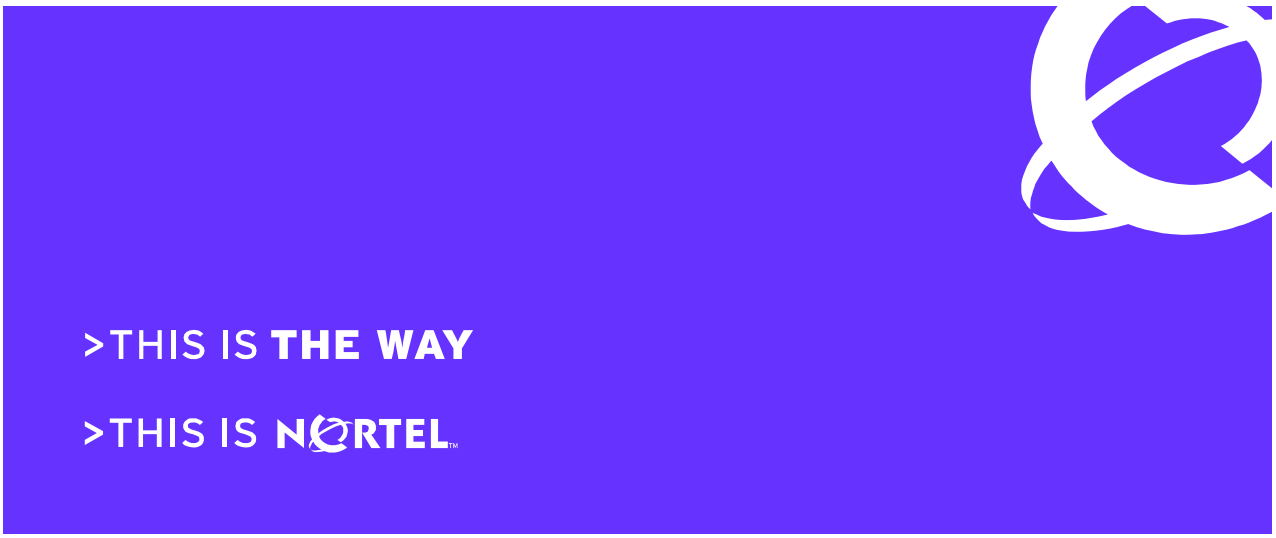


Nortel Multiservice Switch 7400/15000/20000

ATM Queuing and Scheduling Fundamentals

NN10600-707

Document status: Standard
Document issue: 7.1S1
Document date: October 2005
Product release: PCR7.1 and up
Job function: Product Fundamentals
Type: NTP
Language type: U.S. English

Copyright © 2005 Nortel.
All Rights Reserved.

NORTEL, the globemark design, and the NORTEL corporate logo are trademarks of Nortel.

Contents

What's new 7
4-port MR POS and ATM FP 7

Overview of queuing and traffic scheduling 8
Overview of Multiservice Switch multiple priority system 9
Emission priorities 10
Free lists 11
Service category to emission priority mapping 12
Per-VC and common queuing 12
General principles of Multiservice Switch traffic scheduling 14
Link transmit queues 15
Per-VC queue limits and thresholds: general 15
Queuing and traffic shaping 17
Impact of link transmit queue limit on QOS and link usage 17
Per-VC queue limit configuration 18
Discard priority: all function processors 21
Queue discard priorities 22
Free list discard priorities 23
Priority and resource interactions 24
Example of the discard priority process 26
Traffic mapping to internal discard and emission priority 27
Receive mapping - cell relay 28
Receive mapping - frame 28
Software overrides 28
Transmit mapping - cell relay 29
Transmit mapping - frame 30

Queuing and scheduling on CQC-based FPs 32
Overview to CQC queuing and scheduling 32
CQC emission priorities 33
Per-VC queues and common queues 34
Traffic scheduling for CQC-based function processors 35
Traffic scheduling and shaping stacks 36
Mixed queues with a single priority 37

Copyright © 2005, Nortel

Nortel Confidential

CQC discard priority 38
Interaction between emission and discard priorities 39
Service category mapping priorities for CQC 39
CQC port aggregation 40
Free list and queue configurations 42
Free list lengths and congestion thresholds 42
CQC queue limits and congestion thresholds 43
CQC queuing and scheduling for VP termination 49

Queuing and scheduling on ATM IP FPs 50
Overview of ATM IP queuing and scheduling 50
ATM IP emission priorities 51
PQC emission priorities 51
AQM emission priorities 52
Traffic scheduling on the AQM 53
Class scheduling 54
Minimum bandwidth guarantee 55
Minimum bandwidth guarantee compared to bandwidth pools 57
Typical configurations for class scheduling 58
Customized configuration for class scheduling 61
ATM IP connection scheduling 62
Weighted fair queuing for ATM IP FPs 62
Weighted fair queuing and common queuing 65
Per-VC and common queuing in non-port aggregation configurations 66
Discard priorities on ATM IP function processors 66
Interaction between emission and discard priorities 66
Service category mapping to priorities: ATM IP PQC 67
Service category mapping to priorities: ATM IP AQM 67
Priority interactions and minimum bandwidth guarantee 69
Priority interaction without MBG and no free list congestion 70
Priority interaction under free list congestion 70
Priority interaction with MBG and no free list congestion 70
Port aggregation on ATM IP function processors 71
Interaction between congestion control and packet-wise discard levels 72
Per-VC queuing for RT-VBR traffic 73
Traffic allocation to service categories 74
Queue configuration for NRT-VBR and UBR 74
NRT-VBR and UBR allocated to unshaped queues 74
Configuring port aggregation 75
Emulation of port aggregation congestion management 75
Congestion management through MBG 75
Congestion management using the free list state 75
ATM IP queue limits and discard thresholds 76
PQC queue limits and thresholds 76

Copyright © 2005, Nortel
Nortel Confidential

AQM queue limits and thresholds 77
Expanded default queue limits for NRT-VBR and UBR 77
ATM IP queuing and scheduling for basic VP termination 79
ATM IP queuing and scheduling for standard VP termination 79
Standard VPT VCC queuing and scheduling 80
Weighting characteristics for VPT VCCs 82
Queue limits for VPTs 83
VPT queue limit configuration 84

