

Technology Overview

Frequently Asked Questions: MX Series 3D Universal Edge Routers Quality of Service

Release
11.2



Published: 2012-01-27

Juniper Networks, Inc.
1194 North Mathilda Avenue
Sunnyvale, California 94089
USA
408-745-2000
www.juniper.net

This product includes the Envoy SNMP Engine, developed by Epilogue Technology, an Integrated Systems Company. Copyright © 1986-1997, Epilogue Technology Corporation. All rights reserved. This program and its documentation were developed at private expense, and no part of them is in the public domain.

This product includes memory allocation software developed by Mark Moraes, copyright © 1988, 1989, 1993, University of Toronto.

This product includes FreeBSD software developed by the University of California, Berkeley, and its contributors. All of the documentation and software included in the 4.4BSD and 4.4BSD-Lite Releases is copyrighted by the Regents of the University of California. Copyright © 1979, 1980, 1983, 1986, 1988, 1989, 1991, 1992, 1993, 1994. The Regents of the University of California. All rights reserved.

GateD software copyright © 1995, the Regents of the University. All rights reserved. Gate Daemon was originated and developed through release 3.0 by Cornell University and its collaborators. Gated is based on Kirton's EGP, UC Berkeley's routing daemon (routed), and DCN's HELLO routing protocol. Development of Gated has been supported in part by the National Science Foundation. Portions of the GateD software copyright © 1988, Regents of the University of California. All rights reserved. Portions of the GateD software copyright © 1991, D. L. S. Associates.

This product includes software developed by Maker Communications, Inc., copyright © 1996, 1997, Maker Communications, Inc.

Juniper Networks, Junos, Steel-Belted Radius, NetScreen, and ScreenOS are registered trademarks of Juniper Networks, Inc. in the United States and other countries. The Juniper Networks Logo, the Junos logo, and JunosE are trademarks of Juniper Networks, Inc. All other trademarks, service marks, registered trademarks, or registered service marks are the property of their respective owners.

Juniper Networks assumes no responsibility for any inaccuracies in this document. Juniper Networks reserves the right to change, modify, transfer, or otherwise revise this publication without notice.

Products made or sold by Juniper Networks or components thereof might be covered by one or more of the following patents that are owned by or licensed to Juniper Networks: U.S. Patent Nos. 5,473,599, 5,905,725, 5,909,440, 6,192,051, 6,333,650, 6,359,479, 6,406,312, 6,429,706, 6,459,579, 6,493,347, 6,538,518, 6,538,899, 6,552,918, 6,567,902, 6,578,186, and 6,590,785.

Technology Overview Frequently Asked Questions: MX Series 3D Universal Edge Routers Quality of Service

Release 11.2

Copyright © 2012, Juniper Networks, Inc.

All rights reserved.

Revision History

April 2011—R1 Junos OS 11.2

The information in this document is current as of the date listed in the revision history.

YEAR 2000 NOTICE

Juniper Networks hardware and software products are Year 2000 compliant. Junos OS has no known time-related limitations through the year 2038. However, the NTP application is known to have some difficulty in the year 2036.

END USER LICENSE AGREEMENT

The Juniper Networks product that is the subject of this technical documentation consists of (or is intended for use with) Juniper Networks software. Use of such software is subject to the terms and conditions of the End User License Agreement ("EULA") posted at <http://www.juniper.net/support/eula.html>. By downloading, installing or using such software, you agree to the terms and conditions of that EULA.

Table of Contents

MX Series QoS FAQ Overview	1
Dense Port Concentrators (DPCE-X, DPCE-R, DPCE-Q) FAQs	2
Junos Trio Chipset on MX Series FAQs	7

MX Series QoS FAQ Overview

This document presents frequently asked questions about the similarities and differences in implementing quality of service features using Modular Port Concentrators (MPCs), Dense Port Concentrators (DPCs), and the Junos Trio chipset on Juniper Networks MX Series 3D Universal Edge Routers.

The increased demand for sophisticated, media-rich services, the exponential growth of mobile sessions, and the emerging trend of cloud computing require a networking infrastructure that supports massive numbers of subscribers, service types and instances, and bandwidth. A number of features and methods have been developed to address these advanced network requirements, including quality of service (QoS). QoS is a set of mechanisms that helps maintain specified service levels for your network by optimizing and prioritizing network traffic so that demand for resources can meet requirements. Use QoS mechanisms to control the allocation of network attributes such as available bandwidth, latency, jitter, packet drop, and bit rate errors so that resources are managed to levels acceptable to your network customers and applications.

