUG2>VC12	TDM satellite configured in mode t1-chstm4-sfp
SDH	STM4>AUG4>AUG1>VC3>TUG2>VC12	TDM satellite configured in mode ts4-chstm1-sfp
Up to 1,544 kb/s (n*DS0 within DS1 up to DS1)		
SDH	STM4>AUG4>AUG1>VC4>TUG3>TUG2>TU11>VC11	TDM satellite configured in mode ts1-chstm4-sfp
SDH	STM4>AUG4>AUG1>VC4>TUG3>TUG2>TU12>VC11	None
SDH	STM4>AUG4>AUG1>VC3>TUG2>VC11	TDM satellite configured in mode ts4-chstm1-sfp
SONET	OC12>STS12>STS1 SPE>VT GROUP>VT1.5 SPE	TDM satellite configured in mode ts1-choc12-sfp or in mode ts4-choc3-sfp



Note: The E1 encapsulation in the ASAP MDA and in the channelized MDAs is compliant to G.704 and G.703. The G.703 feature allows a user to configure an unstructured E1 channel on deep channel MDAs and ASAP MDAs. In G.704, time slot 0 carries timing information by a service provider and therefore, only 31 slots are available to the end user. In G.703, all 32 time slots are available to the end user. Timing is provided by the end user.

A port ID for channels has one of the following syntax as applicable to channelization and mapping options where the port configuration syntax is slot/mda/port ([Table 29: Channelized port syntax examples](#)).

Table 29: Channelized port syntax examples

Port ID for physical port speed			
Channel speed	OC12/STM4	OC3/STM1	DS3/E3
SONET/SDH			
STS3/STM1	port.sts3-{1 to 4}	port.sts3	—
STS1/STM0	port.sts1-{1 to 4}.{1 to 3}	port.sts1-{1 to 3}	—
TUG3	port.tug3-{1 to 4}.{1 to 3}	port.tug3-{1 to 3}	—

Port ID for physical port speed			
Channel speed	OC12/STM4	OC3/STM1	DS3/E3
VT15/VC1.1 ⁶	port.vt15-{1 to 4}.{1 to 3}.{1 to 4}.{1 to 7}	port.vt15-{1 to 3}.{1 to 4}.{1 to 7}	—
VT2/VC12 ⁶	port.vt2-{1 to 4}.{1 to 3}.{1 to 3}.{1 to 7}	port.vt2-{1 to 3}.{1 to 3}.{1 to 7}	—
TDM			
DS1 in VT2	port.{1 to 4}.{1 to 3}.{1 to 3}.{1 to 7}	port.{1 to 3}.{1 to 3}.{1 to 7}	—
DS1 in VT15	port.{1 to 4}.{1 to 3}.{1 to 4}.{1 to 7}	port.{1 to 3}.{1 to 4}.{1 to 7}	—
E1 in VT2	port.{1 to 4}.{1 to 3}.{1 to 3}.{1 to 7}	port.{1 to 3}.{1 to 3}.{1 to 7}	—

2.17.3.2.6.1 Verify the MDA type

To ensure that you have a channel-capable MDA, verify that the MDA-type you are configuring by entering a **show mda slot-id** command.

The MDAs shown in the MDA Provisioned column in the following output are a 12-port channelized DS3 MDA (m12-ds3) on card 1, MDA slot 1, and a 1-port channelized OC12-SFP MDA (m1-choc12-sfp) on card 1, MDA slot 2.

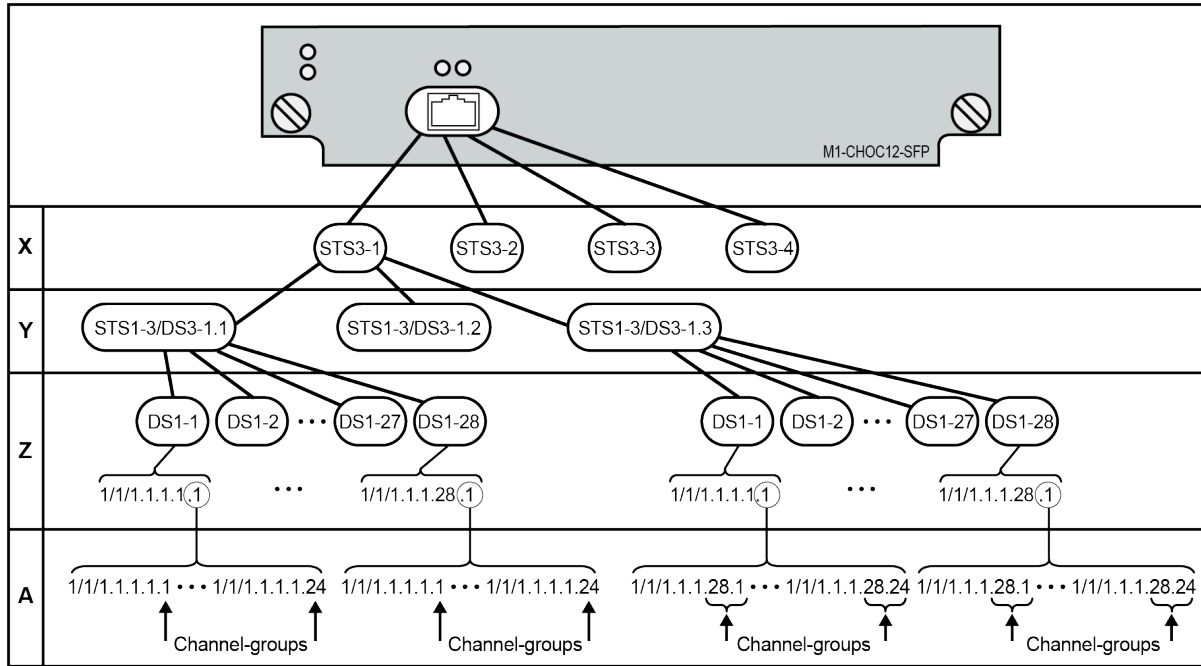
```
A:ALA-A# show mda 1
=====
MDA 1/1
=====
Slot  Mda  Provisioned  Equipped  Admin  Operational
      Mda  Mda-type    Mda-type  State   State
-----
1     1     m12-ds3     m12-ds3   up     provisioned
=====
ALA-A# show mda 2
=====
MDA 1/2
=====
Slot  Mda  Provisioned  Equipped  Admin  Operational
      Mda  Mda-type    Mda-type  State   State
-----
1     2     m1-choc12-sfp  m1-choc12-sfp  up     provisioned
=====
A:ALA-A#
```

2.17.3.2.6.2 Configuring a channelized OC-12-SFP port

[Figure 44: Channelized OC-12 port structure](#) shows the logic of the channelized OC-12 port configuration.

⁶ Supported by TDM satellite.

Figure 44: Channelized OC-12 port structure



hw1516

The following shows an example to configure a channelized port on a 1-port channelized OC-12-SFP MDA:

```
ALA-A>config# port 5/2/1
```

At this level you must choose the tributary. When provisioning nodes on a channelized OC-12 MDA, you must provision the parent STS1-1 SONET path first.

```
A:ALA-A>config>port# sonet-sdh
A:ALA-A>config>port>sonet-sdh# path sts1-1.1
A:ALA-A>config>port>sonet-sdh>path# no shutdown
A:ALA-A>config>port>sonet-sdh>path# exit
```

The following shows the output:

```
A:ALA-A>config>port>sonet-sdh# info
-----
sonet-sdh
  path sts1-1.1
  no shutdown
  exit
  exit
-----
A:ALA-A>config>port>sonet-sdh#
```

To set the channelized mode on a port, the parameter must be in a shutdown state. Clear channel uses out-of-band signaling, not in-band signaling, so the channel's entire bit rate is available. Channelized ports use in-band signaling and must be explicitly enabled.

```
A:ALA-A>config>port# tdm
A:ALA-A>config>port>tdm# ds1 1.1.1
A:ALA-A>config>port>tdm>ds1# shutdown
A:ALA-A>config>port>tdm>ds1# channelized ds1
A:ALA-A>config>port>tdm>ds1# no shutdown
A:ALA-A>config>port>tdm>ds1# exit
```

The following shows an example of the output:

```
A:ALA-A>config>port# info
-----
sonet-sdh
  path sts3-1
    no shutdown
  exit
  path sts1-1.1
    no shutdown
  exit
exit
no shutdown
-----
A:ALA-A>config>port#
```

In the TDM context, configure channel groups parameters. 24 timeslots can be configured per channel group.

```
A:ALA-A>config>port>tdm# ds1 1.1.1
A:ALA-A>config>port>tdm>ds1# no shutdown
A:ALA-A>config>port>tdm>ds1# channel-group 1
A:ALA-A>config>port>tdm>ds1>channel-group# timeslots 1
A:ALA-A>config>port>tdm>ds1>channel-group# no shutdown
A:ALA-A>config>port>tdm>ds1>channel-group# exit
A:ALA-A>config>port>tdm>ds1# no shutdown
A:ALA-A>config>port>tdm>ds1# channel-group 2
A:ALA-A>config>port>tdm>ds1>channel-group# timeslots 2
A:ALA-A>config>port>tdm>ds1>channel-group# no shutdown
A:ALA-A>config>port>tdm>ds1>channel-group# exit
A:ALA-A>config>port>tdm>ds1# exit

A:ALA-A>config>port>tdm# info
-----
sonet-sdh
  path sts3-1
    no shutdown
  exit
  path sts1-1.1
    no shutdown
  exit
exit
tdm
  ds1 ds1-1.1.1
    channel-group 1      (see SAP 5/2/1.1.1.1 below)
    timeslots 1
    no shutdown
  exit
  channel-group 2      (see SAP 5/2/1.1.1.2 below)
```

```

                timeslots 2
                no shutdown
            exit
        no shutdown
    exit
exit
exit
no shutdown
-----
A:ALA-A>config>port>tdm#

```

Services can be applied to the configured channelized ports. The following example shows the CLI usage to configure a customer IES service with interface SAPs on the channelized ports. See the *7450 ESS, 7750 SR, 7950 XRS, and VSR Services Overview Guide* for information about how to configure services.

```

A:ALA-A>config>service# ies 104 customer 1 create
A:ALA-A>config>service>ies$ interface testA create
A:ALA-A>config>service>ies>if$ address 192.168.1.1/24
A:ALA-A>config>service>ies>if# sap 5/2/1.1.1.1.1 create
A:ALA-A>config>service>ies>if>sap$ exit
A:ALA-A>config>service>ies>if# no shutdown
A:ALA-A>config>service>ies>if# exit
A:ALA-A>config>service>ies# interface testB create
A:ALA-A>config>service>ies>if$ address 192.168.2.1/24
A:ALA-A>config>service>ies>if# sap 5/2/1.1.1.1.2 create
A:ALA-A>config>service>ies>if>sap$ exit
A:ALA-A>config>service>ies>if# no shutdown
A:ALA-A>config>service>ies>if# exit
A:ALA-A>config>service>ies# no shutdown

```

The following output shows the channelized ports 5/2/1.1.1.1 and 5/2/1.1.1.2) applied to SAPs on the IES service configuration:

```

A:ALA-A>config>service>ies# info
-----
        interface "testA" create
            address 192.168.1.1/24
            sap 5/2/1.1.1.1.1 create
            exit
        exit
        interface "testB" create
            address 192.168.2.1/24
            sap 5/2/1.1.1.1.2 create
            exit
        exit
        no shutdown
-----
A:ALA-A>config>service>ies#

```

2.17.3.2.7 Configuring channelized STM1/OC3 parameters

The following example shows basic syntax to configure channelized STM1/OC3 parameters:

```

config# port port-id
- sonet-sdh
  - framing {sonet | sdh}
  - group sonet-sdh-index payload {tu3 | vt2 | vt15}
  - path [sonet-sdh-index]
    - payload {sts3 | tug3 | ds3 | e3}
    - trace-string [trace-string]

```

– no shutdown

```

config# port 5/2/1
  – config>port# sonet-sdh
  – config>port>sonet-sdh# framing sdh
  – config>port>sonet-sdh# path sts3
  – config>port>sonet-sdh>path# trace-string "H0-path"
  – config>port>sonet-sdh>path# exit
  – config>port>sonet-sdh# group tug3-1 payload vt2
  – config>port>sonet-sdh# group tug3-3 payload vt2
  – config>port>sonet-sdh# path vt2-1.1.1
  – config>port>sonet-sdh>path# trace-string "L0-path 3.7.3"
  – config>port>sonet-sdh>path# no shutdown
  – config>port>sonet-sdh>path# exit
  – config>port>sonet-sdh# exit
config>port# tdm
config>port>tdm# e1 1.1.1
  config>port>tdm>e1# channel-group 1
  config>port>tdm>e1>channel-group# timeslots 2-32
  config>port>tdm>e1>channel-group# no shutdown
  – config>port>tdm>e1>channel-group# exit
  – config>port>tdm# e1 3.7.3
  config>port>tdm>e1# channel-group 2
  config>port>tdm>e1>channel-group# timeslots 2-32
  config>port>tdm>e1>channel-group# no shutdown
  config>port>tdm>e1>channel-group# exit

```

The following shows the configuration output:

```

A:ALA-49>config>port# info
-----
sonet-sdh
  framing sdh
  path sts3
    trace-string "H0-path"
    no shutdown
  exit
  group tug3-1 payload vt2
  group tug3-3 payload vt2
  path vt2-1.1.1
    trace-string "L0-path 3.7.3"
    no shutdown
  exit
  path vt2-3.7.3
    no shutdown
  exit
exit
tdm
  e1 1.1.1
    channel-group 1
      timeslots 2-32
      no shutdown
    exit
    no shutdown
  exit
  e1 3.7.3
    channel-group 2
      timeslots 2-32
      no shutdown
    exit
    no shutdown
  exit
exit

```

```
no shutdown
-----
```

2.17.3.2.7.1 Example Cpipe port configurations

Before a Cpipe service can be provisioned, the following entities must be configured:

2.17.3.2.7.1.1 Configuring a DS1 port

The following shows an example of a DS1 port configured for CES:

```
A:sim216# show port 1/5/1.1.3.1
=====
TDM DS1 Interface
=====
Description      : DS1
Interface        : 1/5/1.1.3.1
Type             : ds1
Admin Status     : up
Physical Link    : yes
Signal Mode      : none
Last State Change : 10/31/2006 14:23:12
Loopback         : none
Remote Loop respond: false
Load-balance-algo : default
BERT Duration    : N/A
BERT Synched     : 00h00m00s
BERT Errors      : 0
BERT Total Bits  : 0
Cfg Alarm        : ais los
Alarm Status     :
Framing          : esf
Oper Status      : up
Clock Source     : loop-timed
Channel IfIndex  : 580943939
Invert Data      : false
In Remote Loop   : false
Egr. Sched. Pol  : n/a
BERT Pattern     : none
Err Insertion Rate : 0
BERT Status      : idle
=====
```