Queuing and scheduling on APC/PQC-based FPs 85
Buffer management 85
Buffer pool and threshold allocation 86
User-configurable buffer limits for EPs 86
Per-VC queuing on APC/PQC-based FPs 87
Per-VC queue limits 87
Per-VC queue thresholds 88
Overview of APC schedulers 89
APC class scheduling 90
Minimum bandwidth guarantee 92
Minimum bandwidth guarantee compared to bandwidth pools 92
Connection scheduling 92
Rate connection scheduler 92
Weighted fair queuing 93

Queuing and scheduling on GQM-based FPs 94
Buffer management on GQM-based FPs 94
Queuing on GQM-based FPs 95
Queue thresholds on GQM-based FPs 95
Common queuing on GQM-based FPs 95
Per-VC queuing on GQM-based FPs 96
Schedulers for GQM-based FPs 96
GQM link scheduling 97
GQM class scheduling 97
GQM connection scheduling 99
Configuring MBG values for GQM-based FPs 100

Memory management 103
Overview of memory management 103
Memory management for APC-based FPs 103
Memory management for ATM IP FPs 105
PQC buffer space on ATM IP FPs 106
AQM buffer space on ATM IP FPs 107
Memory management on ATM IP FPs 108
PQC CQM memory management 110
AQM CQM memory management 113

Copyright © 2005, Nortel
Nortel Confidential

Memory management for CQC-based FPs 116
Buffer space for CQC-based FPs 116
Memory management on CQC-based FPs 117

Packet-wise discard 125
Overview of packet-wise discard 125
Overview of late packet discard 127
Overview of partial packet discard 128
Overview of early packet discard 130
High and low priority EPD offset 130
Overview of weighted random early detection 131
Applications for packet-wise discard 131
Partial packet discard at the cell relay point 132
Partial packet discard for AAL5 connections 132
Intelligent discard at an AAL5 adaptation point 133
Packet-wise discard and VPTs 133
Congestion notification 134
Overview to congestion notification 134
EFCI header insertion 134
EFCI-FCI mapping 135
Packet-wise discard for CQC-based FPs 135
Characteristics of CQC packet-wise discard 135
LPD on CQC-based FPs 137
PPD at cell relay points: CQC variant 137
PPD for AAL5 connections over CQC-based FPs 137
EPD on CQC-based FPs 137
EFCI on CQC-based FPs 137
Packet-wise discard for ATM IP FPs 138
Characteristics of ATM IP packet-wise discard 138
LPD on ATM IP FPs 142
EPD on ATM IP FPs 142
PPD on ATM IP FPs 142
WRED on ATM IP FPs 143
AAL5 auto-detection 144
Configuring packet-wise discard for ATM IP 145
EFCI on ATM IP FPs 145
Packet-wise discard for APC-based FPs 146
Partial packet discard for APC-based FPs 146
Early packet discard for APC-based FPs 146
Weighted random early detection for APC-based FPs 146
Packet-wise discard for GQM-based FPs 147

Copyright © 2005, Nortel
Nortel Confidential

What's new

The following feature was added to this document:

- 4-port MR POS and ATM FP (page 7)

Attention: To ensure that you are using the most current version of an NTP, check the current NTP list in NN10600-000 *Nortel Multiservice Switch 7400/15000/20000 What's New*.

4-port MR POS and ATM FP

The following sections were updated for this feature:

- Configurable parameters for per-VC queue default limits (page 20)
- Queuing and scheduling on GQM-based FPs (page 94)

Copyright © 2005, Nortel

Nortel Confidential

Overview of queuing and traffic scheduling

Queues and discard priorities apply to the following functions in these function processors (FPs):

- CQC-based
- ATM IP
- GQM-based
- ATM traffic transmission
  - transmit link
  - Nortel Multiservice Switch device backplane
- queuing of traffic processed by the CPU

The service provider requires traffic scheduling policies to meet these requirements and to maintain efficient use of resources. When demand for resources exceeds capacity, discard policy determines which cells to discard so that the node can maintain the required level of service.

This chapter provides an overview to queuing and traffic scheduling for all Multiservice Switch ATM node function processors.

Navigation

- Overview of Multiservice Switch multiple priority system (page 9)
- Emission priorities (page 10)
- Per-VC and common queuing (page 12)
- General principles of Multiservice Switch traffic scheduling (page 14)
- Link transmit queues (page 15)
- Discard priority: all function processors (page 21)
- Traffic mapping to internal discard and emission priority (page 27)

Copyright © 2005, Nortel

Nortel Confidential

Overview of Multiservice Switch multiple priority system

Nortel Multiservice Switch nodes' multiple priority system is based on two sets of multiple priorities:

- multiple emission priorities (degree of traffic urgency)
- multiple discard priorities (degree of traffic importance)

Through a systematic application of this system of priorities, the service provider ensures that the nodes and the network maintain the required level of service for each connection. Emission and discard priorities apply to traffic in all service categories.

The figure General principles of emission and discard priorities on Multiservice Switch nodes (page 10) shows the relationship between the emission and discard priorities. Note how the priority indicators use 0 to indicate highest priority (most urgent or most important). For each emission priority, the node always reserves discard priority 0 for network control traffic. Depending on the traffic carried by the node, discard priority 1 may also be reserved. For example, if a node is carrying DPRS traffic, that node may also reserve discard priority 1 for control traffic.