QoS on Juniper Networks MX Series 3D Universal Edge Routers

MX Series routers are available in a variety of configurations with robust features, including options that provide the level and granularity of the QoS support needed in your network. The MX Series hardware options currently include five models of Modular Port Concentrators (MPCs), using several different 3D Modular Interface Cards (MICs), and three models of Dense Port Concentrators (DPCs). The MPCs and DPCs provide varying degrees of QoS support.

The MPCs are next-generation line modules for advanced Ethernet services edge and broadband edge networks using high capacity, modular 1-Gigabit Ethernet and 10-Gigabit Ethernet hardware. The MPCs house Packet Forwarding Engines that deliver up to 120 Gbps of comprehensive Layer 3 routing (IPv4 and IPv6), Layer 2 switching, inline services, and advanced hierarchical quality of service (H-QoS) per MX Series slot. The MPCs can also take advantage of the high performance Junos Trio chipset.

Key QoS features provided by the MPCs include extensive queue management, scheduler hierarchy, shaping, intelligent oversubscription, weighted round robin (WRR), random early detection (RED), and weighted random early detection (WRED).

The DPCs (DPCE-X, DPCE-R, and DPCE-Q) each provide multiple physical interfaces and Packet Forwarding Engines on a single board that performs packet processing and forwarding. Each Packet Forwarding Engine consists of one I-chip for Layer 3 processing and one network processor for Layer 2. DPCE-Qs offer enhanced queuing capabilities and the QoS features of WRR, RED, and WRED.

Related Documentation

For more information about specific hardware capabilities, see:

- Juniper Networks Datasheet MX Series 3D Universal Edge Routers
- Juniper Networks Datasheet Modular Port Concentrators for the MX Series
- Juniper Networks Datasheet Dense Port Concentrators

Dense Port Concentrators (DPCE-X, DPCE-R, DPCE-Q) FAQs

This section presents frequently asked questions and answers related to Dense Port Concentrators (DPCE-X, DPCE-R, DPCE-Q) on Juniper Networks MX Series routers.

How many Layer 3 and Layer 2 policers are supported on Juniper Networks MX Series devices? What are the limiting factors for the policers?

The Layer 3 interface policer limit for I-chip-based DPCs is 39,000 possible, with 16,000 tested. In most implementations, logical interface limits are reached per port and per Packet Forwarding Engine before policer limits are reached.

The MX Series DPCs support Layer 2 policers indirectly. The Layer 2 policers are supported in the IQ2 Ethernet Services Engine, and the I-chip supports both Layer 3 and Layer 2 policing. The type of policing applied by the I-chip (Layer 3 or Layer 2) depends on the type of packet. For example, if a Layer 2 packet is received on a device using an I-chip, Layer 2 policing is performed by the I-chip.

Are peak information rate (PIR) and committed information rate (CIR) supported at the queue level on MX Series devices?

Support for PIR and CIR on MPCs

On MX Series devices using MPCs, CIR and PIR are supported at the queue level with class-of-service (CoS) schedulers. Use three rates, **transmit-rate**, **shaping-rate**, and **excess-rate**, which can be configured simultaneously, and use the **excess-priority** statement with **excess-rate** to manage the bandwidth in the excess region between CIR and PIR, as follows:

- **excess-rate**—Use to configure the percentage of excess bandwidth traffic that should go into the queue in the excess region.
- **shaping-rate**—Use to configure the maximum bandwidth usage, which sets the PIR.
- **transmit-rate**—Use to configure the minimum bandwidth allocated to a queue, which sets the CIR.
- **excess-priority**—Use to configure the priority of excess traffic in a scheduler as low, medium-low, medium-high, or high.

Support for PIR and CIR on DPCs

On devices using enhanced queuing DPCs, CIR and PIR are not supported at the queue level. PIR on these devices is supported by using the **rate-limit** statement. You can achieve an effect similar to CIR and PIR at the queue level by combining tricolor policers with **rate-limit** and drop profiles. You use the tricolor policers to enforce the CIR and PIR, by configuring the drop profiles to drop the yellow packets (packet loss priority (PLP): medium-high) before the green packets (PLP: low).

Is QoS on GRE tunnels supported when using the MS-DPC (IP services line card) on an MX Series device?

No, QoS on GRE tunnels is not supported when using the MS-DPC line card.

Is it possible to use a common QoS scheduler on a traffic-class group comprised of an aggregate of multiple GRE tunnels?