2.17.3.2.7.1.2 Configuring a channel group

The following shows an example of a DS1 channel group configured for CES:

```
A:sim216# show port 1/5/1.1.3.1
=====
TDM DS0 Chan Group
=====
Description      : DS0GRP
Interface        : 1/5/1.1.3.1
TimeSlots       : 1-12
Speed           : 64
Admin Status     : up
Physical Link    : Yes
Idle Cycle Flags : flags
Egr. Sched. Pol  : n/a
Last State Change : 10/31/2006 14:23:12
Configured mode  : access
Admin MTU        : 4112
Bundle Number    : none
Load-balance-algo : default
CRC              : 16
Oper Status      : up
Chan-Grp IfIndex : 580943940
Encap Type       : cem
Oper MTU         : 4112
=====
A:sim216#
```

2.17.3.2.8 Configuring OTU port parameters

The following example shows an OTU port configuration:

```
*A:ALA-A>config>port>otu# info detail
-----
  otu2-lan-data-rate 11.049
  sf-sd-method fec
  sf-threshold 5
  sd-threshold 7
  fec enhanced
  no report-alarm otu-ais otu-ber-sd otu-tim otu-iae otu-biae fec-sd
  no report-alarm fec-fail fec-uncorr odu-ais odu-oci odu-lck odu-bdi
  no report-alarm odu-tim opu-plm
  report-alarm loc los lof lom otu-ber-sf otu-bdi fec-sf
  sm-tti
    tx auto-generated
    expected auto-generated
    no mismatch-reaction
  exit
  pm-tti
    tx auto-generated
    expected auto-generated
    no mismatch-reaction
  exit
  psi-payload
    tx auto
    expected auto
    no mismatch-reaction
  exit
-----
```

The following example shows the **show port *portId* otu detail** for the default OTU configuration as shown above:

```
*A:ALA-A# show port 3/2/1 otu detail

=====
OTU Interface
=====
OTU Status      : Enabled          FEC Mode       : enhanced
Async Mapping   : Disabled          Data Rate      : 11.049 Gb/s

Cfg Alarms      : loc los lof lom otu-ber-sf otu-bdi fec-sf
Alarm Status    :
SF/SD Method    : FEC                SF Threshold   : 1E-5
                                     SD Threshold   : 1E-7

SM-TTI Tx (auto) : ALA-A:3/2/1/C44
SM-TTI Ex (bytes) : (Not Specified)
SM-TTI Rx       : ALA-A:5/2/1/C34
OTU-TIM reaction : none

PM-TTI Tx (auto) : ALA-A:3/2/1/C44
PM-TTI Ex (bytes) : (Not Specified)
PM-TTI Rx       : ALA-A:5/2/1/C34
ODU-TIM reaction : none

PSI-PT Tx (auto) : 0x03 (syncCbr)
PSI-PT Ex (auto) : 0x03 (syncCbr)
PSI-PT Rx       : 0x03 (syncCbr)
```



```

OPU-PLM reaction   : none
=====
OTU Statistics
=====
Elapsed Seconds                10
-----
Near End Statistics                Count
-----
FEC Corrected 0s                0
FEC Corrected 1s                0
FEC Unrectable Sub-rows         0
FEC ES                           0
FEC SES                           0
FEC UAS                           0
Pre-FEC BER                      0.000E+00
Post-FEC BER                     0.000E+00
-----
SM BIP8                          0
SM ES                             0
SM SES                             0
SM UAS                             0
SM-BIP8-BER                      0.000E+00
-----
PM BIP8                          0
PM ES                             0
PM SES                             0
PM UAS                             0
PM-BIP8-BER                      0.000E+00
-----
NPJ                               0
PPJ                               0
-----
Far End Statistics                Count
-----
SM BEI                           0
PM BEI                           0
=====

```

The window over which the Bit Error Rate (BER) determined is based on the configured threshold level. The higher the error rate the shorter the window and as the error rate decreases the window increases. [Table 30: Configured BER thresholds and window lengths](#) lists the configured BER thresholds and corresponding window lengths.

Table 30: Configured BER thresholds and window lengths

Configured BER threshold	Window length
10 ⁻³	8ms
10 ⁻⁴	8ms
10 ⁻⁵	8ms
10 ⁻⁶	13ms
10 ⁻⁷	100ms
10 ⁻⁸	333ms

Configured BER threshold	Window length
10 ⁻⁹	1.66s

2.17.3.2.9 Configuring LAG parameters

LAG configurations should include at least two ports. Other considerations include:

- A maximum of 64 ports (depending on the lag-id) can be included in a LAG. All ports in the LAG must share the port characteristics inherited from the primary port.
- Autonegotiation must be disabled or set limited mode for ports that are part of a LAG to guarantee a specific port speed.
- Ports in a LAG must be configured as full duplex.

The following example shows LAG configuration output:

```
A:ALA-A>config>lag# info detail
-----
description "LAG2"
mac 04:68:ff:00:00:01
port 1/1/1
port 1/3/1
port 1/5/1
port 1/7/1
port 1/9/1
dynamic-cost
port-threshold 4 action down
-----
A:ALA-A>config>lag#
```

2.17.3.2.9.1 Configuring BFD on LAG links

About this task

BFD can be configured under the LAG context to create and establish the micro-BFD session per link after the LAG and associated links have been configured. An IP interface must be associated with the LAG or a VLAN within the LAG, if dot1q encapsulation is used, before the micro-BFD sessions can be established.

Complete the following steps to enable and configure BFD over the individual LAG links:

When configuring the local and remote IP address for the BFD over LAG link sessions, the *local-ip* parameter should always match an IP address associated with the IP interface to which this LAG is bound. In addition, the *remote-ip* parameter should match an IP address on the remote system and should also be in the same subnet as the *local-ip* address. If the LAG bundle is re-associated with a different IP interface, the *local-ip* and *remote-ip* parameters should be modified to match the new IP subnet. The *local-ip* and *remote-ip* values do not have to match a configured interface in the case of tagged LAG/ports.

The optional parameters that can be configured for the BFD over LAG links include:

- Transmit Interval
- Receive Interval
- Multiplier

- Max-Wait-for-Up-Time, this parameter controls how long a link remains active if BFD is enabled after the LAG and associated links are active and in a forwarding state.
- Max-Time-Admin-Down, this parameter controls how long the system waits before bringing the associated link out of service if an admin down message is received from the far-end.

Procedure

- Step 1.** Enable BFD within the LAG context, which also enters the CLI into the BFD context.
- Step 2.** Configure the address family which is to be used for the micro BFD sessions. Only one address family can be configured per LAG.
- Step 3.** Configure the local-IP address to be used for the BFD sessions.
- Step 4.** Configure the remote-IP address to be used for the BFD sessions.

Example

The following is an example configuration:

```
*A:Dut-C>config>lag# info
-----
      bfd
        family ipv4
          local-ip-address 10.120.1.2
          receive-interval 1000
          remote-ip-address 10.120.1.1
          transmit-interval 1000
          no shutdown
        exit
      exit
    no shutdown
```

2.17.3.2.10 Configuring G.8031 protected Ethernet tunnels

Ethernet tunnel configuration can include at most two paths. Other considerations include:

- A path contains one member port and one control-tag (backbone VLAN ID/BVID).
- If the user wants to replace an existing member port or a control-tag, the whole path needs to be shutdown first. The alternate path is activated as a result keeping the traffic interruption to a minimum. Then the whole path must be deleted and re-created. To replace an existing member port or control tag, the whole path needs to be shutdown first. The alternate path is activated as a result keeping traffic interruption to a minimum. Then the whole path must be deleted, the alternate path precedence modified to primary before re-creating the new path.
- The Ethernet tunnel inherits the configuration from the first member port. The following port-level configuration needs to be the same between member ports of an Ethernet tunnel:
 - **config>port>ethernet>access>{ingress|egress}>queue-group**
 - **config>port>ethernet>egress-scheduler-policy**
 - **config>port>access>egress>pool**
 - **config>port>ethernet>dot1q-etype**
 - **config>port>ethernet>qinq-etype**
 - **config>port>ethernet>pbb-etype**

– **config>port>ethernet> mtu**

- The user can update these port parameters only if the port is the sole member of an Ethernet tunnel. This means that in the example below, the user needs to remove port 1/1/4 and port 1/1/5 before being allowed to modify 1/1/1 for the above parameters.

```
eth-tunnel 1
  - path 1
    - member 1/1/1
  - path 2
    - member 1/1/4
- eth-tunnel 2
  - path 1
    - member 1/1/1
  - path 2
    - member 1/1/5
```

The following example shows eth-tunnel configuration output:

```
- port 1/1/1
  - ethernet
    - encap-type dot1q
- port 2/2/2
  - ethernet
    - encap-type dot1q
-
- config eth-tunnel 1
  - path 1
    - member 1/1/1
    - control-tag 100
    - precedence primary
    - eth-cfm
      - mep 51 domain 1 association 1
      - ccm-enable
      - low-priority-defect allDef
      - mac-address 00:AE:AE:AE:AE:AE
      - control-mep
      - no shutdown
  - no shutdown
  - path 2
    - member 2/2/2
    - control-tag 200
- eth-cfm
  - mep
    - mep 52 domain 1 association 2 direction down
    - ccm-enable
    - low-priority-defect allDef
    - mac-address 00:BE:BE:BE:BE:BE
    - control-mep
    - no shutdown
- no shutdown
```

2.17.3.2.11 Configuring connectors and connector ports

Some assemblies have support for QSFP28 or QSFP-DD transceiver modules. These modules have different variants, some of which provide multiple physical ports out of a single module (breakout modules). There is a QSFP28 breakout module that supports ten physical 10 Gb Ethernet ports. On assemblies that support these breakout variants, the front panel cages are modeled as connectors instead of as direct

ports. The connector must be configured for the type of breakout module that is to be inserted and then the appropriate ports are created and can be configured. The options for breakout on specific connectors depend on both the card type and level (or XMA type and level). See the applicable installation guides for details.

The connector reference is in the format *slot/mda/connector* (for example, 1/1/c3) and the ports owned by the connector use the format *slot/mda/connector/port*. For example, in a 7750 SR-1 with the 6-port QSFP28 mda-e-xp installed in the first MDA slot, initially there are no ports available, only six connectors:

```
A:bkvm18# show mda
=====
MDA Summary
=====
Slot  Mda  Provisioned Type           Admin  Operational
      Mda  Equipped Type (if different) State   State
-----
1     1     me6-100gb-qsfp28          up     up
*A:bkvm18# show port
=====
Ports on Slot 1
=====
Port   Admin Link Port  Cfg  Oper  LAG/  Port  Port  Port  C/QS/S/XFP/
Id     State  State State MTU  MTU  Bndl  Mode  Encp  Type  MDIMDX
-----
1/1/c1  Up    No   Down  -    -    -    unkn unkn conn
1/1/c2  Up    No   Down  -    -    -    unkn unkn conn
1/1/c3  Up    No   Down  -    -    -    unkn unkn conn
1/1/c4  Up    No   Down  -    -    -    unkn unkn conn
1/1/c5  Up    No   Down  -    -    -    unkn unkn conn
1/1/c6  Up    No   Down  -    -    -    unkn unkn conn
=====
```

After configuring a module with four 10 Gb breakout ports in connector position 1/1/c1 and a module with one 100 Gb breakout port in connector position 1/1/c2, the physical ports are created:

```
A:bkvm18# configure port 1/1/c1 connector breakout c4-10g
A:bkvm18# configure port 1/1/c2 connector breakout c1-100g
A:bkvm18>show port
=====
Ports on Slot 1
=====
Port   Admin Link Port  Cfg  Oper  LAG/  Port  Port  Port  C/QS/S/XFP/
Id     State  State State MTU  MTU  Bndl  Mode  Encp  Type  MDIMDX
-----
1/1/c1  Up    No   Link Up  -    -    -    unkn unkn conn  40GBASE-SR4
1/1/c1/1 Down  No   Down  9212 9212 -    netw null xgige
1/1/c1/2 Down  No   Down  9212 9212 -    netw null xgige
1/1/c1/3 Down  No   Down  9212 9212 -    netw null xgige
1/1/c1/4 Down  No   Down  9212 9212 -    netw null xgige
1/1/c2  Down  No   Link Up  -    -    -    unkn unkn conn  100GBASE-LR4
1/1/c2/1 Down  No   Down  1578 1578 -    netw null cgige
1/1/c3  Up    No   Down  -    -    -    unkn unkn conn
1/1/c4  Up    No   Down  -    -    -    unkn unkn conn
1/1/c5  Up    No   Down  -    -    -    unkn unkn conn
1/1/c6  Up    No   Down  -    -    -    unkn unkn conn
=====
```

These physical ports can now be used as Ethernet port references in other commands.

The transceiver information is shown under the connector while Ethernet-related items are shown under the connector ports.

```
*A:Dut-A# show port 1/1/c1
=====
QSFP28 Connector
=====
Description      : QSFP28 Connector
Interface        : 1/1/c1
Admin State     : up
Oper State      : up
IfIndex         : 104939520
Last State Change : 10/31/2017 13:23:22
Last Cleared Time : N/A
Breakout        : c4-10g
DDM Events      : Enabled

Transceiver Data
Transceiver Status : operational
Transceiver Type   : QSFP28
Model Number       : 3HE10551AARA01 NOK IPU3BFVEAA
TX Laser Wavelength: 850 nm
Number of Lanes    : 4
Connector Code     : MP0 1x12
Manufacture date   : 2017/03/12
Serial Number      : INHAG8480890
Part Number        : TR-FC85S-NN0
Optical Compliance : 100GBASE-SR4 or 25GBASE-SR
Link Length support: 70m for OM3; 100m for OM4

=====
Transceiver Digital Diagnostic Monitoring (DDM)
=====
Value High Alarm High Warn Low Warn Low Alarm
-----
Temperature (C)      +29.8 +80.0 +75.0 -5.0 -10.0
Supply Voltage (V)   3.29  3.63  3.46  3.14  2.97
=====

Transceiver Lane Digital Diagnostic Monitoring (DDM)
=====
High Alarm High Warn Low Warn Low Alarm
-----
Lane Tx Bias Current (mA)      12.0  10.0  4.5  3.0
Lane Tx Output Power (dBm)     5.40  2.40  -8.40 -11.40
Lane Rx Optical Pwr (avg dBm)  5.40  2.40 -10.30 -13.30
-----
Lane ID Temp(C)/Alm Tx Bias(mA)/Alm Tx Pwr(dBm)/Alm Rx Pwr(dBm)/Alm
-----
1 - 7.1 -0.08 -4.56
2 - 0.0/L-WA -40.00/L-WA -40.00/L-WA
3 - 0.0/L-WA -40.00/L-WA -40.00/L-WA
4 - 0.0/L-WA -40.00/L-WA -40.00/L-WA
=====
*A:Dut-A#
*A:Dut-A#
*A:Dut-A# show port 1/1/c1/1
=====
```

```

Ethernet Interface
=====
Description      : 10-Gig Ethernet
Interface        : 1/1/c1/1
Link-level      : Ethernet
Admin State     : up
Oper State      : up
Physical Link   : Yes
Single Fiber Mode : No
IfIndex         : 104939521
Last State Change : 10/31/2017 13:23:23
Last Cleared Time : N/A
Phys State Chng Cnt: 1

Oper Speed      : 10 Gbps
Config Speed   : N/A
Oper Duplex    : full
Config Duplex  : N/A
MTU            : 9212
Min Frame Length : 64 Bytes
Hold time up   : 0 seconds
Hold time down : 0 seconds

RS-FEC Mode    : None

Configured Mode : network
Dot1Q Ethertype : 0x8100
PBB Ethertype   : 0x88e7
Ing. Pool % Rate : 100
Ing. Pool Policy : n/a
Egr. Pool Policy : n/a
Net. Egr. Queue Pol: default
Egr. Sched. Pol : n/a
HS Scheduler Plcy : default
HS Port Pool Plcy : default
Monitor Port Sched : Disabled
Monitor Agg Q Stats: Disabled
Auto-negotiate   : N/A
Oper Phy-tx-clock : not-applicable
Accounting Policy : None
Acct Plcy Eth Phys : None
Egress Rate      : Default
Load-balance-algo : Default
Access Bandwidth : Not-Applicable
Access Available BW: 0
Access Booked BW : 0
Sflow           : Disabled

Encap Type      : null
QinQ Ethertype  : 0x8100
Egr. Pool % Rate : 100

MDI/MDX        : N/A
Collect-stats  : Disabled
Collect Eth Phys : Disabled
Ingress Rate   : Default
LACP Tunnel    : Disabled
Booking Factor : 100

Suppress Threshold : 2000
Max Penalties     : 16000
Half Life        : 5 seconds

Reuse Threshold   : 1000
Max Suppress Time: 20 seconds

Down-when-looped : Disabled
Loop Detected    : False
Use Broadcast Addr : False

Keep-alive       : 10
Retry            : 120

Sync. Status Msg. : Disabled
Tx DUS/DNU       : Disabled
SSM Code Type    : sdh

Rx Quality Level : N/A
Tx Quality Level : N/A

Down On Int. Error : Disabled
DOIE Tx Disable   : N/A

CRC Mon SD Thresh : Disabled
CRC Mon SF Thresh : Disabled
CRC Mon Window    : 10 seconds

EFM OAM          : Disabled
Ignr EFM OAM State : False
EFM OAM Link Mon : Disabled

Configured Address : 00:03:fa:2b:ef:1a
Hardware Address   : 00:03:fa:2b:ef:1a
Cfg Alarm          : remote local
=====
Traffic Statistics
=====