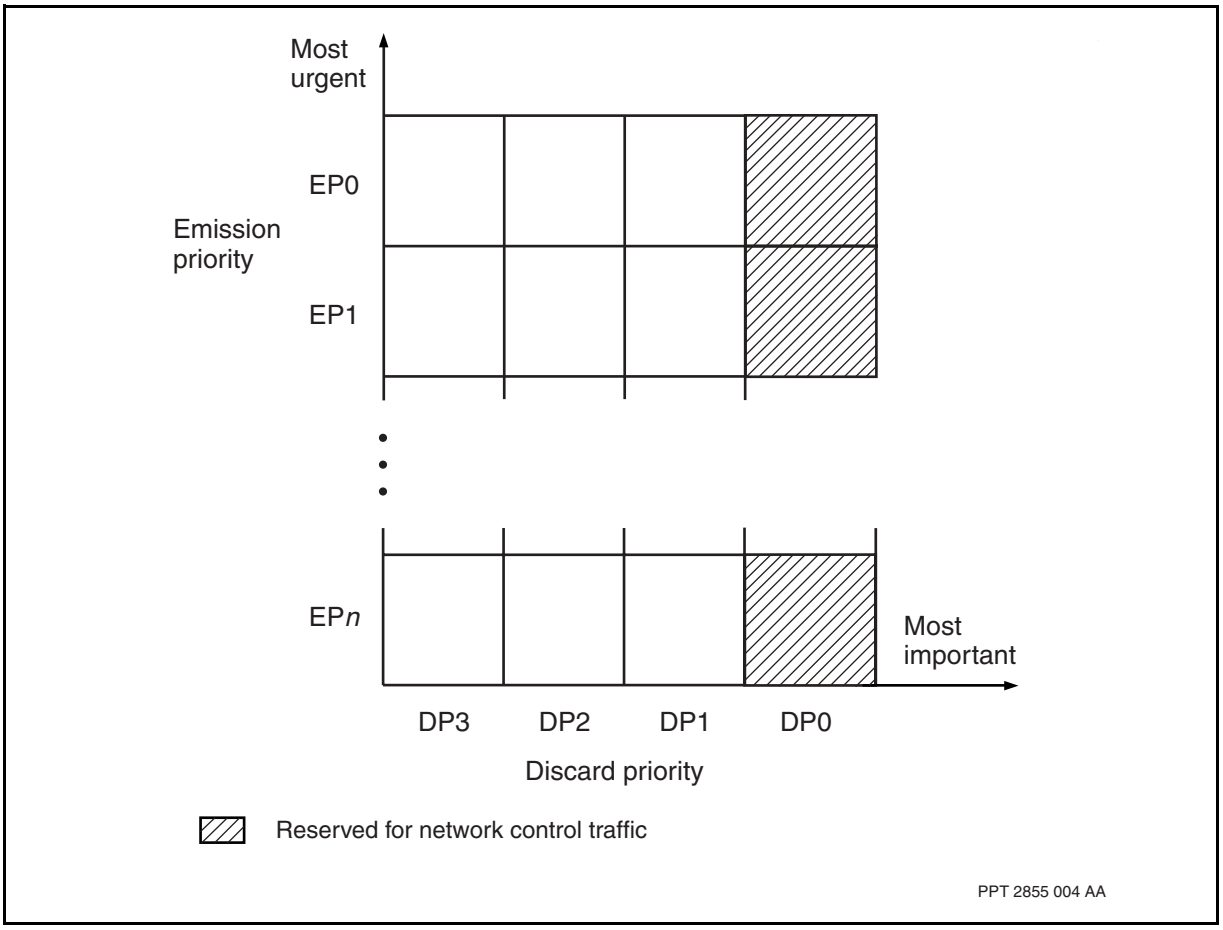
This system of emission and discard priorities appears on the following resources that have queuing requirements:

- the CPU
- the backplane
- the link

Each resource consists of a set of queues (the number depends on the resource and function processor type) where each queue has four congestion control levels. The congestion control levels are the thresholds against which the node evaluates the discard eligibility of incoming cells, which are marked as having one of the four discard priorities.

Copyright © 2005, Nortel
Nortel Confidential

General principles of emission and discard priorities on Multiservice Switch nodes

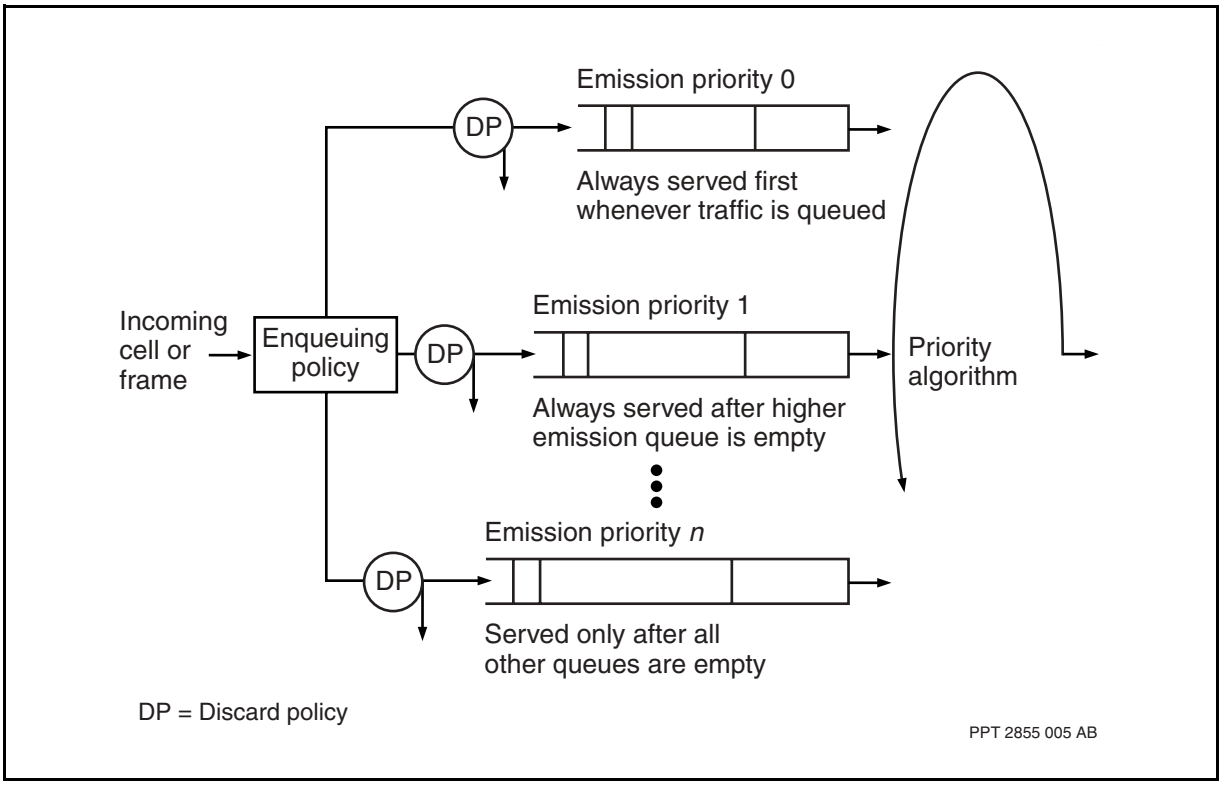


Emission priorities

For all traffic on a connection, the node determines an emission priority for each cell or frame based on the ATM service category for the connection. The number of available emission priorities depends on the type of function processor. The figure General principles of traffic on multiple queues (page 11) shows a generic scenario for enqueuing ATM traffic.