Support for using a common QoS scheduler on an aggregate group of GRE tunnels is accomplished by using a per-unit scheduler for GRE tunnels, which provides fine-grained queuing by using a single scheduler for a set of queues. As of Junos OS Release 10.1 and later, support for per-unit scheduler for GRE tunnels is added for:

- M Series Multiservice Edge Routers: M320 with SFPC, M120, M7i, and M10i for non-enhanced (ABC-based) Compact Forwarding Engine Board (CFEB) and enhanced CFEB (I-chip-based)
- T Series Core Routers, including TX Matrix devices

This feature adds all of the functionality of tunnel PICs to Gigabit Ethernet Intelligent Queuing 2 (IQ2) and Enhanced IQ2 (IQ2E) PICs. The QoS for the GRE tunnel traffic is applied as the traffic is looped through the IQ2 or IQ2E PIC. Shaping is performed on all packets that pass through the GRE tunnel.

Is it possible to create a strict-high priority queue running at the same time as several high priority queues?

Yes, this can be done on all of the DPCs, those with enhanced queuing and those without enhanced queuing. This cannot be done on devices using IQ2 PICs.

Are the QoS features handled the same on the 10-Gigabit Ethernet DPCs as on the 1-Gigabit Ethernet DPCs?

Yes, the QoS features are the same for both types of DPCs.

Is hierarchical QoS per VLAN supported on aggregated Ethernet (AE) interfaces?

Hierarchical QoS allows you to control QoS at multiple levels: the physical level, the logical level, and fine-grained control at the command-line level. It is useful for managing bandwidth congestion and link sharing in multi-service networks. For MX Series devices with AE interfaces, hierarchical QoS per VLAN is available over a link aggregation group (LAG) in Junos OS Release 9.4 and later. This support is only one-to-one active/backup. However, it can be supported across DPCs.

What are the differences between the CoS traffic-manager options on the MX Series?

Use **traffic-manager** mode with the following options to configure CoS traffic manager mode of operation:

- **egress-only**—Enables CoS queuing and scheduling on the egress side for the PIC that houses the interface. This is the default mode for an Enhanced Queuing (EQ) DPC on MX Series routers.

- **ingress-and-egress**—Enables CoS queuing and scheduling on both the egress and ingress sides for the PIC. This is the default mode for IQ2 and IQ2E PICs on M Series and T Series routers. For EQ DPCs, you must configure the **traffic-manager** statement with **ingress-and-egress** mode to enable ingress CoS on the EQ DPC.
 - When **ingress-and-egress** is turned on, the classification is done in the IQ2 or IQ2E PIC.
 - When **ingress-and-egress** is not turned on, the classification is done at the I-chip.

Is rate-limit at the physical interface level supported on Enhanced Queuing DPCs (DPCE-R-Qs)?

Yes. In Junos OS Release 10.0 and later, **rate-limit** is supported at the physical interface level on Enhanced Queuing DPCs.

How is the shaping rate calculated on the Enhanced Queuing DPCs (DPCE-R-Qs)?

The shaping rate calculation on the Enhanced Queuing DPC is similar to the shaping rate and WRR calculations performed on the Gigabit Ethernet IQ2 PICs:

For ingress and egress: Layer 3 header + Layer 2 header + frame check sequence (FCS)

Are ingress queuing and rate-limit supported with the Enhanced Queuing DPCs (DPCE-R-Qs)?

Ingress queuing and **rate-limit** are not supported with the Enhanced Queuing DPCs. However, as of Junos OS Release 10.2 and later, ingress policing on a Layer 2 Gigabit Ethernet logical interface is supported on the MX Series devices, enabling several two-color and three-color policer types on input and output interfaces. Also supported are Layer 2 policers at the VLAN tag level on MX Series platforms with Junos Trio MPC and MIC interfaces.

What is the queuing buffer size on MX Series DPCs?

- On port queuing DPCs, the delay buffer is 100 ms per port on egress. This delay calculation is based on a 64 byte average packet size.
- On Enhanced Queuing DPCs, the delay buffer is 500 ms per port on egress and ingress. This delay calculation is based on a 512 byte average packet size.

Are fine-grained queuing capabilities supported on Enhanced Queuing DPCs (DPCE-R-Q)?

Yes. DPCE-Rs and DPCE-Qs support up to 64,000 individual queues across both 1-Gigabit Ethernet and 10-Gigabit Ethernet ports. Additional features supported on these DPC types include:

- Four-level hierarchical WRR
- Four levels of per-VLAN queue priority
- Priority propagation

-
- Drop statistics per VLAN, color, or queue
 - Changeable allocation of schedulers per port for up to 8,000 scheduler nodes with eight queues each or 16,000 nodes with four queues each

What are the QoS properties of the DPCE-R and DPCE-Q line cards?