```

```

=====
                                     Input          Output
-----
Octets                               0            0
Packets                              0            0
Errors                               0            0
Utilization (300 seconds)           0.00%        0.00%
=====

Port Statistics
=====
                                     Input          Output
-----
Unicast Packets                      0            0
Multicast Packets                    0            0
Broadcast Packets                     0            0
Discards                             0            0
Unknown Proto Discards                0
=====

Ethernet-like Medium Statistics
=====
Alignment Errors :                   0  Sngl Collisions :                   0
FCS Errors       :                   0  Mult Collisions :                   0
SQE Test Errors :                   0  Late Collisions :                   0
CSE              :                   0  Excess Collisns :                   0
Too long Frames :                   0  Int MAC Tx Errs :                   0
Symbol Errors   :                   0  Int MAC Rx Errs :                   0
In Pause Frames :                   0  Out Pause Frames :                   0
=====

```

2.18 Service management tasks

This section discusses basic procedures to complete service management tasks.

2.18.1 Modifying or deleting an MDA or XMA

To change an MDA or XMA type already provisioned for a specific slot or card, first you must shut down the slot/MDA/port configuration and then delete the MDA or the XMA from the configuration.

To modify or delete XMA, use the MDA command structure.

Use the following CLI syntax to modify an MDA on the 7450 ESS and 7750 SR platforms (or an XMA on the 7950 XRS platforms):

```
config> port port-id
- shutdown
```

```
config> card slot-number
- shutdown
- [no] mda mda-number
- [no] mda-type mda-type
- shutdown
```


2.18.2 Modifying a card type

To modify the card type already provisioned for a specific slot, you must shutdown existing port configurations and shutdown and remove all MDA or XMA configurations.

You must reset the IOM after changing the MDA type from MS-ISA to any other MDA type.

Use the following CLI syntax to modify a card type already provisioned for a specific slot:

```
config> port port-id
  - [no] shutdown
```

```
config> card slot-number
  - mda mda-number
    - [no] mda-type mda-type
    - [no] shutdown
```

2.18.3 Deleting a card

To delete a card type provisioned for a specific slot, you must shutdown existing port configurations and shutdown and remove all MDA or XMA configurations.

Use the following CLI syntax to delete a card provisioned for a specific slot:

```
config> port port-id
  - shutdown
```

```
config> card slot-number
  - card-type card-type
  - mda mda-number
    - no mda-type mda-type
    - no shutdown
```

2.18.4 Deleting port parameters

Use the following CLI syntax to delete a port provisioned for a specific card:

```
config>port port-id
  - shutdown
  - no port port-id
```

2.18.5 Soft IOM reset

This section provides basic procedures for soft IOM reset service management tasks.

2.18.5.1 Soft reset

Soft reset is an advanced high availability feature that greatly reduces the impact of IOM/IMM resets either during a software upgrade or during other maintenance or debug operations. The combination of In Service Software Upgrade (ISSU) and Soft reset maximizes service availability in an operational network.

A soft reset re-initializes the control plane while the data plane continues operation with only very minimal impact to data forwarding. During the soft reset some processes that rely on the IOM control plane do not run for a duration that is similar to the duration of an IOM Hard reset. These processes include the updating of the IP forwarding table on the IOM (IP FIB downloads from the CPM), Layer 2 learning of new MAC addresses on the IOM, updating of the MAC forwarding table (for MAC addresses learned from other IOMs), ARP, Ethernet OAM 802.3ah, LLDP and handling for specific ICMP functions such as Can't Fragment, Redirect, Host Unreachable, Network Unreachable and TTL Expired. Note that protocols and processes on the CPM continue to operate during a Soft Reset (BGP continues to learn new routes from peers, and the new routes are downloaded to the IOM after the Soft Reset has completed).

The combination of the very small data plane impact and special soft reset enhancements for protocols ensures that most protocols do not go down and no visible impacts to most protocols are detected externally to the SR/ESS platforms. BFD timers are temporarily increased for the duration of a soft reset to keep BFD sessions up. Protocols such as BGP, OSPF, IS-IS, PIM, and so on with default timers remain up. A protocol using aggressive timers may go down momentarily during a soft reset.

Although the majority of protocols stay up during a Soft Reset, there are some limitations for a few protocols. See *Known Limitations* in the *Release Notes* for the relevant release for details.

Configuration changes are not allowed while any card is in the process of a soft reset.

The soft IOM reset procedure is applicable during the ISSU process and for a manual soft reset procedure.

To manually perform a soft IOM reset, enter the **clear card slot-number soft** command.

Soft Reset is supported on Ethernet IMMs and on IOMs that have Ethernet MDAs provisioned. The operator can optionally force a Soft Reset on an IOM that contains at least one MDA that supports Soft Reset but also has an MDA that does not support Soft Reset or is operationally down. To force Soft Reset in this case the **hard-reset-unsupported-mdas** is used and the supported MDAs and the card itself are soft reset while the MDAs that do not support soft reset (or are operationally down) are hard reset.

The **show card** and **show mda** commands indicate that a soft IOM reset is occurring during the soft reset process.

2.18.5.2 Deferred MDA reset

As part of an ISSU, soft reset is supported even if the (old) firmware version on the MDAs is not the same as the (new) firmware version in the software load to which the user is upgrading. The soft reset is allowed to proceed by leaving the previous version of the firmware running while upgrading the rest of the MDA/IOM/IMM. The user can then issue a hard reset of the MDA/IMM at some time in the future to upgrade the firmware.

The soft reset is only allowed to proceed if the older firmware is compatible with the new IOM/IMM software load. Otherwise the soft reset is blocked and a hard reset must be used instead.

After a soft reset has been completed, a log event is raised to warn the user that the MDA (or IMM) is running older firmware and that they can perform a hard reset of the MDA (or IMM) at some point if required.

If the MDA/IMM is not hard reset by the user, and then a software upgrade is performed, and the older firmware is no longer compatible with the newest load being upgraded to, then the soft reset is blocked (or an automatic hard reset occurs for ISSU).

The user can see whether they are running with older MDA/IMM firmware at any time by using the **show mda detail** command.

2.19 DWDM provisioning

This section provides information to provision the DWDM coherent optic, frequency, and coherent parameters.



Note: Different procedures are required if the DWDM port is managed using the OpenConfig component YANG model.

2.19.1 Provisioning the DWDM coherent optic

About this task

This procedure describes how to provision the DWDM coherent optic. To provision the DWDM coherent optic, the user must provision the connector breakout type and enable the transceiver Digital Coherent Optics (DCO) on the cage where the device is present. Use the following command to show connector information.

```
show port 1/1/c1
```

Output example

The following example shows that the starting point is a new default connector 1/1/c1 that is not provisioned, is administratively shutdown, and has a DWDM coherent optic transceiver installed.

```
=====
QSFP-DD Connector
=====
Description      : QSFP-DD Connector
Interface        : 1/1/c1
FP Number        : 1                      MAC Chip Number : 1
Licensed         : Yes
Admin State      : down
Oper State       : down
IfIndex          : 1610899520
Last State Change : 03/01/2022 18:10:14
Last Cleared Time  : 03/01/2022 18:20:32    DDM Events       : Enabled
Breakout         : no breakout
RS-FEC Config Mode : None

Transceiver Data

Transceiver Status : operational
Transceiver Type   : QSFP-DD                DCO              : Disabled
Model Number       : 3HE16565AARA01 NOK INUIAPJHAA
TX Laser Wavelength: 1547 nm                Diag Capable     : yes
Number of Lanes    : 1
Connector Code     : LC                      Vendor OUI       : 20:20:20
Manufacture date   : 2021/06/03             Media            : Ethernet
=====
```

```
Serial Number      : 1111111111
Part Number       : DP04QSD- E30-090
Optical Compliance : 400G-ZR-Amp 400G-ZR-Unamp 400G-ZR-Amp Custom (197) Custo*
Link Length support: Unknown
```

Procedure

Step 1. Use the following command to configure the connector breakout.

```
configure port connector breakout
```

Example

MD-CLI

```
*[ex:/configure port 1/1/c1 connector]
A:admin@node-2# breakout c1-400g
```

Example

classic CLI

```
*A:node-2# configure port 1/1/c1 connector breakout c1-400g
```

Step 2. Use the following command to configure the DCO.

```
configure port transceiver digital-coherent-optics
```

Example

MD-CLI

```
*[ex:/configure port 1/1/c1 transceiver]
A:admin@node-2# digital-coherent-optics true
```

Example

classic CLI

```
*A:node-2# configure port 1/1/c1 transceiver digital-coherent-optics
```

2.19.2 Provisioning DWDM frequency

Prerequisites

Before provisioning DWDM frequency, the user must configure the connector breakout and apply the DCO setting to the transceiver (as described in [Provisioning the DWDM coherent optic](#)). After the DCO is enabled, and the physical optical transceiver is present in the front panel cage, the optic transceiver data shows the optic DWDM details. The provisionable frequency range is also displayed here with the **show port 1/1/c1** command (see the "Model Number", "Fine Tune Range", "Fine Tune Resolu**", and "Supported Grids" fields below). To initially provision the frequency or change what was previously configured, the user must shut down or disable the 1/1/c1 connector before provisioning.

```
Transceiver Type   : QSFP-DD                DCO                : Enabled
Model Number      : 3HE16565AARA01  NOK  INUIAPJHAA
TX Laser Wavelength: 1547 nm
```

```
Laser Tunability      : flex-tunable
Config Freq (MHz)    : 0
Oper Freq (MHz)      : not-operational
Fine Tune Range      : 6000 MHz
Supported Grids      : 100GHz 75GHz 50GHz 25GHz 12.5GHz 6.25GHz
Diag Capable         : yes
Number of Lanes      : 1
Connector Code       : LC
Manufacture date     : 2021/06/03
Serial Number        : 1111111111
Part Number          : DP04QSDD-E30-090
Optical Compliance   : 400G-ZR-Amp 400G-ZR-Unamp
Link Length support  : Unknown
Vendor OUI           : 20:20:20
Media                : Ethernet
```

=====
Transceiver Digital Diagnostic Monitoring (DDM)
=====

	Value	High Alarm	High Warn	Low Warn	Low Alarm
Temperature (C)	+36.0	+80.0	+75.0	+15.0	-5.0
Supply Voltage (V)	3.30	3.46	3.43	3.17	3.13

=====
Transceiver Lane Digital Diagnostic Monitoring (DDM)
=====

	High Alarm	High Warn	Low Warn	Low Alarm
Lane Tx Output Power (dBm)	0.00	-2.00	-16.00	-18.01
Lane Rx Optical Pwr (avg dBm)	2.00	0.00	-23.01	-28.24

Lane ID	Temp(C)/Alm	Tx Bias(mA)/Alm	Tx Pwr(dBm)/Alm	Rx Pwr(dBm)/Alm
1	-	-	-40.00	-40.00

=====
Coherent Optical Module
=====

```
Cfg Tx Target Power: 1.00 dBm
Cfg Rx LOS Thresh : -23.00 dBm
Present Rx Channel : N/A
Cfg Rx Channel : 0 (auto)

Disp Control Mode : automatic
Cfg Dispersion : 0 ps/nm
CPR Window Size : 32 symbols
Compatibility : longHaul
Cfg Tx Power Min : -22.90 dBm
Sweep Start Disp : -25500 ps/nm
Sweep End Disp : 2000 ps/nm
Rx LOS Reaction : squelch
Cfg Tx Power Max : 4.00 dBm

Cfg Alarms : modflt mod netrx nettx hosttx
Alarm Status :
Defect Points :

Rx Q Margin : 0.0 dB
SNR X Polar : 0.0 dB
SNR Y Polar : 0.0 dB
Chromatic Disp : 0 ps/nm
Diff Group Delay : 0 ps
Pre-FEC BER : 0.000E+00

Module State : lowPower
Tx Turn-Up States :
Rx Turn-Up States :
```

About this task

This procedure describes how to provision the DWDM frequency.

Procedure

Step 1. Configure the frequency using the following context.

By default, no frequency is configured. A frequency must be provisioned. The following example displays a frequency configuration of 194400000:

Example

configure port 1/1/c1 dwdm frequency 194400000 (classic CLI)

```
*A:Dut-A# configure port 1/1/c1 dwdm frequency
- frequency <MHz>
- no frequency

<MHz>                : [191100000..196150000]

*A:Dut-A# configure port 1/1/c1 dwdm frequency 194400000
```

Example

configure port 1/1/c1 dwdm frequency 194400000 (MD-CLI)

```
*[pr:/configure port 1/1/c1 dwdm]
A:admin@cses-V27# frequency ?

frequency <number>
<number> - <191100000..196150000> - megahertz

Center frequency for tunable DWDM optical interface

*[pr:/configure port 1/1/c1 dwdm]
A:admin@cses-V27# frequency 194400000
```

Step 2. Enable frequency by configuring either of the following:

- In the classic CLI, use the **configure port 1/1/c1 no shutdown** command to enable the connector.
- In the MD-CLI, use the **configure port 1/1/c1 admin-state enable** command to enable the connector.