Copyright © 2005, Nortel

General principles of traffic on multiple queues



After determining emission priority, the node places cells into queues at the following points:

- function processor CPU
- backplane transmit
- link transmit

Free lists

Each function processor has a finite amount of shared memory. Memory allocation is configurable for the amount of connection space that the function processor requires. Also, the node reserves a fixed amount of memory for internal control functions. The remaining memory is available to the queues; this memory is known as the free list.

Queues access the memory (in cell or frame blocks) that you allocate to the cell and frame free lists. As a queue receives a cell or frame, it uses a block from the free list. On a typical function processor with several queues that service a constant flow of traffic, each queue constantly receives cells or frames and for each takes a block in the free list. As the scheduler processes enqueued cells for transmission on the backplane or link, the block becomes free (returned to the free list).

Copyright © 2005, Nortel
Nortel Confidential

Free list configuration is essentially a memory management issue, where the free list is configured to support frame or cell traffic, or both. For information on memory management, see Overview of memory management (page 103).

Service category to emission priority mapping

The table Service category and emission priority mapping by FP type (page 12) summarizes the default mapping for service category to emission priority by FP type. Note how the ATM IP FP can map UBR connections to an independent emission priority.

Service category and emission priority mapping by FP type

ATM service category	ATM IP and GQM-based FPs	CQC-based FPs	GQM-based FPs
CBR	1 per EP	1 per EP	1 per EP
RT-VBR	1 per EP	1 per EP	1 per EP
NRT-VBR	1 per EP	shared over 1 EP	1 per EP
UBR	1 per EP	shared over 1 EP	shared over 1 EP
UBR with MDCR	N/A	N/A	shared over 1 EP
This mapping does not apply to shaped emission priorities.			

Per-VC and common queuing

There are two methods for serving transmit traffic:

- per-VC queuing (also referred to as per-connection queuing)
- common queuing

The table Comparison of per-VC and common queuing (page 13) summarizes the characteristics of each type of queuing.

Copyright © 2005, Nortel Nortel Confidential

Comparison of per-VC and common queuing

Per-VC queuing	Common queuing
Greater fairness (equal access to the scheduler), since the node services all queues within an emission priority in round robin fashion.	Not necessarily fair, since cells from multiple connections are enqueued and served on a first-in first-out (FIFO) basis. For example, the first connection have cells enqueued may have sufficient time to fill 60% of the queue before cells from subsequent connections arrive at the node. The node does not serve cells from subsequent connections until it serves the cells from the first connection.
Less control over delay, since multiple per-VC queues have, in effect, a greater number of cells enqueued.	More control over delay characteristics, since you can anticipate usage of the queues, regardless of the number of connections using that queue.

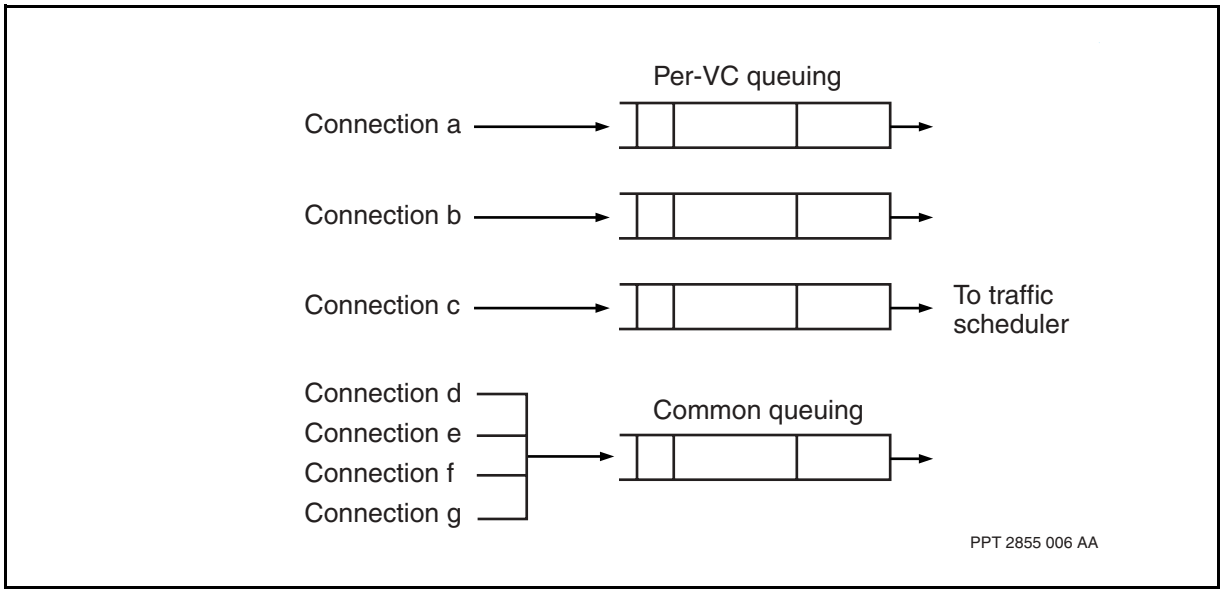
When configuring the queuing under an ATM interface, it is important to strike a balance between the application of per-VC queuing and common queuing. Typical configurations assign common queuing to real-time traffic and per-VC queuing to non-real-time traffic.

The figure General principles of per-VC and common queuing on Multiservice Switch nodes (page 14) shows how connections can be mapped to per-VC and common queues. The specific mechanisms and parameters that achieve this mapping depends on the function processor type. For specific information by function processor type, see the following sections:

- Queuing and scheduling on CQC-based FPs (page 32)
- Queuing and scheduling on ATM IP FPs (page 50)
- Queuing and scheduling on APC/PQC-based FPs (page 85)
- Queuing and scheduling on GQM-based FPs (page 94)

Copyright © 2005, Nortel
Nortel Confidential

General principles of per-VC and common queuing on Multiservice Switch nodes



General principles of Multiservice Switch traffic scheduling

A critical aspect of good queue management is the scheduling algorithm. Scheduling gives the service provider direct control over how the network serves every customer. The service policy must achieve the following:

- supports the quality of service (QOS) that each service category requires
- ensures a minimum bandwidth for every connection

Nortel Multiservice Switch nodes' scheduling algorithm ensures that all Multiservice Switch nodes in the network can use every cell transmit opportunity to achieve maximum network efficiency. The scheduler also provides an option to shape any connection to smooth out bursty traffic and ensure efficient use of network resources.