The following are the major QoS properties and their features:

- Queuing at the VLAN level, per Packet Forwarding Engine using I-chip
 - 4,000 schedulers with four queues
 - 2,000 schedulers with eight queues
- Hierarchical QoS
 - Traffic shaping at the physical port and at the customer VLAN or set of VLANs with the same service VLAN.
 - The **traffic-control-profiles** configuration statement is extended to support QoS at the **interface-set** level.
 - Support for priority propagation.
 - Shape traffic at the physical port and the customer VLAN and set of customer VLANs with the same service VLAN.
 - More customer VLAN schedulers than previous solutions: 2,000 eight-queue schedulers or 4,000 four-queue schedulers per 10 Gigabit Ethernet, versus 1,000 any-queue schedulers per 10 Gigabit Ethernet for IQ2.
 - Three levels of priority versus only two levels with IQ and IQ2.
 - Priority propagation—Priority from the queue level is preserved/demoted when passing logical interface or interface set stages.
 - Shaping and scheduling at inner/outer VLAN tag levels using a logical interface or interface set.
- Queues and forwarding classes:
 - Eight queues per port
 - 16 forwarding classes
 - Four scheduling priorities per queue
 - Four WRED profiles per queue with flexible RED profiles

What are the QoS properties of the DPCE-R line cards?

- Full Junos OS Layer 3 routing feature set
 - Eight queues per port
 - Layer 2 Ethernet switching features

- Per-VLAN policing
- Per-VLAN rewrite
- Per-VLAN tricolor marking
- Per-VLAN classification
- Per-VLAN accounting
- Per-VLAN filtering
- Classification per VLAN
 - 802.1p of inner or outer tag
 - MPLS EXP
 - IPv4 type of service (ToS) firewall filters
 - Both Layer 3 and Layer 2 fields for VPLS and bridge traffic
 - Hierarchical policers per VLAN, two-rate tricolor marking (TCM), single-rate TCM, single-rate policers
 - Packet header rewrite per-VLAN queuing and scheduler per port
- Queues and forwarding classes
 - Eight queues per port
 - 16 forwarding classes
 - Four scheduling priorities per queue
 - Four WRED profiles per queue with flexible RED profiles

What are the QoS properties of the Enhanced Queuing DPCs (DPCE-R-Q)?

- Support ingress queuing, scheduling, and shaping
- Classification using EXP for VPLS without tunnel
- ACL-based classification for ingress QoS Layer 2 policers: per-VLAN ingress policers and per-VLAN egress policers
- Match 802.1p and PLP in a firewall filter
- Rewrite inner packets 802.1p
- Rate limit per queue
- Includes DEI7 bit in 802.1p-based classification
- Double the number of subscribers, schedulers, shapers, and queues per DPC
- Multiple VLAN bundling (interface sets within interface sets)
- Class-aware hierarchical policers

Is DiffServ code point (DSCP) classification of MPLS-tagged packets supported on the I-chip-based DPCs?

The DSCP classifier is supported on I-chip DPCs as shown in the following table.

Table 1: DSCP Classifier Configuration

MPLS Configuration	Supported Platforms	DSCP Classifier Configuration
Layer 3 VPNs and VPLSs using an LSI routing instance	M320, M120, and MX Series	Configured under class-of-service routing-instances on the egress PE router.
Layer 3 VPNs using a virtual tunnel (VT) routing instance	M320, M120, and MX Series	Configured on the core-facing interface under class-of-service interfaces on the egress PE router.
MPLS forwarding	M320, M120, and MX Series (not supported on IQE and MX Series when ingress queuing is used)	Configured on the ingress core facing interface under class-of-service interfaces on the P or egress PE router.
VPLS using a VT routing instance	Not supported	
MPLS forwarding when number of labels in the MPLS label stack is more than two	Not supported	

Junos Trio Chipset on MX Series FAQs

Junos Trio is a chipset with revolutionary 3D scaling technology that enables networks to scale dynamically for more bandwidth, subscribers, and services. This section answers questions related to this chipset used with Junos OS Release 10.1 and later.

Which product platforms use the Junos Trio chipset?

- MX 3D MPC line cards

The MX 3D MPC provides the connection between the customer's Ethernet interfaces and the routing fabric of the MX Series chassis. The board features two Junos Trio chipsets combined with 40 SFP Ethernet interfaces.

- 16 port 10-Gigabit Ethernet MPC

This MPC provides the connection between 10-Gigabit Ethernet LAN interfaces and the routing fabric of the MX Series chassis. It has four identical packet processing paths combined with a control plane.

- MX80 (standalone chassis)

What is the total number of Packet Forwarding Engines in each of the DPCs using the Junos Trio chipset?