Expected outcome

In the following example, the “Config Freq” and “Oper Freq” fields in the **show port 1/1/c1** output reflect the frequency setting, the “Coherent Optical Module” values are present, and the “Module State” displays the “txOff” value until the connector ports are administratively set to **no shutdown** or set to **enable**. For any DWDM frequency changes, after port 1/1/c1 is administratively set to **no shutdown** or set to **enable** again, the coherent module initializes and transitions through various states as visible in the “Module State” field.

```
Transceiver Type   : QSFP-DD                               DCO           : Enabled
Model Number      : 3HE16565AARA01  NOK  INUIAPJHAA
TX Laser Wavelength: 1542.142 nm
Laser Tunability  : flex-tunable
Config Freq (MHz) : 194400000                               Min Freq (MHz) : 191300000
Oper Freq (MHz)   : 194400000                               Max Freq (MHz) : 196100000
Fine Tune Range   : 6000 MHz                               Fine Tune Resolu*: 1 MHz
Supported Grids   : 100GHz 75GHz 50GHz 25GHz 12.5GHz 6.25GHz
Diag Capable      : yes
Number of Lanes   : 1
Connector Code    : LC                                   Vendor OUI     : 7c:b2:5c
Manufacture date  : 2020/01/11                           Media          : Ethernet
```

```

Serial Number      : 11111111
Part Number       : DP04QSDD-E30-090
Optical Compliance : 400G-ZR-Amp 400G-ZR-Unamp 400G-ZR-Amp Custom (197) Custo*
Link Length support: Unknown

=====
Transceiver Digital Diagnostic Monitoring (DDM)
=====
Value High Alarm  High Warn   Low Warn   Low Alarm
-----
Temperature (C)   +63.0    +80.0     +75.0     +15.0     -5.0
Supply Voltage (V) 3.25     3.46     3.43     3.17     3.13
=====

=====
Transceiver Lane Digital Diagnostic Monitoring (DDM)
=====
High Alarm  High Warn   Low Warn   Low Alarm
-----
Lane Tx Output Power (dBm)      0.00     -2.00     -13.00     -14.00
Lane Rx Optical Pwr (avg dBm)   2.00     0.00     -21.02     -23.01
-----
Lane ID Temp(C)/Alm    Tx Bias(mA)/Alm    Tx Pwr(dBm)/Alm    Rx Pwr(dBm)/Alm
-----
1          -          -          -40.00          -9.78
=====

=====
Coherent Optical Module
=====
Cfg Tx Target Power:  1.00 dBm
Cfg Rx LOS Thresh   : -23.00 dBm

Disp Control Mode   : automatic           Sweep Start Disp   : -25500 ps/nm
Cfg Dispersion      : 0 ps/nm             Sweep End Disp     : 2000 ps/nm
CPR Window Size    : 32 symbols           Rx LOS Reaction    : squelch
Compatibility       : longHaul
Cfg Tx Power Min    : -22.90 dBm           Cfg Tx Power Max   : 4.00 dBm

Cfg Alarms          : modflt mod netrx nettx hosttx
Alarm Status        :
Defect Points       :

Rx Q Margin         : 4.1 dB             Chromatic Disp     : 1 ps/nm
SNR/OSNR X Polar   : 19.2 dB / 35.7 dB  Diff Group Delay   : 3 ps
SNR/OSNR Y Polar   : 19.2 dB / 35.7 dB  Pre-FEC BER        : 3.340E-04

Module State        : tx0ff
Tx Turn-Up States   : init laserTurnUp laserReadyOff laserReady
                    modulatorConverge
Rx Turn-Up States   : init laserReady waitforInput adcSignal opticalLoc
                    demodLock

```

Step 3. To enable the laser and have the optic reach the “ready” state, the connector port as per the configured breakout must be enabled. Configure each connector port to be administratively set to **no shutdown** or set to **enable** as follows:

- In the classic CLI, use the **configure port 1/1/c1 no shutdown** command to enable each connector port.
- In the MD-CLI, use the **configure port 1/1/c1 admin-state enable** command to enable each connector port.

2.19.3 Provisioning DWDM coherent parameters

About this task

This procedure describes how to configure the following DWDM coherent parameters.
Classic CLI command

```
*A:Dut-A# configure port 1/1/c1 dwdm coherent
- coherent

      compatibility - Configure the optical mode and rate of operation
      cpr-window-size - Configure the Carrier Phase Recovery window size
      dispersion - Configure the dispersion compensation
      mode - Configure the dispersion algorithm mode
[no] report-alarms - Enable/disable logging of alarms
[no] rx-los-reaction - Configure the reaction to an RX LOS
      rx-los-thresh - Configure the average input power LOS (Loss of Signal)
                    threshold
      sweep - Configure the dispersion sweep start and end values
      target-power - Configure the desired average output power
```

MD-CLI command

```
*[pr:/configure port 1/1/c1 dwdm]
A:admin@csees-V27# coherent ?

coherent

apply-groups - Apply a configuration group at this level
apply-groups-exclude - Exclude a configuration group at this level
compatibility - Optical mode and rate of operation
cpr-window-size - Window size for the carrier phase recovery
dispersion - Residual chromatic dispersion compensation
mode - Mode used to compensate for chromatic dispersion
report-alarm + Enter the report-alarm context
rx-los-reaction - Reaction to an RX LOS
rx-los-thresh - Average input power LOS threshold
sweep + Enter the sweep context
target-power - Average output power target for the port
```



Note: For ZR and ZR+ optics, only the **compatibility**, **report-alarms**, **rx-los-thresh**, and **target-power** commands are applicable.

Procedure

Step 1. Configure the coherent **compatibility** command using the following context. The example below displays a **compatibility metro** configuration.

Example

configure port 1/1/c1 dwdm coherent compatibility metro (classic CLI)

```
*A:Dut-A# configure port 1/1/c1 dwdm coherent compatibility
- compatibility <mode>

<mode>                : long-haul|metro|access|interop|interop2|interop3|
                       long-haul-non-diff

*A:Dut-B# configure port 1/1/c1 dwdm coherent compatibility metro
```


Example

configure port 1/1/c1 dwdm coherent compatibility metro (MD-CLI)

```
*[pr:/configure port 1/1/c1 dwdm coherent]
A:admin@cses-V27# compatibility ?

compatibility <keyword>
<keyword> - (long-haul|metro|access|interop|interop2|interop3|long-haul-
            non-differential)
Default   - long-haul

            Optical mode and rate of operation

*[pr:/configure port 1/1/c1 dwdm coherent]
A:admin@cses-V27# compatibility metro
```

The coherent compatibility applies different optic application codes. The command default is **long-haul** and to change what has been previously configured, the user must shut down or disable the 1/1/c1 connector before provisioning for coherent compatibility.

For any DWDM compatibility changes, after the 1/1/c1 connector is configured as **no shutdown** in the classic CLI or **admin-state enable** in the MD-CLI, the coherent module initializes and transitions through various states as visible in the “Module State” field until it is ready again.

- Step 2.** Configure the **report-alarms** or **report-alarm** command using the following context. The example below displays a configuration where the alarms are not reported.

configure port 1/1/c1 dwdm coherent no report-alarms modflt mod netrx nettx hosttx (classic CLI)

Example

```
*A:Dut-A# configure port 1/1/c1 dwdm coherent report-alarms
- no report-alarms [modflt] [mod] [netrx] [nettx] [hosttx]
- report-alarms [modflt] [mod] [netrx] [nettx] [hosttx]

*A:Dut-B# configure port 1/1/c1 dwdm coherent no report-alarms modflt
mod netrx nettx hosttx
```

configure port 1/1/c1 dwdm coherent report-alarm (MD-CLI)

Example

```
*[pr:/configure port 1/1/c1 dwdm coherent]
A:admin@cses-V27# report-alarm ?

report-alarm

hosttx - Report the host (electrical side) transmit alarm
mod    - Report the module alarm
modflt - Report the module fault alarm
netrx  - Report the network (optical side) receive alarm
nettx  - Report the network (optical side) transmit alarm

*[pr:/configure port 1/1/c1 dwdm coherent report-alarm]
A:admin@cses-V27# info
modflt false
mod false
netrx false
nettx false
hosttx false
```

When an alarm is reported, it is logged in the event logs and a trap is sent to the Network Management System (NMS). For details about the alarms and defect points, see the TIMETRA-PORT-MIB.mib. By default, all alarms are reported.

- Step 3.** Configure the **rx-los-thresh** command using the following context. The example below displays a configuration of -18 dBm.

configure port 1/1/c1 dwdm coherent rx-los-thresh -18 (classic CLI)

Example

```
*A:Dut-A# configure port 1/1/c1 dwdm coherent rx-los-thresh
- rx-los-thresh <threshold>

<threshold>          : [-30.00..-13.00] dBm - Default: -23.00

*A:Dut-B# configure port 1/1/c1 dwdm coherent rx-los-thresh -18
```

configure port 1/1/c1 dwdm coherent rx-los-thresh -18 (MD-CLI)

Example

```
*[pr:/configure port 1/1/c1 dwdm coherent]
A:admin@cses-V27# rx-los-thresh ?

rx-los-thresh <decimal-number>
<decimal-number> - <-30.0..-13.0> - decibel-milliwatts, 2 fraction digits
Default          - -23

Average input power LOS threshold

*[pr:/configure port 1/1/c1 dwdm coherent]
A:admin@cses-V27# rx-los-thresh -18
```

This command configures the received optical power threshold at which the Loss of Signal (LOS) alarm is declared. To clear the LOS alarm, the received optical power must be higher than the configured **rx-los-thresh** command.

- Step 4.** Configure the **target-power** command using the following context. The example below displays a configuration of -12 dBm.

configure port 1/1/c1 dwdm coherent target-power -12 (classic CLI)

Example

```
*A:Dut-A# configure port 1/1/c1 dwdm coherent target-power
- target-power <power>

<power>              : [-20.00..3.00] dBm - Default: 1.00

*A:Dut-A# configure port 1/1/c1 dwdm coherent target-power -12
```

configure port 1/1/c1 dwdm coherent target-power -12 (MD-CLI)

Example

```
*[pr:/configure port 1/1/c1 dwdm coherent]
A:admin@cses-V27# target-power ?

target-power <decimal-number>
<decimal-number> - <-20.0..3.0> - decibel-milliwatts, 2 fraction digits
```

```
Default - 1
Average output power target for the port
```

```
*[pr:/configure port 1/1/c1 dwdm coherent]
A:admin@cses-V27# target-power -12
```

The **target-power** command configures the average output power target for the port.

Expected outcome

The actual measured Tx power along with the configured power is displayed by using the **show port 1/1/c1 detail** command.

```
=====
Transceiver Lane Digital Diagnostic Monitoring (DDM)
=====
High Alarm   High Warn   Low Warn   Low Alarm
-----
Lane Tx Output Power (dBm)      0.00      -2.00      -13.00     -14.00
Lane Rx Optical Pwr (avg dBm)   2.00       0.00      -21.02     -23.01
-----
Lane ID Temp(C)/Alm   Tx Bias(mA)/Alm   Tx Pwr(dBm)/Alm   Rx Pwr(dBm)/Alm
-----
1 - - -11.91 -9.86
=====
* indicates that the corresponding row element may have been truncated.
=====
Coherent Optical Module
=====
Cfg Tx Target Power: -12.00 dBm
Cfg Rx LOS Thresh : -23.00 dBm

Disp Control Mode : automatic           Sweep Start Disp : -25500 ps/nm
Cfg Dispersion    : 0 ps/nm             Sweep End Disp   : 2000 ps/nm
CPR Window Size  : 32 symbols           Rx LOS Reaction  : squelch
Compatibility     : longHaul
Cfg Tx Power Min  : -22.90 dBm          Cfg Tx Power Max : 4.00 dBm

Cfg Alarms        : modflt mod netrx nettx hosttx
Alarm Status      :
Defect Points     :

Rx Q Margin       : 4.1 dB               Chromatic Disp   : 1 ps/nm
SNR/OSNR X Polar  : 19.1 dB / 35.7 dB     Diff Group Delay : 2 ps
SNR/OSNR Y Polar  : 19.1 dB / 35.7 dB     Pre-FEC BER      : 3.430E-04

Module State      : ready
Tx Turn-Up States : init laserTurnUp laserReadyOff laserReady
                  : modulatorConverge outputPowerAdjust
Rx Turn-Up States : init laserReady waitforInput adcSignal opticalLock
                  : demodLock
-----
Coherent Optical Port Statistics (Elapsed Seconds: 588)
-----
Statistic          Current      Average      Minimum      Maximum
-----
Rx BER              3.430E-04   3.415E-04   3.050E-04   3.780E-04
Rx SNR (dB)         19.2        19.1         19.1         19.2
Rx OSNR (dB)        35.7        35.7         35.7         35.7
Rx Chromatic Disp (ps/nm) 1              2              1              3
-----
```

Rx Diff Group Delay (ps)	2	2	2	3
Rx Freq Offset (MHz)	287	284	182	387
Rx Q (dB)	10.6	10.5	10.5	10.6
Rx Per-Channel Power (dBm)	-9.69	-9.63	-9.78	-9.49
Tx Power (dBm)	-11.93	-10.69	-11.93	-9.72

3 Standards and protocol support



Note:

The information provided in this chapter is subject to change without notice and may not apply to all platforms.

Nokia assumes no responsibility for inaccuracies.