The scheduler services the queues, starting with the highest priority queue, according to the following rules:

- Regardless of the number of queues that apply to a point in the data path, the node always serves high priority traffic first, then lower priority traffic.
- If cells or frames arrive at a queue with a higher priority than the queue currently serviced, the scheduler moves back to the higher priority queue, beginning the process from the start.
- Within each emission priority, the scheduler serves queues using a fair algorithm such as round-robin or weighted fair queuing (WFQ).
- If minimum bandwidth guarantees (MBG) are in place, lower priority traffic may receive service over higher priority traffic, where the scheduler grants

Copyright © 2005, Nortel
Nortel Confidential

service opportunities on a statistical basis (ATM IP function processors only).

At points where there is severe congestion, the lower priority queues begin to fill up and eventually reach a congestion level.

The figure General principles of traffic on multiple queues (page 11) illustrates a general model of how the traffic scheduler serves queues assigned to different emission priorities. The specifics of how the scheduler serves the queues depends on the type of function processor. For specific information by function processor type, see the following sections:

- Queuing and scheduling on CQC-based FPs (page 32)
- Queuing and scheduling on ATM IP FPs (page 50)
- Queuing and scheduling on APC/PQC-based FPs (page 85)
- Queuing and scheduling on GQM-based FPs (page 94)

Link transmit queues

Link transmit queues require consideration in the following areas:

- Per-VC queue limits and thresholds: general (page 15)
- Queuing and traffic shaping (page 17)
- Impact of link transmit queue limit on QOS and link usage (page 17)
- Per-VC queue limit configuration (page 18)

The figure Transmit queuing approaches on Multiservice Switch nodes (page 16) summarizes the transmit queuing approaches available.

Per-VC queue limits and thresholds: general

Queue limits are configurable for all service categories. The principles for defining queue limits are:

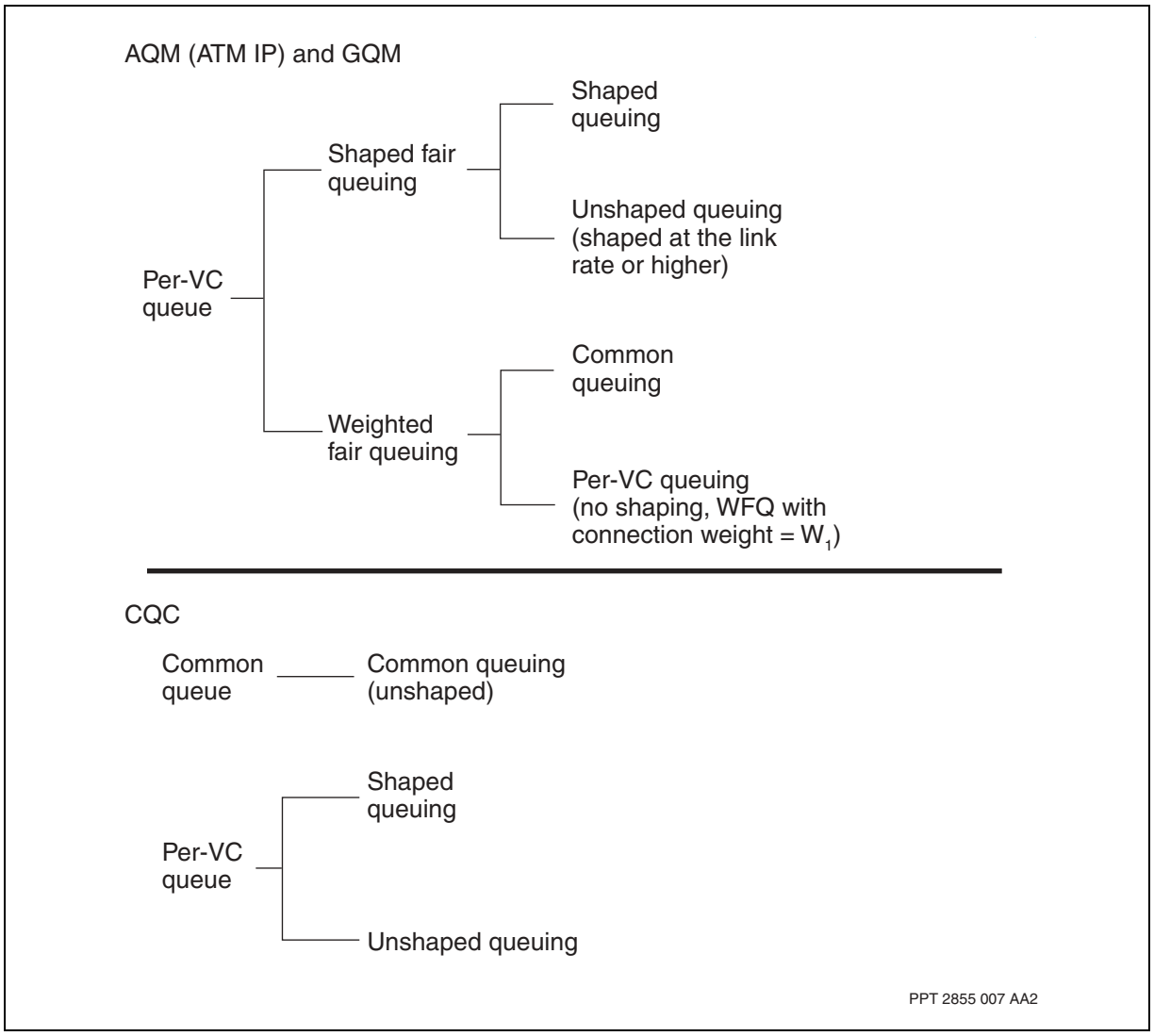
- Queue limits for real time services (CBR and RT-VBR) are governed by delay and CDV constraints.
- For non-real-time services where cell loss ratio (CLR) is important, queue limits are designed to absorb the maximum traffic burst which can arrive from an application.

  There may be a limit to the amount of buffer space which an application can use. Configuring the queue limit larger than this amount is wasteful since the queue can never be filled. Some data applications may operate under delay constraints which also limit the useful queue length. Configuring a queue limit which causes protocol time-outs and re-transmissions is wasteful of network bandwidth.

Copyright © 2005, Nortel

- Common queues use a single queue per emission priority. Per-VC queues use one queue per VCC or VPC, or per VCC within a VPC. Limits are configured by specific values for attributes or through auto-configuration.
- There is a trade-off between per-VC fairness and cell loss. If per-VC queues are too small, cell loss may become an issue on a given queue. If there are many per-VC queues which are very large, there is an increased chance for free list exhaustion. Free list exhaustion effectively negates the value of per-VC queues and results in common queue behavior for service categories.