- The 12-16x10-Gigabit Ethernet MPC on MX Series has a total of four Packet Forwarding Engines per MPC.
- The 3D MPCs each have two Packet Forwarding Engines.

What is the default power-on sequence when line cards with Junos Trio chipsets are in the same chassis with other types of line cards?

If the **set chassis network-services** attribute is not configured, the following is the line card power-up rule:

- If the first powered up Flexible PIC Concentrator (FPC) is a DPC, then only DPCs within the box are allowed to power up.
- If the first powered up FPC is an MPC, then only MPCs are allowed to power up.

As of Junos OS Release 10.2 and later, the power-up rule has been changed to the following:

- If the **set chassis network-services** attribute is configured as **ip** at start time, any MX Series platform-supported boards (such as ADPC/SDPC/FPC/MPC/ and the like), except XDPC, will boot.
- If the **set chassis network-services** attribute is configured as **ethernet** at start time, any MX Series platform-supported boards (such as ADPC/SDPC/FPC/MPC/XDPC and the like) will boot.
- If the **set chassis network-services** attribute is configured as **enhanced-ip** or **enhanced-ethernet** at start time, only Junos Trio chipset boards and SDPC will boot.

What are the QoS differences between the 16 Port 10-Gigabit Ethernet MPC and the I-chip-based DPC?

- Dynamic memory is not supported on the 16x10-Gigabit Ethernet MPC.
- A buffer configured on a 3D MPC queue is treated as the maximum. However, it is treated as the minimum on the I-chip DPC.
- The Junos Trio chipset (3D MPC) maintains packets in 128-byte chunks.
- Port shaping is supported on all MPCs.
- Queues can have unique shaping and guaranteed rate configuration.

What is the difference in buffer management on the MPCs with the Junos Trio chipset compared with the I-chip-based DPC?

On port-queuing I-chip DPCs, 64 byte-per-unit dynamic buffers are available per queue. If a queue is using more than its allocated bandwidth share due to excess bandwidth left over from other queues, its buffers are dynamically increased. This is feasible because the I-chip DPCs primarily perform WRED drops at the head of the queue, as opposed to “tail-assisted” drops, which are performed only when a temporal buffer is configured or when the queue becomes full. When a temporal buffer is not configured, the allocated buffer is treated as the minimum for that queue, and can expand if other queues are not using their share.

On MPCs with the Junos Trio chipset, WRED drops are performed at the tail of the queue. The packet buffer is organized into 128-byte units. Before a packet is queued, buffer and WRED checks are performed, and the decision to drop is made at this time. Once a packet is queued, it is not dropped. As a result, dynamic buffer allocation is not supported on the Junos Trio chipset Packet Forwarding Engines. The buffer allocation per queue on the Junos Trio chipset Packet Forwarding Engines is considered the maximum for that queue. Once the allocated buffer becomes full, subsequent packets are dropped until space is available, even if other queues are idle. Buffering is only required during oversubscription.

To provide larger buffers on Junos Trio chipset Packet Forwarding Engines, the delay buffer can be increased from the default 100 ms to 200 ms of the port speed, and can also be oversubscribed using the **delay-buffer-rate** configuration per port.

What are the supported QoS features of 16-port 10-Gigabit Ethernet MPC?

The 16x10-Gigabit Ethernet MPC supports the following QoS functionality:

- Port-based queuing
 - Per-port shaping
 - Eight queues per port
 - 100 ms of delay buffer by default per port
 - 200 ms of delay buffer configurable per port
 - Ability to oversubscribe the delay buffer beyond 200 ms per port
- Queue-level shaping and guaranteed rate
 - Separate guaranteed and shaping rates
 - Rate limit option to police a queue to act as a Low Latency Queue (LLQ)
- Four WRED profiles per queue
- Multiple queue priority levels
- Strict High, High, Medium, and Low guaranteed priority levels
 - Strict High and High are at the same hardware priority level
 - Round robin at each guaranteed priority level
- High and Low excess priority levels
 - Queues perform WRR at the excess priority levels
 - Strict priority scheduling at each excess priority level
- Classification per VLAN
 - MPLS EXP
 - IPv6, IPv4 ToS

- Inner and outer tag 802.1p (and DEI7)
- MF classifiers per VLAN
- Policers per VLAN
 - Single rate, single-rate tricolor marking, two-rate tricolor marking, hierarchical policers
 - Class-aware intelligent hierarchical policers
- Rewrites per VLAN
 - MPLS EXP, IP DSCP/PREC
 - Inner and outer tag 802.1p (and DEI7)
 - Ingress DSCP rewrite