3.1 Access Node Control Protocol (ANCP)

draft-ietf-ancp-protocol-02, *Protocol for Access Node Control Mechanism in Broadband Networks*

RFC 5851, *Framework and Requirements for an Access Node Control Mechanism in Broadband Multi-Service Networks*

3.2 Application Assurance (AA)

3GPP Release 12, *ADC rules over Gx interfaces*

RFC 3507, *Internet Content Adaptation Protocol (ICAP)*

3.3 Bidirectional Forwarding Detection (BFD)

draft-ietf-idr-bgp-ls-sbfd-extensions-01, *BGP Link-State Extensions for Seamless BFD*

RFC 5880, *Bidirectional Forwarding Detection (BFD)*

RFC 5881, *Bidirectional Forwarding Detection (BFD) IPv4 and IPv6 (Single Hop)*

RFC 5882, *Generic Application of Bidirectional Forwarding Detection (BFD)*

RFC 5883, *Bidirectional Forwarding Detection (BFD) for Multihop Paths*

RFC 7130, *Bidirectional Forwarding Detection (BFD) on Link Aggregation Group (LAG) Interfaces*

RFC 7880, *Seamless Bidirectional Forwarding Detection (S-BFD)*

RFC 7881, *Seamless Bidirectional Forwarding Detection (S-BFD) for IPv4, IPv6, and MPLS*

RFC 7883, *Advertising Seamless Bidirectional Forwarding Detection (S-BFD) Discriminators in IS-IS*

RFC 7884, *OSPF Extensions to Advertise Seamless Bidirectional Forwarding Detection (S-BFD) Target Discriminators*

3.4 Border Gateway Protocol (BGP)

draft-gredler-idr-bgplu-epe-14, *Egress Peer Engineering using BGP-LU*

draft-hares-idr-update-attr-low-bits-fix-01, *Update Attribute Flag Low Bits Clarification*
draft-ietf-idr-add-paths-guidelines-08, *Best Practices for Advertisement of Multiple Paths in IBGP*
draft-ietf-idr-best-external-03, *Advertisement of the best external route in BGP*
draft-ietf-idr-bgp-flowspec-oid-03, *Revised Validation Procedure for BGP Flow Specifications*
draft-ietf-idr-bgp-gr-notification-01, *Notification Message support for BGP Graceful Restart*
draft-ietf-idr-bgp-ls-app-specific-attr-16, *Application-Specific Attributes Advertisement with BGP Link-State*
draft-ietf-idr-bgp-ls-flex-algo-06, *Flexible Algorithm Definition Advertisement with BGP Link-State*
draft-ietf-idr-bgp-optimal-route-reflection-10, *BGP Optimal Route Reflection (BGP-ORR)*
draft-ietf-idr-error-handling-03, *Revised Error Handling for BGP UPDATE Messages*
draft-ietf-idr-flowspec-interfaceset-03, *Applying BGP flowspec rules on a specific interface set*
draft-ietf-idr-flowspec-path-redirect-05, *Flowspec Indirection-id Redirect – localised ID*
draft-ietf-idr-flowspec-redirect-ip-02, *BGP Flow-Spec Redirect to IP Action*
draft-ietf-idr-link-bandwidth-03, *BGP Link Bandwidth Extended Community*
draft-ietf-idr-long-lived-gr-00, *Support for Long-lived BGP Graceful Restart*
RFC 1772, *Application of the Border Gateway Protocol in the Internet*
RFC 1997, *BGP Communities Attribute*
RFC 2385, *Protection of BGP Sessions via the TCP MD5 Signature Option*
RFC 2439, *BGP Route Flap Damping*
RFC 2545, *Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing*
RFC 2858, *Multiprotocol Extensions for BGP-4*
RFC 2918, *Route Refresh Capability for BGP-4*
RFC 4271, *A Border Gateway Protocol 4 (BGP-4)*
RFC 4360, *BGP Extended Communities Attribute*
RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*
RFC 4456, *BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)*
RFC 4486, *Subcodes for BGP Cease Notification Message*
RFC 4659, *BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN*
RFC 4684, *Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)*
RFC 4724, *Graceful Restart Mechanism for BGP – helper mode*
RFC 4760, *Multiprotocol Extensions for BGP-4*
RFC 4798, *Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)*
RFC 5004, *Avoid BGP Best Path Transitions from One External to Another*
RFC 5065, *Autonomous System Confederations for BGP*
RFC 5291, *Outbound Route Filtering Capability for BGP-4*
RFC 5396, *Textual Representation of Autonomous System (AS) Numbers – asplain*
RFC 5492, *Capabilities Advertisement with BGP-4*

RFC 5668, *4-Octet AS Specific BGP Extended Community*
RFC 6286, *Autonomous-System-Wide Unique BGP Identifier for BGP-4*
RFC 6793, *BGP Support for Four-Octet Autonomous System (AS) Number Space*
RFC 6810, *The Resource Public Key Infrastructure (RPKI) to Router Protocol*
RFC 6811, *Prefix Origin Validation*
RFC 6996, *Autonomous System (AS) Reservation for Private Use*
RFC 7311, *The Accumulated IGP Metric Attribute for BGP*
RFC 7606, *Revised Error Handling for BGP UPDATE Messages*
RFC 7607, *Codification of AS 0 Processing*
RFC 7674, *Clarification of the Flowspec Redirect Extended Community*
RFC 7752, *North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP*
RFC 7854, *BGP Monitoring Protocol (BMP)*
RFC 7911, *Advertisement of Multiple Paths in BGP*
RFC 7999, *BLACKHOLE Community*
RFC 8092, *BGP Large Communities Attribute*
RFC 8097, *BGP Prefix Origin Validation State Extended Community*
RFC 8212, *Default External BGP (EBGP) Route Propagation Behavior without Policies*
RFC 8277, *Using BGP to Bind MPLS Labels to Address Prefixes*
RFC 8571, *BGP - Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions*
RFC 8950, *Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop*
RFC 8955, *Dissemination of Flow Specification Rules*
RFC 8956, *Dissemination of Flow Specification Rules for IPv6*
RFC 9086, *Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing BGP Egress Peer Engineering*

3.5 Broadband Network Gateway (BNG) Control and User Plane Separation (CUPS)

3GPP 23.007, *Restoration procedures*
3GPP 29.244, *Interface between the Control Plane and the User Plane nodes*
3GPP 29.281, *General Packet Radio System (GPRS) Tunneling Protocol User Plane (GTPv1-U)*
BBF TR-459, *Control and User Plane Separation for a Disaggregated BNG*
RFC 8300, *Network Service Header (NSH)*

3.6 Certificate management

- RFC 4210, *Internet X.509 Public Key Infrastructure Certificate Management Protocol (CMP)*
- RFC 4211, *Internet X.509 Public Key Infrastructure Certificate Request Message Format (CRMF)*
- RFC 5280, *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*
- RFC 6712, *Internet X.509 Public Key Infrastructure -- HTTP Transfer for the Certificate Management Protocol (CMP)*
- RFC 7030, *Enrollment over Secure Transport*
- RFC 7468, *Textual Encodings of PKIX, PKCS, and CMS Structures*

3.7 Circuit emulation

- RFC 4553, *Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)*
- RFC 5086, *Structure-Aware Time Division Multiplexed (TDM) Circuit Emulation Service over Packet Switched Network (CESoPSN)*
- RFC 5287, *Control Protocol Extensions for the Setup of Time-Division Multiplexing (TDM) Pseudowires in MPLS Networks*

3.8 Ethernet

- IEEE 802.1AB, *Station and Media Access Control Connectivity Discovery*
- IEEE 802.1ad, *Provider Bridges*
- IEEE 802.1ag, *Connectivity Fault Management*
- IEEE 802.1ah, *Provider Backbone Bridges*
- IEEE 802.1ak, *Multiple Registration Protocol*
- IEEE 802.1aq, *Shortest Path Bridging*
- IEEE 802.1ax, *Link Aggregation*
- IEEE 802.1D, *MAC Bridges*
- IEEE 802.1p, *Traffic Class Expediting*
- IEEE 802.1Q, *Virtual LANs*
- IEEE 802.1s, *Multiple Spanning Trees*
- IEEE 802.1w, *Rapid Reconfiguration of Spanning Tree*
- IEEE 802.1X, *Port Based Network Access Control*
- IEEE 802.3ac, *VLAN Tag*
- IEEE 802.3ad, *Link Aggregation*
- IEEE 802.3ah, *Ethernet in the First Mile*
- IEEE 802.3x, *Ethernet Flow Control*

ITU-T G.8031/Y.1342, *Ethernet Linear Protection Switching*
ITU-T G.8032/Y.1344, *Ethernet Ring Protection Switching*
ITU-T Y.1731, *OAM functions and mechanisms for Ethernet based networks*

3.9 Ethernet VPN (EVPN)

draft-ietf-bess-evpn-ipvpn-interworking-06, *EVPN Interworking with IPVPN*
draft-ietf-bess-evpn-irb-mcast-04, *EVPN Optimized Inter-Subnet Multicast (OISM) Forwarding – ingress replication*
draft-ietf-bess-evpn-pref-df-06, *Preference-based EVPN DF Election*
draft-ietf-bess-evpn-unequal-lb-16, *Weighted Multi-Path Procedures for EVPN Multi-Homing – section 9*
draft-ietf-bess-evpn-virtual-eth-segment-06, *EVPN Virtual Ethernet Segment*
draft-ietf-bess-pbb-evpn-isid-cmacflush-00, *PBB-EVPN ISID-based CMAC-Flush*
draft-sajassi-bess-evpn-ip-aliasing-05, *EVPN Support for L3 Fast Convergence and Aliasing/Backup Path – IP Prefix routes*
RFC 7432, *BGP MPLS-Based Ethernet VPN*
RFC 7623, *Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)*
RFC 8214, *Virtual Private Wire Service Support in Ethernet VPN*
RFC 8317, *Ethernet-Tree (E-Tree) Support in Ethernet VPN (EVPN) an Provider Backbone Bridging EVPN (PBB-EVPN)*
RFC 8365, *A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)*
RFC 8560, *Seamless Integration of Ethernet VPN (EVPN) with Virtual Private LAN Service (VPLS) and Their Provider Backbone Bridge (PBB) Equivalents*
RFC 8584, *DF Election and AC-influenced DF Election*
RFC 9136, *IP Prefix Advertisement in Ethernet VPN (EVPN)*
RFC 9161, *Operational Aspects of Proxy ARP/ND in Ethernet Virtual Private Networks*
RFC 9251, *Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Proxies for Ethernet VPN (EVPN)*

3.10 gRPC Remote Procedure Calls (gRPC)

cert.proto version 0.1.0, *gRPC Network Operations Interface (gNOI) Certificate Management Service*
file.proto version 0.1.0, *gRPC Network Operations Interface (gNOI) File Service*
gnmi.proto version 0.7.0, *gRPC Network Management Interface (gNMI) Service Specification*
PROTOCOL-HTTP2, *gRPC over HTTP2*
system.proto Version 1.0.0, *gRPC Network Operations Interface (gNOI) System Service*

3.11 Intermediate System to Intermediate System (IS-IS)

draft-ietf-isis-mi-02, *IS-IS Multi-Instance*

draft-kaplan-isis-ext-eth-02, *Extended Ethernet Frame Size Support*

ISO/IEC 10589:2002 Second Edition, *Intermediate system to Intermediate system intra-domain routing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473)*

RFC 1195, *Use of OSI IS-IS for Routing in TCP/IP and Dual Environments*

RFC 2973, *IS-IS Mesh Groups*

RFC 3359, *Reserved Type, Length and Value (TLV) Codepoints in Intermediate System to Intermediate System*

RFC 3719, *Recommendations for Interoperable Networks using Intermediate System to Intermediate System (IS-IS)*

RFC 3787, *Recommendations for Interoperable IP Networks using Intermediate System to Intermediate System (IS-IS)*

RFC 4971, *Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information*

RFC 5120, *M-ISIS: Multi Topology (MT) Routing in IS-IS*

RFC 5130, *A Policy Control Mechanism in IS-IS Using Administrative Tags*

RFC 5301, *Dynamic Hostname Exchange Mechanism for IS-IS*

RFC 5302, *Domain-wide Prefix Distribution with Two-Level IS-IS*

RFC 5303, *Three-Way Handshake for IS-IS Point-to-Point Adjacencies*

RFC 5304, *IS-IS Cryptographic Authentication*

RFC 5305, *IS-IS Extensions for Traffic Engineering TE*

RFC 5306, *Restart Signaling for IS-IS – helper mode*

RFC 5307, *IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*

RFC 5308, *Routing IPv6 with IS-IS*

RFC 5309, *Point-to-Point Operation over LAN in Link State Routing Protocols*

RFC 5310, *IS-IS Generic Cryptographic Authentication*

RFC 6119, *IPv6 Traffic Engineering in IS-IS*

RFC 6213, *IS-IS BFD-Enabled TLV*

RFC 6232, *Purge Originator Identification TLV for IS-IS*

RFC 6233, *IS-IS Registry Extension for Purges*

RFC 6329, *IS-IS Extensions Supporting IEEE 802.1aq Shortest Path Bridging*

RFC 7775, *IS-IS Route Preference for Extended IP and IPv6 Reachability*

RFC 7794, *IS-IS Prefix Attributes for Extended IPv4 and IPv6 Reachability*

RFC 7987, *IS-IS Minimum Remaining Lifetime*

RFC 8202, *IS-IS Multi-Instance – single topology*

RFC 8570, *IS-IS Traffic Engineering (TE) Metric Extensions – Min/Max Unidirectional Link Delay metric for flex-algo, RSVP, SR-TE*

RFC 8919, *IS-IS Application-Specific Link Attributes*

3.12 Internet Protocol (IP) Fast Reroute (FRR)

draft-ietf-rtgwg-lfa-manageability-08, *Operational management of Loop Free Alternates*

RFC 5286, *Basic Specification for IP Fast Reroute: Loop-Free Alternates*

RFC 7431, *Multicast-Only Fast Reroute*

RFC 7490, *Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)*

RFC 8518, *Selection of Loop-Free Alternates for Multi-Homed Prefixes*

3.13 Internet Protocol (IP) general

draft-grant-tacacs-02, *The TACACS+ Protocol*

RFC 768, *User Datagram Protocol*

RFC 793, *Transmission Control Protocol*

RFC 854, *Telnet Protocol Specifications*

RFC 1350, *The TFTP Protocol (revision 2)*

RFC 2347, *TFTP Option Extension*

RFC 2348, *TFTP Blocksize Option*

RFC 2349, *TFTP Timeout Interval and Transfer Size Options*

RFC 2428, *FTP Extensions for IPv6 and NATs*

RFC 2617, *HTTP Authentication: Basic and Digest Access Authentication*

RFC 2784, *Generic Routing Encapsulation (GRE)*

RFC 2818, *HTTP Over TLS*

RFC 2890, *Key and Sequence Number Extensions to GRE*

RFC 3164, *The BSD syslog Protocol*

RFC 4250, *The Secure Shell (SSH) Protocol Assigned Numbers*

RFC 4251, *The Secure Shell (SSH) Protocol Architecture*

RFC 4252, *The Secure Shell (SSH) Authentication Protocol – publickey, password*

RFC 4253, *The Secure Shell (SSH) Transport Layer Protocol*

RFC 4254, *The Secure Shell (SSH) Connection Protocol*

RFC 4511, *Lightweight Directory Access Protocol (LDAP): The Protocol*

RFC 4513, *Lightweight Directory Access Protocol (LDAP): Authentication Methods and Security Mechanisms – TLS*

RFC 4632, *Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan*

RFC 5082, *The Generalized TTL Security Mechanism (GTSM)*

RFC 5246, *The Transport Layer Security (TLS) Protocol Version 1.2 – TLS client, RSA public key*

RFC 5425, *Transport Layer Security (TLS) Transport Mapping for Syslog – RFC 3164 with TLS*

RFC 5656, *Elliptic Curve Algorithm Integration in the Secure Shell Transport Layer – ECDSA*

RFC 5925, *The TCP Authentication Option*

RFC 5926, *Cryptographic Algorithms for the TCP Authentication Option (TCP-AO)*

RFC 6398, *IP Router Alert Considerations and Usage – MLD*

RFC 6528, *Defending against Sequence Number Attacks*

RFC 7011, *Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information*

RFC 7012, *Information Model for IP Flow Information Export*

RFC 7230, *Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing*

RFC 7231, *Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content*

RFC 7232, *Hypertext Transfer Protocol (HTTP/1.1): Conditional Requests*

RFC 7301, *Transport Layer Security (TLS) Application Layer Protocol Negotiation Extension*