Transmit queuing approaches on Multiservice Switch nodes



Copyright © 2005, Nortel
Nortel Confidential

Queuing and traffic shaping

If traffic shaping is enabled for a connection, the node always serves traffic on the per-VC queues. If shaping is not allowed for the ATM interface, the node serves traffic for common queues only.

If shapers are allocated to a port but traffic shaping is disabled (that is, connections are unshaped), you can configure the node to use one of the following unshaped transmit queuing methods:

- per-VC, in which the node allocates a per-VC queue to each unshaped connection, and services each queue with the same fairness as any shaped connections.
- common, in which the node directs any unshaped connection to a common queue by service category.
- auto-configured, in which the node determines queue method based on the ATM interface configuration for per-VC queues

On a port that has traffic shaping enabled, all switched connections use per-VC queuing. On a port with traffic shaping disabled, all switched connections use the configuration for unshaped transmit queuing (per-VC or common). On CQC-based FPs, this behavior depends on the configuration for unshaped transmit queuing.

Impact of link transmit queue limit on QOS and link usage

Changing the transmit queue limits of each service category affects the QOS and the link utilization.

For real-time applications (CBR and RT-VBR), the transmit queue limit can control the amount of cell delay variation (CDV) that the link transmit queue introduces into the connection:

$$\text{CDV} = \text{transmit queue threshold} \div \text{service rate}$$

The transmit queue threshold is a percentage of the transmit queue limit. It is set at approximately 90% for CBR and RT-VBR traffic. The exact value can be displayed through the operational attribute transmit queue threshold under the VCC. The larger the queue limit the larger the CDV due to cell buffering. For CBR, the service rate is equal to the link rate (no shaping). For RT-VBR, the service rate is a function of the amount of admitted CBR traffic and the use of RT-VBR per-VC queuing (shaped or unshaped).

If common queuing is used for RT-VBR, then the CDV is calculated as follows:

$$\text{CDV} = \text{transmit queue threshold} \div \text{link rate} - \Sigma(x)$$

where x is the PCR for all CBR connections.

Copyright © 2005, Nortel Nortel Confidential

If traffic shaping is used for RT-VBR, the CDV for a given connection is calculated as:

CDV = transmit queue threshold ÷ shaping rate

If per-VC queuing without traffic shaping is used for RT-VBR, then the CDV is calculated as:

CDV = transmit queue threshold ÷ [(link rate - Σ(*x*)) ÷ *y*]

where *x* is the PCR for all CBR connections and y is the number of RT-VBR connections.

As the queue limit increases, CLR due to buffering loss decreases and link usage increases. For CBR, RT-VBR, and NRT-VBR, the transmit queue threshold is used in the equivalent cell rate (ECR) calculation to determine how much bandwidth a given connection requires to meet its CLR requirements based on the traffic descriptor for the connection. As the queue limit increases, ECR decreases. As a result, a higher link utilization is achievable with a larger queue limit.

For non-real-time applications (NRT-VBR, and UBR), the transmit queue limit can control the amount of CLR introduced by the link transmit queue and hence, can control the achievable link utilization.

Per-VC queue limit configuration

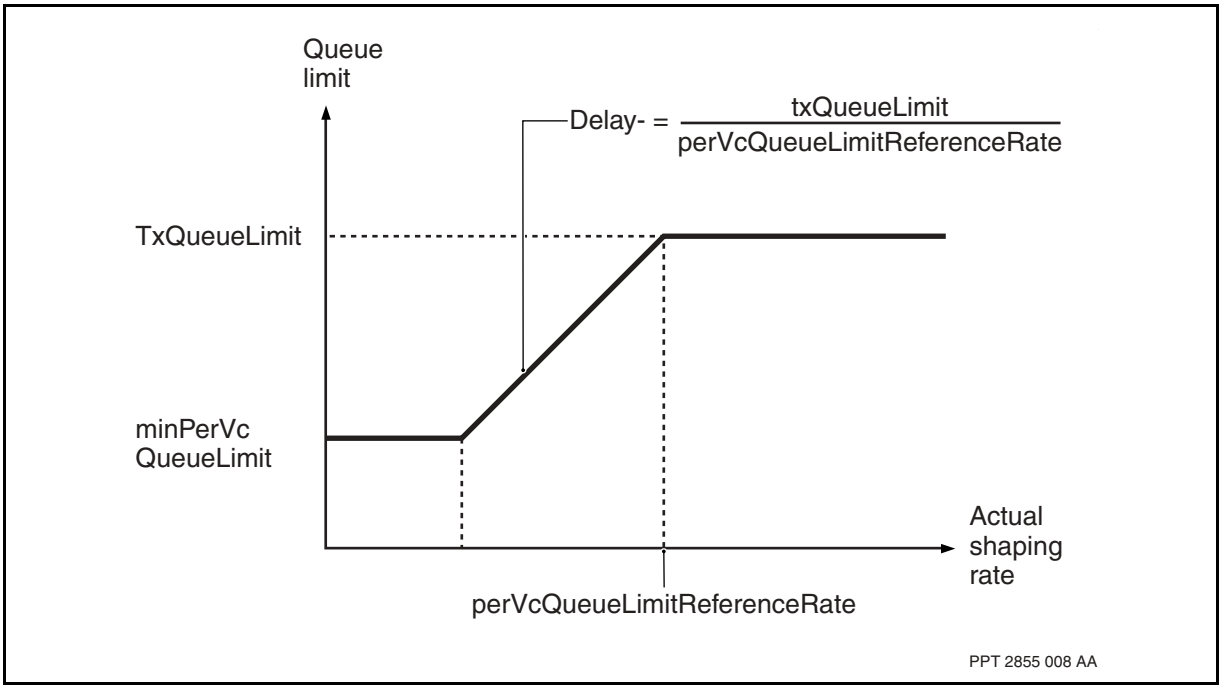
The per-VC queue limit is a function of four configurable parameters:

- transmit queue limit (*txQueueLimit*), which defines the maximum queue length for the queues used to buffer the traffic of the corresponding service category
- actual shaping rate
- the minimum per-VC queue limit (*minPerVcQueueLimit*), which defines the minimum queue limit for the per-VC queues for connections of the corresponding service category
- the reference rate for the per-VC queue limit (*perVcQueueLimitReferenceRate*), which defines the shaping rate used to calculate the tolerable delay for per-VC queues for connections of the corresponding service category (delay is the transmit queue limit divided by the reference rate)
- For any ATM FP, an attempt to set up a per-VC queue beyond the limit will fail with alarm 7039 2001 generating against the component *Lp Eng Arc* for resource exhaustion.