RFC 7616, *HTTP Digest Access Authentication*

RFC 8446, *The Transport Layer Security (TLS) Protocol Version 1.3*

3.14 Internet Protocol (IP) multicast

cisco-ipmulticast/pim-autorp-spec01, *Auto-RP: Automatic discovery of Group-to-RP mappings for IP multicast – version 1*

draft-ietf-bier-pim-signaling-08, *PIM Signaling Through BIER Core*

draft-ietf-idmr-traceroute-ipm-07, *A "traceroute" facility for IP Multicast*

draft-ietf-l2vpn-vpls-pim-snooping-07, *Protocol Independent Multicast (PIM) over Virtual Private LAN Service (VPLS)*

RFC 1112, *Host Extensions for IP Multicasting*

RFC 2236, *Internet Group Management Protocol, Version 2*

RFC 2365, *Administratively Scoped IP Multicast*

RFC 2375, *IPv6 Multicast Address Assignments*

RFC 2710, *Multicast Listener Discovery (MLD) for IPv6*

RFC 3306, *Unicast-Prefix-based IPv6 Multicast Addresses*

RFC 3376, *Internet Group Management Protocol, Version 3*

RFC 3446, *Anycast Rendezvous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)*

RFC 3590, *Source Address Selection for the Multicast Listener Discovery (MLD) Protocol*

RFC 3618, *Multicast Source Discovery Protocol (MSDP)*

RFC 3810, *Multicast Listener Discovery Version 2 (MLDv2) for IPv6*

RFC 3956, *Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address*

RFC 3973, *Protocol Independent Multicast - Dense Mode (PIM-DM): Protocol Specification (Revised) – auto-RP groups*

RFC 4541, *Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches*

RFC 4604, *Using Internet Group Management Protocol Version 3 (IGMPv3) and Multicast Listener Discovery Protocol Version 2 (MLDv2) for Source-Specific Multicast*

RFC 4607, *Source-Specific Multicast for IP*

RFC 4608, *Source-Specific Protocol Independent Multicast in 232/8*

RFC 4610, *Anycast-RP Using Protocol Independent Multicast (PIM)*

RFC 4611, *Multicast Source Discovery Protocol (MSDP) Deployment Scenarios*

RFC 5059, *Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)*

RFC 5186, *Internet Group Management Protocol Version 3 (IGMPv3) / Multicast Listener Discovery Version 2 (MLDv2) and Multicast Routing Protocol Interaction*

RFC 5384, *The Protocol Independent Multicast (PIM) Join Attribute Format*

RFC 5496, *The Reverse Path Forwarding (RPF) Vector TLV*

RFC 6037, *Cisco Systems' Solution for Multicast in MPLS/BGP IP VPNs*

RFC 6512, *Using Multipoint LDP When the Backbone Has No Route to the Root*

RFC 6513, *Multicast in MPLS/BGP IP VPNs*

RFC 6514, *BGP Encodings and Procedures for Multicast in MPLS/IP VPNs*

RFC 6515, *IPv4 and IPv6 Infrastructure Addresses in BGP Updates for Multicast VPNs*

RFC 6516, *IPv6 Multicast VPN (MVPN) Support Using PIM Control Plane and Selective Provider Multicast Service Interface (S-PMSI) Join Messages*

RFC 6625, *Wildcards in Multicast VPN Auto-Discover Routes*

RFC 6826, *Multipoint LDP In-Band Signaling for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Path*

RFC 7246, *Multipoint Label Distribution Protocol In-Band Signaling in a Virtual Routing and Forwarding (VRF) Table Context*

RFC 7385, *IANA Registry for P-Multicast Service Interface (PMSI) Tunnel Type Code Points*

RFC 7716, *Global Table Multicast with BGP Multicast VPN (BGP-MVPN) Procedures*

RFC 7761, *Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)*

RFC 8279, *Multicast Using Bit Index Explicit Replication (BIER)*

RFC 8296, *Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks – MPLS encapsulation*

RFC 8401, *Bit Index Explicit Replication (BIER) Support via IS-IS*

RFC 8444, *OSPFv2 Extensions for Bit Index Explicit Replication (BIER)*

RFC 8487, *Mtrace Version 2: Traceroute Facility for IP Multicast*

RFC 8534, *Explicit Tracking with Wildcard Routes in Multicast VPN – (C-*,C-*) wildcard*
RFC 8556, *Multicast VPN Using Bit Index Explicit Replication (BIER)*

3.15 Internet Protocol (IP) version 4

RFC 791, *Internet Protocol*
RFC 792, *Internet Control Message Protocol*
RFC 826, *An Ethernet Address Resolution Protocol*
RFC 951, *Bootstrap Protocol (BOOTP) – relay*
RFC 1034, *Domain Names - Concepts and Facilities*
RFC 1035, *Domain Names - Implementation and Specification*
RFC 1191, *Path MTU Discovery – router specification*
RFC 1519, *Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy*
RFC 1534, *Interoperation between DHCP and BOOTP*
RFC 1542, *Clarifications and Extensions for the Bootstrap Protocol*
RFC 1812, *Requirements for IPv4 Routers*
RFC 1918, *Address Allocation for Private Internets*
RFC 2003, *IP Encapsulation within IP*
RFC 2131, *Dynamic Host Configuration Protocol*
RFC 2132, *DHCP Options and BOOTP Vendor Extensions*
RFC 2401, *Security Architecture for Internet Protocol*
RFC 3021, *Using 31-Bit Prefixes on IPv4 Point-to-Point Links*
RFC 3046, *DHCP Relay Agent Information Option (Option 82)*
RFC 3768, *Virtual Router Redundancy Protocol (VRRP)*
RFC 4884, *Extended ICMP to Support Multi-Part Messages – ICMPv4 and ICMPv6 Time Exceeded*

3.16 Internet Protocol (IP) version 6

RFC 2464, *Transmission of IPv6 Packets over Ethernet Networks*
RFC 2529, *Transmission of IPv6 over IPv4 Domains without Explicit Tunnels*
RFC 3122, *Extensions to IPv6 Neighbor Discovery for Inverse Discovery Specification*
RFC 3315, *Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*
RFC 3587, *IPv6 Global Unicast Address Format*
RFC 3596, *DNS Extensions to Support IP version 6*
RFC 3633, *IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6*
RFC 3646, *DNS Configuration options for Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*

RFC 3736, *Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6*
RFC 3971, *SEcure Neighbor Discovery (SEND)*
RFC 3972, *Cryptographically Generated Addresses (CGA)*
RFC 4007, *IPv6 Scoped Address Architecture*
RFC 4193, *Unique Local IPv6 Unicast Addresses*
RFC 4291, *Internet Protocol Version 6 (IPv6) Addressing Architecture*
RFC 4443, *Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification*
RFC 4861, *Neighbor Discovery for IP version 6 (IPv6)*
RFC 4862, *IPv6 Stateless Address Autoconfiguration – router functions*
RFC 4890, *Recommendations for Filtering ICMPv6 Messages in Firewalls*
RFC 4941, *Privacy Extensions for Stateless Address Autoconfiguration in IPv6*
RFC 5007, *DHCPv6 Leasequery*
RFC 5095, *Deprecation of Type 0 Routing Headers in IPv6*
RFC 5722, *Handling of Overlapping IPv6 Fragments*
RFC 5798, *Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6 – IPv6*
RFC 5952, *A Recommendation for IPv6 Address Text Representation*
RFC 6092, *Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service – Internet Control and Management, Upper-Layer Transport Protocols, UDP Filters, IPsec and Internet Key Exchange (IKE), TCP Filters*
RFC 6106, *IPv6 Router Advertisement Options for DNS Configuration*
RFC 6164, *Using 127-Bit IPv6 Prefixes on Inter-Router Links*
RFC 6437, *IPv6 Flow Label Specification*
RFC 6603, *Prefix Exclude Option for DHCPv6-based Prefix Delegation*
RFC 8021, *Generation of IPv6 Atomic Fragments Considered Harmful*
RFC 8200, *Internet Protocol, Version 6 (IPv6) Specification*
RFC 8201, *Path MTU Discovery for IP version 6*

3.17 Internet Protocol Security (IPsec)

draft-ietf-ipsec-isakmp-mode-cfg-05, *The ISAKMP Configuration Method*
draft-ietf-ipsec-isakmp-xauth-06, *Extended Authentication within ISAKMP/Oakley (XAUTH)*
RFC 2401, *Security Architecture for the Internet Protocol*
RFC 2403, *The Use of HMAC-MD5-96 within ESP and AH*
RFC 2404, *The Use of HMAC-SHA-1-96 within ESP and AH*
RFC 2405, *The ESP DES-CBC Cipher Algorithm With Explicit IV*
RFC 2406, *IP Encapsulating Security Payload (ESP)*

RFC 2407, *IPsec Domain of Interpretation for ISAKMP (IPsec DoI)*
RFC 2408, *Internet Security Association and Key Management Protocol (ISAKMP)*
RFC 2409, *The Internet Key Exchange (IKE)*
RFC 2410, *The NULL Encryption Algorithm and Its Use With IPsec*
RFC 2560, *X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP*
RFC 3526, *More Modular Exponential (MODP) Diffie-Hellman group for Internet Key Exchange (IKE)*
RFC 3566, *The AES-XCBC-MAC-96 Algorithm and Its Use With IPsec*
RFC 3602, *The AES-CBC Cipher Algorithm and Its Use with IPsec*
RFC 3706, *A Traffic-Based Method of Detecting Dead Internet Key Exchange (IKE) Peers*
RFC 3947, *Negotiation of NAT-Traversal in the IKE*
RFC 3948, *UDP Encapsulation of IPsec ESP Packets*
RFC 4106, *The Use of Galois/Counter Mode (GCM) in IPsec ESP*
RFC 4109, *Algorithms for Internet Key Exchange version 1 (IKEv1)*
RFC 4301, *Security Architecture for the Internet Protocol*
RFC 4303, *IP Encapsulating Security Payload*
RFC 4307, *Cryptographic Algorithms for Use in the Internet Key Exchange Version 2 (IKEv2)*
RFC 4308, *Cryptographic Suites for IPsec*
RFC 4434, *The AES-XCBC-PRF-128 Algorithm for the Internet Key Exchange Protocol (IKE)*
RFC 4543, *The Use of Galois Message Authentication Code (GMAC) in IPsec ESP and AH*
RFC 4754, *IKE and IKEv2 Authentication Using the Elliptic Curve Digital Signature Algorithm (ECDSA)*
RFC 4835, *Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload (ESP) and Authentication Header (AH)*
RFC 4868, *Using HMAC-SHA-256, HMAC-SHA-384, and HMAC-SHA-512 with IPsec*
RFC 4945, *The Internet IP Security PKI Profile of IKEv1/ISAKMP, IKEv2 and PKIX*
RFC 5019, *The Lightweight Online Certificate Status Protocol (OCSP) Profile for High-Volume Environments*
RFC 5282, *Using Authenticated Encryption Algorithms with the Encrypted Payload of the IKEv2 Protocol*
RFC 5903, *ECP Groups for IKE and IKEv2*
RFC 5996, *Internet Key Exchange Protocol Version 2 (IKEv2)*
RFC 5998, *An Extension for EAP-Only Authentication in IKEv2*
RFC 6379, *Suite B Cryptographic Suites for IPsec*
RFC 6380, *Suite B Profile for Internet Protocol Security (IPsec)*
RFC 6960, *X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP*
RFC 7296, *Internet Key Exchange Protocol Version 2 (IKEv2)*
RFC 7321, *Cryptographic Algorithm Implementation Requirements and Usage Guidance for Encapsulating Security Payload (ESP) and Authentication Header (AH)*
RFC 7383, *Internet Key Exchange Protocol Version 2 (IKEv2) Message Fragmentation*

RFC 7427, *Signature Authentication in the Internet Key Exchange Version 2 (IKEv2)*

3.18 Label Distribution Protocol (LDP)

draft-pdutta-mpls-ldp-adj-capability-00, *LDP Adjacency Capabilities*

draft-pdutta-mpls-ldp-v2-00, *LDP Version 2*

draft-pdutta-mpls-ldp-up-redundancy-00, *Upstream LSR Redundancy for Multi-point LDP Tunnels*

draft-pdutta-mpls-multi-ldp-instance-00, *Multiple LDP Instances*

draft-pdutta-mpls-ldp-hello-reduce-04, *Targeted LDP Hello Reduction*

RFC 3037, *LDP Applicability*

RFC 3478, *Graceful Restart Mechanism for Label Distribution Protocol – helper mode*

RFC 5036, *LDP Specification*

RFC 5283, *LDP Extension for Inter-Area Label Switched Paths (LSPs)*

RFC 5443, *LDP IGP Synchronization*

RFC 5561, *LDP Capabilities*

RFC 5919, *Signaling LDP Label Advertisement Completion*

RFC 6388, *Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths*

RFC 6512, *Using Multipoint LDP When the Backbone Has No Route to the Root*

RFC 6826, *Multipoint LDP in-band signaling for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths*

RFC 7032, *LDP Downstream-on-Demand in Seamless MPLS*

RFC 7473, *Controlling State Advertisements of Non-negotiated LDP Applications*

RFC 7552, *Updates to LDP for IPv6*

3.19 Layer Two Tunneling Protocol (L2TP) Network Server (LNS)

draft-mammoliti-l2tp-accessline-avp-04, *Layer 2 Tunneling Protocol (L2TP) Access Line Information Attribute Value Pair (AVP) Extensions*

RFC 2661, *Layer Two Tunneling Protocol "L2TP"*

RFC 2809, *Implementation of L2TP Compulsory Tunneling via RADIUS*

RFC 3438, *Layer Two Tunneling Protocol (L2TP) Internet Assigned Numbers: Internet Assigned Numbers Authority (IANA) Considerations Update*

RFC 3931, *Layer Two Tunneling Protocol - Version 3 (L2TPv3)*

RFC 4719, *Transport of Ethernet Frames over Layer 2 Tunneling Protocol Version 3 (L2TPv3)*

RFC 4951, *Fail Over Extensions for Layer 2 Tunneling Protocol (L2TP) "failover"*

3.20 Multiprotocol Label Switching (MPLS)

draft-ietf-mpls-lsp-ping-ospfv3-codepoint-02, *OSPFv3 CodePoint for MPLS LSP Ping*
RFC 3031, *Multiprotocol Label Switching Architecture*
RFC 3032, *MPLS Label Stack Encoding*
RFC 3270, *Multi-Protocol Label Switching (MPLS) Support of Differentiated Services – E-LSP*
RFC 3443, *Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks*
RFC 4023, *Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)*
RFC 4182, *Removing a Restriction on the use of MPLS Explicit NULL*
RFC 5332, *MPLS Multicast Encapsulations*
RFC 5884, *Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)*
RFC 6374, *Packet Loss and Delay Measurement for MPLS Networks – Delay Measurement, Channel Type 0x000C*
RFC 6424, *Mechanism for Performing Label Switched Path Ping (LSP Ping) over MPLS Tunnels*
RFC 6425, *Detecting Data Plane Failures in Point-to-Multipoint Multiprotocol Label Switching (MPLS) - Extensions to LSP Ping*
RFC 6790, *The Use of Entropy Labels in MPLS Forwarding*
RFC 7510, *Encapsulating MPLS in UDP*
RFC 7746, *Label Switched Path (LSP) Self-Ping*
RFC 7876, *UDP Return Path for Packet Loss and Delay Measurement for MPLS Networks – Delay Measurement*
RFC 8029, *Detecting Multiprotocol Label Switched (MPLS) Data-Plane Failures*