Copyright © 2005, Nortel Nortel Confidential

The figure General principles of interaction between parameters for defining per-VC queue limits (page 19) shows the relationships between these parameters.

General principles of interaction between parameters for defining per-VC queue limits



When traffic from two or more service categories share an emission priority, the rules for setting these parameters differs depending on FP type. For specific information by FP type, see the following sections:

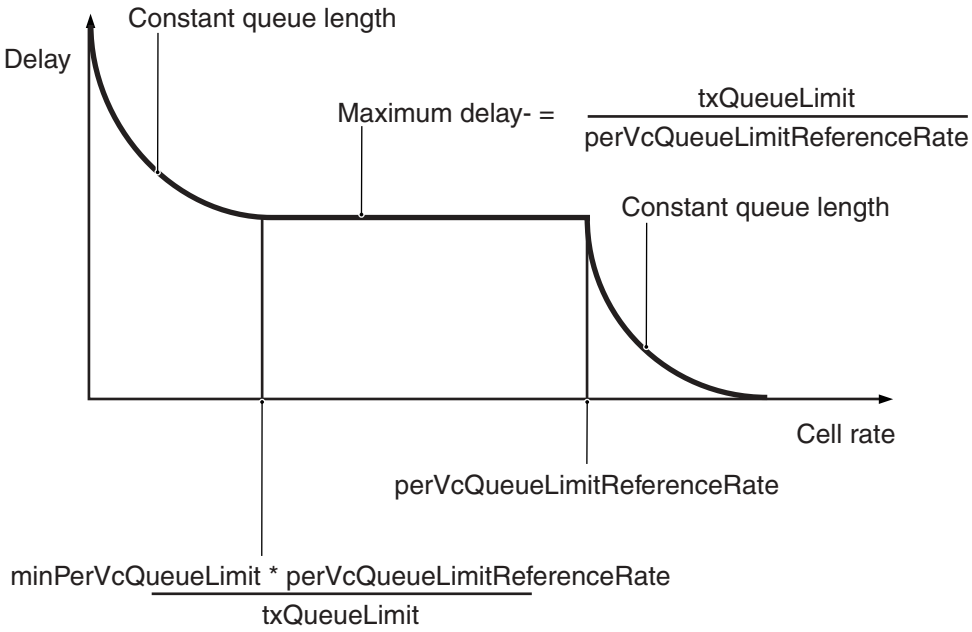
- Queuing and scheduling on CQC-based FPs (page 32)
- Queuing and scheduling on ATM IP FPs (page 50)
- Queuing and scheduling on APC/PQC-based FPs (page 85)
- Queuing and scheduling on GQM-based FPs (page 94)

The value for the transmit queue limit is the common queue limit that acts as a reference point from which the node derives per-VC queue limits. Further, this value is an upper bound for the per-VC queue limit.

The figure General principles of interaction between per-VC queuing parameters: delay over cell rate (page 20) shows the relationships between delay, cell rate (shaping rate or link rate), transmit queue limit, minimum per-VC queue limit, and reference rate.

Copyright © 2005, Nortel
Nortel Confidential

General principles of interaction between per-VC queuing parameters: delay over cell rate



The table Configurable parameters for per-VC queue default limits (page 20) summarizes the values of these parameters for these FPs:

- ATM IP
- CQC-based
- APC-based
- GQM-based

Configurable parameters for per-VC queue default limits

FP type and parameter	CBR	RT-VBR	NRT-VBR	UBR
ATM IP high-speed FPs				
txQueueLimit	96 cells	480 cells	10 240 cells	
minPerVcQueueLimit	88 cells			
perVcQueueLimitReferenceRate	65 511 cells/s	14 740 cells/s	65 511 cells/s	
CQC-based high-speed FPs				
txQueueLimit	96 cells	480 cells	2304 cell	
minPerVcQueueLimit	88 cells			
perVcQueueLimitReferenceRate	n/a	14 740 cells/s		

(1 of 2)

Copyright © 2005, Nortel
Nortel Confidential

Configurable parameters for per-VC queue default limits (continued)

FP type and parameter	CBR	RT-VBR	NRT-VBR	UBR
CQC-based low-speed FPs				
txQueueLimit	96 cells	480 cells	1792 cells	
minPerVcQueueLimit	88 cells			
perVcQueueLimitReferenceRate	n/a	3685 cells/s		
APC-based 16pOC3 FP				
txQueueLimit	96 cells	480 cells	10 240 cells	10 240 cells
minPerVcQueueLimit	88 cells	88 cells	88 cells	92 cells
APC-based 4pOC12 FP				
txQueueLimit	384 cells	1 920 cells	10 240 cells	10 240 cells
minPerVcQueueLimit	88 cells	88 cells	88 cells	92 cells
GQM-based 16pOC3PosAtm FP				
txQueueLimit	96 cells	1280 cells	10 240 cells	10 240 cells
minPerVcQueueLimit	88 cells	88 cells	88 cells	88 cells
perVcQueueLimitReferenceRate	65 511 cells/s	39 307 cells/s	65 511 cells/s	65 511 cells/s
GQM-based 4pMRPosAtm FP in 4-port OC12 mode				
txQueueLimit	386 cells	5120 cells	40 960 cells	40 960 cells
minPerVcQueueLimit	88 cells	88 cells	88 cells	88 cells
perVcQueueLimitReferenceRate	262 044 cells/s	157 228 cells/s	262 044 cells/s	262 044 cells/s
GQM-based 4pMRPosAtm FP in 1-port OC48 mode				
txQueueLimit	1544 cells	20480 cells	40 960 cells	40 960 cells
minPerVcQueueLimit	88 cells	88 cells	88 cells	88 cells
perVcQueueLimitReferenceRate	1 048 176 cells/s	628 912 cells/s	262 044 cells/s	262 044 cells/s

(2 of 2)

Discard priority: all function processors

Discard policy, as it applies to queues and free lists, ensures that traffic does not overload function processor (FP) memory resources.