3.21 Multiprotocol Label Switching - Transport Profile (MPLS-TP)

RFC 5586, *MPLS Generic Associated Channel*
RFC 5921, *A Framework for MPLS in Transport Networks*
RFC 5960, *MPLS Transport Profile Data Plane Architecture*
RFC 6370, *MPLS Transport Profile (MPLS-TP) Identifiers*
RFC 6378, *MPLS Transport Profile (MPLS-TP) Linear Protection*
RFC 6426, *MPLS On-Demand Connectivity and Route Tracing*
RFC 6427, *MPLS Fault Management Operations, Administration, and Maintenance (OAM)*
RFC 6428, *Proactive Connectivity Verification, Continuity Check and Remote Defect indication for MPLS Transport Profile*
RFC 6478, *Pseudowire Status for Static Pseudowires*
RFC 7213, *MPLS Transport Profile (MPLS-TP) Next-Hop Ethernet Addressing*

3.22 Network Address Translation (NAT)

draft-ietf-behave-address-format-10, *IPv6 Addressing of IPv4/IPv6 Translators*
draft-ietf-behave-v6v4-xlate-23, *IP/ICMP Translation Algorithm*
draft-miles-behave-l2nat-00, *Layer2-Aware NAT*
draft-nishitani-cgn-02, *Common Functions of Large Scale NAT (LSN)*
RFC 4787, *Network Address Translation (NAT) Behavioral Requirements for Unicast UDP*
RFC 5382, *NAT Behavioral Requirements for TCP*
RFC 5508, *NAT Behavioral Requirements for ICMP*
RFC 6146, *Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers*
RFC 6333, *Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion*
RFC 6334, *Dynamic Host Configuration Protocol for IPv6 (DHCPv6) Option for Dual-Stack Lite*
RFC 6887, *Port Control Protocol (PCP)*
RFC 6888, *Common Requirements For Carrier-Grade NATs (CGNs)*
RFC 7753, *Port Control Protocol (PCP) Extension for Port-Set Allocation*
RFC 7915, *IP/ICMP Translation Algorithm*

3.23 Network Configuration Protocol (NETCONF)

RFC 5277, *NETCONF Event Notifications*
RFC 6020, *YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)*
RFC 6022, *YANG Module for NETCONF Monitoring*
RFC 6241, *Network Configuration Protocol (NETCONF)*
RFC 6242, *Using the NETCONF Protocol over Secure Shell (SSH)*
RFC 6243, *With-defaults Capability for NETCONF*
RFC 8342, *Network Management Datastore Architecture (NMDA) – Startup, Candidate, Running and Intended datastores*
RFC 8525, *YANG Library*
RFC 8526, *NETCONF Extensions to Support the Network Management Datastore Architecture – <get-data> operation*

3.24 Open Shortest Path First (OSPF)

RFC 1765, *OSPF Database Overflow*
RFC 2328, *OSPF Version 2*
RFC 3101, *The OSPF Not-So-Stubby Area (NSSA) Option*
RFC 3509, *Alternative Implementations of OSPF Area Border Routers*

RFC 3623, *Graceful OSPF Restart Graceful OSPF Restart – helper mode*
RFC 3630, *Traffic Engineering (TE) Extensions to OSPF Version 2*
RFC 4203, *OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)*
RFC 4222, *Prioritized Treatment of Specific OSPF Version 2 Packets and Congestion Avoidance*
RFC 4552, *Authentication/Confidentiality for OSPFv3*
RFC 4576, *Using a Link State Advertisement (LSA) Options Bit to Prevent Looping in BGP/MPLS IP Virtual Private Networks (VPNs)*
RFC 4577, *OSPF as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)*
RFC 5185, *OSPF Multi-Area Adjacency*
RFC 5187, *OSPFv3 Graceful Restart – helper mode*
RFC 5243, *OSPF Database Exchange Summary List Optimization*
RFC 5250, *The OSPF Opaque LSA Option*
RFC 5309, *Point-to-Point Operation over LAN in Link State Routing Protocols*
RFC 5340, *OSPF for IPv6*
RFC 5642, *Dynamic Hostname Exchange Mechanism for OSPF*
RFC 5709, *OSPFv2 HMAC-SHA Cryptographic Authentication*
RFC 5838, *Support of Address Families in OSPFv3*
RFC 6549, *OSPFv2 Multi-Instance Extensions*
RFC 6987, *OSPF Stub Router Advertisement*
RFC 7471, *OSPF Traffic Engineering (TE) Metric Extensions – Min/Max Unidirectional Link Delay metric for flex-algo, RSVP, SR-TE*
RFC 7684, *OSPFv2 Prefix/Link Attribute Advertisement*
RFC 7770, *Extensions to OSPF for Advertising Optional Router Capabilities*
RFC 8362, *OSPFv3 Link State Advertisement (LSA) Extensibility*
RFC 8920, *OSPF Application-Specific Link Attributes*

3.25 OpenFlow

TS-007 Version 1.3.1, *OpenFlow Switch Specification* – OpenFlow-hybrid switches

3.26 Path Computation Element Protocol (PCEP)

draft-ietf-pce-binding-label-sid-15, *Carrying Binding Label/Segment Identifier (SID) in PCE-based Networks*. – MPLS binding SIDs

draft-alvarez-pce-path-profiles-04, *PCE Path Profiles*

draft-dhs-spring-pce-sr-p2mp-policy-00, *PCEP extensions for p2mp sr policy*

RFC 5440, *Path Computation Element (PCE) Communication Protocol (PCEP)*
RFC 8231, *Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE*
RFC 8281, *PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model*
RFC 8408, *Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages*
RFC 8664, *Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing*

3.27 Point-to-Point Protocol (PPP)

RFC 1332, *The PPP Internet Protocol Control Protocol (IPCP)*
RFC 1377, *The PPP OSI Network Layer Control Protocol (OSINLCP)*
RFC 1661, *The Point-to-Point Protocol (PPP)*
RFC 1662, *PPP in HDLC-like Framing*
RFC 1877, *PPP Internet Protocol Control Protocol Extensions for Name Server Addresses*
RFC 1989, *PPP Link Quality Monitoring*
RFC 1990, *The PPP Multilink Protocol (MP)*
RFC 1994, *PPP Challenge Handshake Authentication Protocol (CHAP)*
RFC 2153, *PPP Vendor Extensions*
RFC 2516, *A Method for Transmitting PPP Over Ethernet (PPPoE)*
RFC 2615, *PPP over SONET/SDH*
RFC 2686, *The Multi-Class Extension to Multi-Link PPP*
RFC 2878, *PPP Bridging Control Protocol (BCP)*
RFC 4638, *Accommodating a Maximum Transit Unit/Maximum Receive Unit (MTU/MRU) Greater Than 1492 in the Point-to-Point Protocol over Ethernet (PPPoE)*
RFC 5072, *IP Version 6 over PPP*

3.28 Policy management and credit control

3GPP TS 29.212 Release 11, *Policy and Charging Control (PCC); Reference points – Gx support as it applies to wireline environment (BNG)*
RFC 4006, *Diameter Credit-Control Application*
RFC 6733, *Diameter Base Protocol*

3.29 Pseudowire (PW)

draft-ietf-l2vpn-vpws-iw-oam-04, *OAM Procedures for VPWS Interworking*
MFA Forum 12.0.0, *Multiservice Interworking - Ethernet over MPLS*
MFA Forum 13.0.0, *Fault Management for Multiservice Interworking v1.0*

MFA Forum 16.0.0, *Multiservice Interworking - IP over MPLS*
RFC 3916, *Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)*
RFC 3985, *Pseudo Wire Emulation Edge-to-Edge (PWE3)*
RFC 4385, *Pseudo Wire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN*
RFC 4446, *IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)*
RFC 4447, *Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)*
RFC 4448, *Encapsulation Methods for Transport of Ethernet over MPLS Networks*
RFC 5085, *Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires*
RFC 5659, *An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge*
RFC 5885, *Bidirectional Forwarding Detection (BFD) for the Pseudowire Virtual Circuit Connectivity Verification (VCCV)*
RFC 6073, *Segmented Pseudowire*
RFC 6310, *Pseudowire (PW) Operations, Administration, and Maintenance (OAM) Message Mapping*
RFC 6391, *Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network*
RFC 6575, *Address Resolution Protocol (ARP) Mediation for IP Interworking of Layer 2 VPNs*
RFC 6718, *Pseudowire Redundancy*
RFC 6829, *Label Switched Path (LSP) Ping for Pseudowire Forwarding Equivalence Classes (FECs) Advertised over IPv6*
RFC 6870, *Pseudowire Preferential Forwarding Status bit*
RFC 7023, *MPLS and Ethernet Operations, Administration, and Maintenance (OAM) Interworking*
RFC 7267, *Dynamic Placement of Multi-Segment Pseudowires*
RFC 7392, *Explicit Path Routing for Dynamic Multi-Segment Pseudowires – ER-TLV and ER-HOP IPv4 Prefix*
RFC 8395, *Extensions to BGP-Signaled Pseudowires to Support Flow-Aware Transport Labels*

3.30 Quality of Service (QoS)

RFC 2430, *A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)*
RFC 2474, *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*
RFC 2597, *Assured Forwarding PHB Group*
RFC 3140, *Per Hop Behavior Identification Codes*
RFC 3246, *An Expedited Forwarding PHB (Per-Hop Behavior)*

3.31 Remote Authentication Dial In User Service (RADIUS)

draft-oscca-cfrg-sm3-02, *The SM3 Cryptographic Hash Function*
RFC 2865, *Remote Authentication Dial In User Service (RADIUS)*

RFC 2866, *RADIUS Accounting*
RFC 2867, *RADIUS Accounting Modifications for Tunnel Protocol Support*
RFC 2868, *RADIUS Attributes for Tunnel Protocol Support*
RFC 2869, *RADIUS Extensions*
RFC 3162, *RADIUS and IPv6*
RFC 4818, *RADIUS Delegated-IPv6-Prefix Attribute*
RFC 5176, *Dynamic Authorization Extensions to RADIUS*
RFC 6613, *RADIUS over TCP – with TLS*
RFC 6614, *Transport Layer Security (TLS) Encryption for RADIUS*
RFC 6929, *Remote Authentication Dial-In User Service (RADIUS) Protocol Extensions*
RFC 6911, *RADIUS attributes for IPv6 Access Networks*

3.32 Resource Reservation Protocol - Traffic Engineering (RSVP-TE)

draft-newton-mpls-te-dynamic-overbooking-00, A Diffserv-TE Implementation Model to dynamically change booking factors during failure events
RFC 2702, *Requirements for Traffic Engineering over MPLS*
RFC 2747, *RSVP Cryptographic Authentication*
RFC 2961, *RSVP Refresh Overhead Reduction Extensions*
RFC 3097, *RSVP Cryptographic Authentication -- Updated Message Type Value*
RFC 3209, *RSVP-TE: Extensions to RSVP for LSP Tunnels*
RFC 3473, *Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions – IF_ID RSVP_HOP object with unnumbered interfaces and RSVP-TE graceful restart helper procedures*
RFC 3477, *Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)*
RFC 3564, *Requirements for Support of Differentiated Services-aware MPLS Traffic Engineering*
RFC 3906, *Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels*
RFC 4090, *Fast Reroute Extensions to RSVP-TE for LSP Tunnels*
RFC 4124, *Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering*
RFC 4125, *Maximum Allocation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*
RFC 4127, *Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*
RFC 4561, *Definition of a Record Route Object (RRO) Node-Id Sub-Object*
RFC 4875, *Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)*
RFC 4950, *ICMP Extensions for Multiprotocol Label Switching*
RFC 5151, *Inter-Domain MPLS and GMPLS Traffic Engineering -- Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions*

RFC 5712, *MPLS Traffic Engineering Soft Preemption*

RFC 5817, *Graceful Shutdown in MPLS and Generalized MPLS Traffic Engineering Networks*

3.33 Routing Information Protocol (RIP)

RFC 1058, *Routing Information Protocol*

RFC 2080, *RIPng for IPv6*

RFC 2082, *RIP-2 MD5 Authentication*

RFC 2453, *RIP Version 2*

3.34 Segment Routing (SR)

draft-bashandy-rtgwg-segment-routing-uloop-06, *Loop avoidance using Segment Routing*

draft-filsfils-spring-net-pgm-extension-srv6-usid-13, *Network Programming extension: SRv6 uSID instruction*

draft-filsfils-spring-srv6-net-pgm-insertion-04, *SRv6 NET-PGM extension: Insertion*

draft-ietf-6man-spring-srv6-oam-10, *Operations, Administration, and Maintenance (OAM) in Segment Routing Networks with IPv6 Data plane (SRv6)*

draft-ietf-idr-bgp-ls-segment-routing-ext-16, *BGP Link-State extensions for Segment Routing*

draft-ietf-idr-segment-routing-te-policy-11, *Advertising Segment Routing Policies in BGP*

draft-ietf-isis-mpls-elc-10, *Signaling Entropy Label Capability and Entropy Readable Label Depth Using IS-IS – advertising ELC*

draft-ietf-lsr-flex-algo-16, *IGP Flexible Algorithm*

draft-ietf-lsr-isis-srv6-extensions-14, *IS-IS Extension to Support Segment Routing over IPv6 Dataplane*

draft-ietf-ospf-mpls-elc-12, *Signaling Entropy Label Capability and Entropy Readable Label-stack Depth Using OSPF – advertising ELC*

draft-ietf-rtgwg-segment-routing-ti-lfa-01, *Topology Independent Fast Reroute using Segment Routing*

draft-ietf-spring-conflict-resolution-05, *Segment Routing MPLS Conflict Resolution*

draft-ietf-spring-segment-routing-policy-08, *Segment Routing Policy Architecture*

draft-ietf-teas-sr-rsvp-coexistence-rec-02, *Recommendations for RSVP-TE and Segment Routing LSP co-existence*

draft-voyer-6man-extension-header-insertion-10, *Deployments With Insertion of IPv6 Segment Routing Headers*

draft-voyer-pim-sr-p2mp-policy-02, *Segment Routing Point-to-Multipoint Policy*

draft-voyer-spring-sr-p2mp-policy-03, *SR Replication Policy for P2MP Service Delivery*

RFC 8287, *Label Switched Path (LSP) Ping/Traceroute for Segment Routing (SR) IGP-Prefix and IGP-Adjacency Segment Identifiers (SIDs) with MPLS Data Planes*

RFC 8476, *Signaling Maximum SID Depth (MSD) Using OSPF – node MSD*

RFC 8491, *Signaling Maximum SID Depth (MSD) Using IS-IS – node MSD*
RFC 8660, *Segment Routing with the MPLS Data Plane*
RFC 8661, *Segment Routing MPLS Interworking with LDP*
RFC 8663, *MPLS Segment Routing over IP – BGP SR with SR-MPLS-over-UDP/IP*
RFC 8665, *OSPF Extensions for Segment Routing*
RFC 8666, *OSPFv3 Extensions for Segment Routing*
RFC 8667, *IS-IS Extensions for Segment Routing*
RFC 8669, *Segment Routing Prefix Segment Identifier Extensions for BGP*
RFC 8754, *IPv6 Segment Routing Header (SRH)*
RFC 8814, *Signaling Maximum SID Depth (MSD) Using the Border Gateway Protocol - Link State*
RFC 8986, *Segment Routing over IPv6 (SRv6) Network Programming*
RFC 9252, *BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)*

3.35 Simple Network Management Protocol (SNMP)

draft-blumenthal-aes-usm-04, *The AES Cipher Algorithm in the SNMP's User-based Security Model – CFB128-AES-192 and CFB128-AES-256*
draft-ietf-isis-wg-mib-06, *Management Information Base for Intermediate System to Intermediate System (IS-IS)*
draft-ietf-mboned-msdp-mib-01, *Multicast Source Discovery protocol MIB*
draft-ietf-mpls-ldp-mib-07, *Definitions of Managed Objects for the Multiprotocol Label Switching, Label Distribution Protocol (LDP)*
draft-ietf-mpls-lsr-mib-06, *Multiprotocol Label Switching (MPLS) Label Switching Router (LSR) Management Information Base Using SMIv2*
draft-ietf-mpls-te-mib-04, *Multiprotocol Label Switching (MPLS) Traffic Engineering Management Information Base*
draft-ietf-ospf-mib-update-08, *OSPF Version 2 Management Information Base*
draft-ietf-rrp-unified-mib-06, *Definitions of Managed Objects for the VRRP over IPv4 and IPv6 – IPv6*
ESO-CONSORTIUM-MIB revision 200406230000Z, *esoConsortiumMIB*
IANA-ADDRESS-FAMILY-NUMBERS-MIB revision 200203140000Z, *ianaAddressFamilyNumbers*
IANAifType-MIB revision 200505270000Z, *ianaifType*
IANA-RTPROTO-MIB revision 200009260000Z, *ianaRtProtoMIB*
IEEE8021-CFM-MIB revision 200706100000Z, *ieee8021CfmMib*
IEEE8021-PAE-MIB revision 200101160000Z, *ieee8021paeMIB*
IEEE8023-LAG-MIB revision 200006270000Z, *lagMIB*
LLDP-MIB revision 200505060000Z, *lldpMIB*
RFC 1157, *A Simple Network Management Protocol (SNMP)*
RFC 1212, *Concise MIB Definitions*

RFC 1215, *A Convention for Defining Traps for use with the SNMP*
RFC 1724, *RIP Version 2 MIB Extension*
RFC 1901, *Introduction to Community-based SNMPv2*
RFC 2021, *Remote Network Monitoring Management Information Base Version 2 using SMIv2*
RFC 2206, *RSVP Management Information Base using SMIv2*
RFC 2213, *Integrated Services Management Information Base using SMIv2*
RFC 2494, *Definitions of Managed Objects for the DS0 and DS0 Bundle Interface Type*
RFC 2578, *Structure of Management Information Version 2 (SMIv2)*
RFC 2579, *Textual Conventions for SMIv2*
RFC 2580, *Conformance Statements for SMIv2*
RFC 2787, *Definitions of Managed Objects for the Virtual Router Redundancy Protocol*
RFC 2819, *Remote Network Monitoring Management Information Base*
RFC 2856, *Textual Conventions for Additional High Capacity Data Types*
RFC 2863, *The Interfaces Group MIB*
RFC 2864, *The Inverted Stack Table Extension to the Interfaces Group MIB*
RFC 2933, *Internet Group Management Protocol MIB*
RFC 3014, *Notification Log MIB*
RFC 3165, *Definitions of Managed Objects for the Delegation of Management Scripts*
RFC 3231, *Definitions of Managed Objects for Scheduling Management Operations*
RFC 3273, *Remote Network Monitoring Management Information Base for High Capacity Networks*
RFC 3410, *Introduction and Applicability Statements for Internet Standard Management Framework*
RFC 3411, *An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks*
RFC 3412, *Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)*
RFC 3413, *Simple Network Management Protocol (SNMP) Applications*
RFC 3414, *User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)*
RFC 3415, *View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)*
RFC 3416, *Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP)*
RFC 3417, *Transport Mappings for the Simple Network Management Protocol (SNMP) – SNMP over UDP over IPv4*
RFC 3418, *Management Information Base (MIB) for the Simple Network Management Protocol (SNMP)*
RFC 3419, *Textual Conventions for Transport Addresses*
RFC 3498, *Definitions of Managed Objects for Synchronous Optical Network (SONET) Linear Automatic Protection Switching (APS) Architectures*
RFC 3584, *Coexistence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework*

RFC 3592, *Definitions of Managed Objects for the Synchronous Optical Network/Synchronous Digital Hierarchy (SONET/SDH) Interface Type*

RFC 3593, *Textual Conventions for MIB Modules Using Performance History Based on 15 Minute Intervals*

RFC 3635, *Definitions of Managed Objects for the Ethernet-like Interface Types*

RFC 3637, *Definitions of Managed Objects for the Ethernet WAN Interface Sublayer*

RFC 3826, *The Advanced Encryption Standard (AES) Cipher Algorithm in the SNMP User-based Security Model*

RFC 3877, *Alarm Management Information Base (MIB)*

RFC 3895, *Definitions of Managed Objects for the DS1, E1, DS2, and E2 Interface Types*

RFC 3896, *Definitions of Managed Objects for the DS3/E3 Interface Type*

RFC 4001, *Textual Conventions for Internet Network Addresses*

RFC 4022, *Management Information Base for the Transmission Control Protocol (TCP)*

RFC 4113, *Management Information Base for the User Datagram Protocol (UDP)*

RFC 4220, *Traffic Engineering Link Management Information Base*

RFC 4273, *Definitions of Managed Objects for BGP-4*

RFC 4292, *IP Forwarding Table MIB*

RFC 4293, *Management Information Base for the Internet Protocol (IP)*

RFC 4631, *Link Management Protocol (LMP) Management Information Base (MIB)*

RFC 4878, *Definitions and Managed Objects for Operations, Administration, and Maintenance (OAM) Functions on Ethernet-Like Interfaces*

RFC 7420, *Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module*

RFC 7630, *HMAC-SHA-2 Authentication Protocols in the User-based Security Model (USM) for SNMPv3*

SFLOW-MIB revision 200309240000Z, *sFlowMIB*

3.36 Timing

GR-1244-CORE Issue 3, *Clocks for the Synchronized Network: Common Generic Criteria*

GR-253-CORE Issue 3, *SONET Transport Systems: Common Generic Criteria*

IEEE 1588-2008, *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*

ITU-T G.781, *Synchronization layer functions*

ITU-T G.813, *Timing characteristics of SDH equipment slave clocks (SEC)*

ITU-T G.8261, *Timing and synchronization aspects in packet networks*

ITU-T G.8262, *Timing characteristics of synchronous Ethernet equipment slave clock (EEC)*

ITU-T G.8262.1, *Timing characteristics of an enhanced synchronous Ethernet equipment slave clock (eEEEC)*

ITU-T G.8264, *Distribution of timing information through packet networks*

ITU-T G.8265.1, *Precision time protocol telecom profile for frequency synchronization*

ITU-T G.8275.1, *Precision time protocol telecom profile for phase/time synchronization with full timing support from the network*

RFC 3339, *Date and Time on the Internet: Timestamps*

RFC 5905, *Network Time Protocol Version 4: Protocol and Algorithms Specification*

3.37 Two-Way Active Measurement Protocol (TWAMP)

RFC 5357, *A Two-Way Active Measurement Protocol (TWAMP) – server, unauthenticated mode*

RFC 5938, *Individual Session Control Feature for the Two-Way Active Measurement Protocol (TWAMP)*

RFC 6038, *Two-Way Active Measurement Protocol (TWAMP) Reflect Octets and Symmetrical Size Features*

RFC 8545, *Well-Known Port Assignments for the One-Way Active Measurement Protocol (OWAMP) and the Two-Way Active Measurement Protocol (TWAMP) – TWAMP*

RFC 8762, *Simple Two-Way Active Measurement Protocol – unauthenticated*

RFC 8972, *Simple Two-Way Active Measurement Protocol Optional Extensions – unauthenticated*

3.38 Virtual Private LAN Service (VPLS)

RFC 4761, *Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling*

RFC 4762, *Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling*

RFC 5501, *Requirements for Multicast Support in Virtual Private LAN Services*

RFC 6074, *Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs)*

RFC 7041, *Extensions to the Virtual Private LAN Service (VPLS) Provider Edge (PE) Model for Provider Backbone Bridging*

RFC 7117, *Multicast in Virtual Private LAN Service (VPLS)*

3.39 Voice and video

DVB BlueBook A86, *Transport of MPEG-2 TS Based DVB Services over IP Based Networks*

ETSI TS 101 329-5 Annex E, *QoS Measurement for VoIP - Method for determining an Equipment Impairment Factor using Passive Monitoring*

ITU-T G.1020 Appendix I, *Performance Parameter Definitions for Quality of Speech and other Voiceband Applications Utilizing IP Networks - Mean Absolute Packet Delay Variation & Markov Models*

ITU-T G.107, *The E Model - A computational model for use in planning*

ITU-T P.564, *Conformance testing for voice over IP transmission quality assessment models*

RFC 3550, *RTP: A Transport Protocol for Real-Time Applications – Appendix A.8*

RFC 4585, *Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)*

RFC 4588, *RTP Retransmission Payload Format*

3.40 Wireless Local Area Network (WLAN) gateway

3GPP TS 23.402, *Architecture enhancements for non-3GPP accesses – S2a roaming based on GPRS*

3.41 Yet Another Next Generation (YANG)

RFC 6991, *Common YANG Data Types*

RFC 7950, *The YANG 1.1 Data Modeling Language*

RFC 7951, *JSON Encoding of Data Modeled with YANG*

3.42 Yet Another Next Generation (YANG) OpenConfig Modules

openconfig-aaa.yang version 0.4.0, *OpenConfig AAA Module*

openconfig-aaa-radius.yang version 0.3.0, *OpenConfig AAA RADIUS Module*

openconfig-aaa-tacacs.yang version 0.3.0, *OpenConfig AAA TACACS+ Module*

openconfig-acl.yang version 1.0.0, *OpenConfig ACL Module*

openconfig-bfd.yang version 0.1.0, *OpenConfig BFD Module*

openconfig-bgp.yang version 3.0.1, *OpenConfig BGP Module*

openconfig-bgp-common.yang version 3.0.1, *OpenConfig BGP Common Module*

openconfig-bgp-common-multiprotocol.yang version 3.0.1, *OpenConfig BGP Common Multiprotocol Module*

openconfig-bgp-common-structure.yang version 3.0.1, *OpenConfig BGP Common Structure Module*

openconfig-bgp-global.yang version 3.0.1, *OpenConfig BGP Global Module*

openconfig-bgp-neighbor.yang version 3.0.1, *OpenConfig BGP Neighbor Module*

openconfig-bgp-peer-group.yang version 3.0.1, *OpenConfig BGP Peer Group Module*

openconfig-bgp-policy.yang version 4.0.1, *OpenConfig BGP Policy Module*

openconfig-if-aggregate.yang version 2.0.0, *OpenConfig Interfaces Aggregated Module*

openconfig-if-ethernet.yang version 2.0.0, *OpenConfig Interfaces Ethernet Module*

openconfig-if-ip.yang version 2.0.0, *OpenConfig Interfaces IP Module*

openconfig-if-ip-ext.yang version 2.0.0, *OpenConfig Interfaces IP Extensions Module*

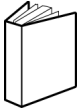
openconfig-igmp.yang version 0.2.0, *OpenConfig IGMP Module*

openconfig-interfaces.yang version 2.0.0, *OpenConfig Interfaces Module*

openconfig-isis.yang version 0.3.2, *OpenConfig IS-IS Module*

openconfig-isis-policy.yang version 0.3.2, *OpenConfig IS-IS Policy Module*
openconfig-isis-routing.yang version 0.3.2, *OpenConfig IS-IS Routing Module*
openconfig-lacp.yang version 1.1.0, *OpenConfig LACP Module*
openconfig-lldp.yang version 0.1.0, *OpenConfig LLDP Module*
openconfig-local-routing.yang version 1.0.1, *OpenConfig Local Routing Module*
openconfig-mpls.yang version 2.3.0, *OpenConfig MPLS Module*
openconfig-mpls-ldp.yang version 3.0.2, *OpenConfig MPLS LDP Module*
openconfig-mpls-rsvp.yang version 2.3.0, *OpenConfig MPLS RSVP Module*
openconfig-mpls-te.yang version 2.3.0, *OpenConfig MPLS TE Module*
openconfig-network-instance.yang version 1.1.0, *OpenConfig Network Instance Module*
openconfig-network-instance-l3.yang version 0.11.1, *OpenConfig L3 Network Instance Module – static routes*
openconfig-packet-match.yang version 1.0.0, *OpenConfig Packet Match Module*
openconfig-pim.yang version 0.2.0, *OpenConfig PIM Module*
openconfig-platform.yang version 0.12.2, *OpenConfig Platform Module*
openconfig-platform-fan.yang version 0.1.1, *OpenConfig Platform Fan Module*
openconfig-platform-linecard.yang version 0.1.2, *OpenConfig Platform Linecard Module*
openconfig-platform-port.yang version 0.3.3, *OpenConfig Port Module*
openconfig-platform-transceiver.yang version 0.7.1, *OpenConfig Transceiver Module*
openconfig-procmon.yang version 0.4.0, *OpenConfig Process Monitoring Module*
openconfig-relay-agent.yang version 0.1.0, *OpenConfig Relay Agent Module*
openconfig-routing-policy.yang version 3.0.0, *OpenConfig Routing Policy Module*
openconfig-rsvp-sr-ext.yang version 0.1.0, *OpenConfig RSVP-TE and SR Extensions Module*
openconfig-system.yang version 0.9.1, *OpenConfig System Module*
openconfig-system-logging.yang version 0.3.1, *OpenConfig System Logging Module*
openconfig-system-terminal.yang version 0.3.0, *OpenConfig System Terminal Module*
openconfig-telemetry.yang version 0.5.0, *OpenConfig Telemetry Module*
openconfig-terminal-device.yang version 1.7.3, *OpenConfig Terminal Optics Device Module*
openconfig-vlan.yang version 2.0.0, *OpenConfig VLAN Module*

Customer document and product support



Customer documentation

[Customer documentation welcome page](#)



Technical support

[Product support portal](#)



Documentation feedback

[Customer documentation feedback](#)