Discard policy consists of a set of priorities that are assigned to individual data units (frames or ATM cells) based on configured parameters or signaled parameters in the information elements. Discard priority determines if the node must discard the incoming cell given the congestion state of the target

Copyright © 2005, Nortel Nortel Confidential

queue. Each cell that requires buffering in one of the queues has a discard priority that the Nortel Multiservice Switch node assigns. Each queue can hold a random mixture of cells with different discard priorities.

There are four discard priorities ranging from 0 (most important traffic) through to 3 (least important traffic). For example, the node assigns CBR CLP0 traffic a discard priority=1 and assigns CBR CLP1 a discard priority=3. The node reserves the range above level 0 for internal management traffic.

For some applications, the application configuration can override the default discard priority setting. These applications include:

- path-oriented routing system (PORS)
- frame relay
- dynamic packet routing system (DPRS)

If the FP supports DPRS services, it uses level 1 for DPRS control traffic (the node also uses level 1 for CBR CLP0 traffic).

Queue control (limits and thresholds) based on cell and frame discard priority and queue congestion control states apply to per-VC queuing, common queuing, and free lists.

Queue discard priorities

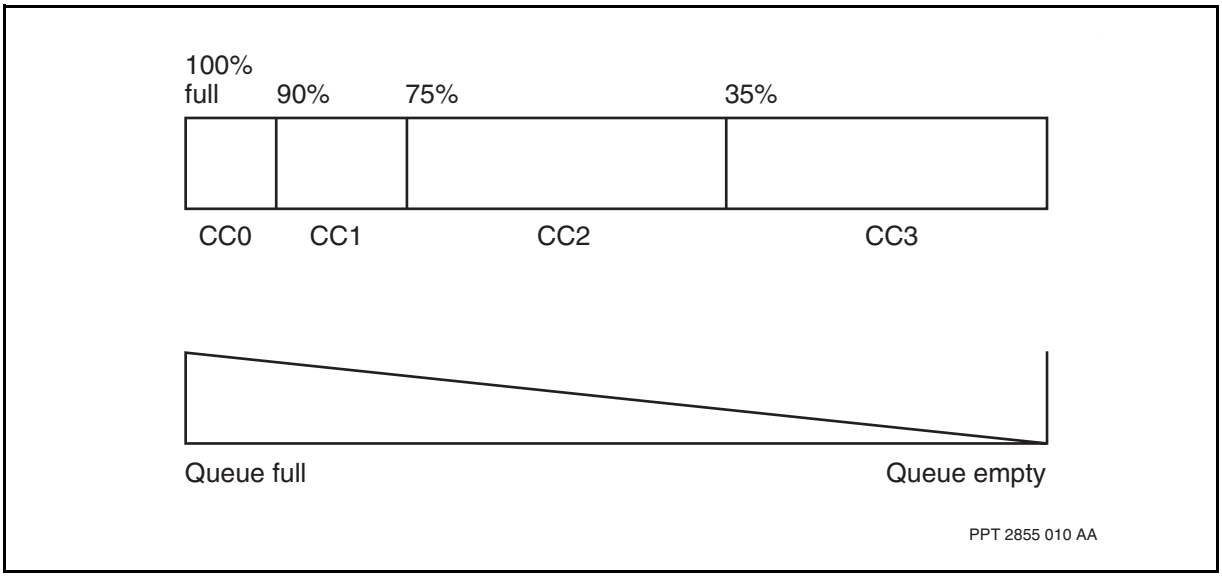
The node controls discard priority for each queue through a set of four congestion control levels. These levels define the congestion state of the queue in terms of percentage filled:

- congestion control level 3 (CC3) represents a queue level that is less than 35% of the available memory available to the queue
- congestion control level 2 (CC2) represents a queue level that is between 35% and 75%
- congestion control level 1 (CC1) represents a queue level that is between 75% and 90%
- congestion control level 0 (CC0) represents a queue level that is greater than 90%

CC0 indicates the most severe congestion state, such that the queue is subject to the most drastic actions available for reducing congestion. CC3 indicates little or no congestion such that no remedial action is necessary.

Copyright © 2005, Nortel Nortel Confidential

Implementation of queue congestion control levels



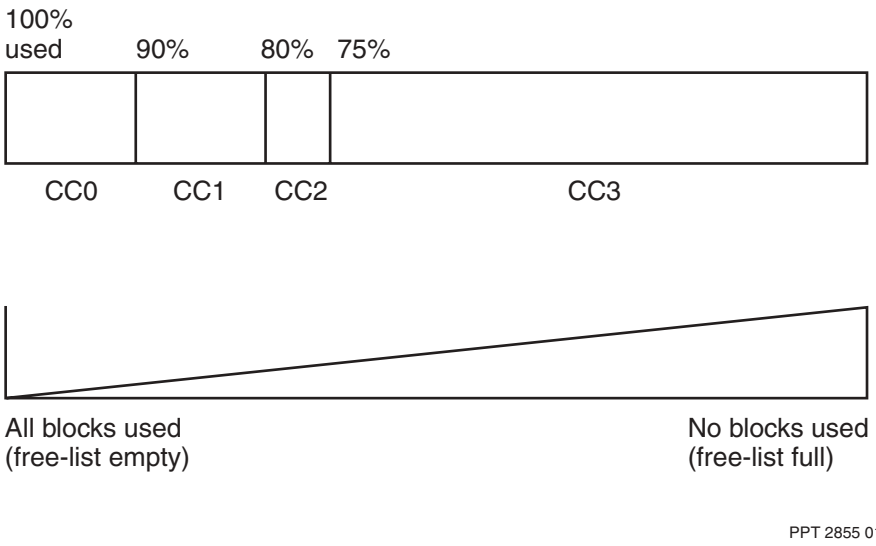
Free list discard priorities

The node controls discard priority for the cell and frame free list through a set of four congestion control (CC) levels. As with per-VC and common queues, these levels define the congestion state of the free list in terms of percentage filled:

- CC3 represents a free list level that is less than 75% of the available memory available to the free list
- CC2 represents a free list level that is between 75% and 80%
- CC1 represents a free list level that is between 80% and 90%
- CC0 represents a free list level that is greater than 90%

Copyright © 2005, Nortel
Nortel Confidential

Implementation of free list congestion control levels



For free lists, congestion control levels provide similar indications as the levels for the queues. Compare the free list implementation in this figure with the queue implementation in the figure Implementation of queue congestion control levels (page 23). Note that the threshold percentages differ.

Depending on the queue configuration, congestion has the following effects on the free lists:

- if the queues are allocated more memory than the free lists define, the free lists are exhausted
- if queues are allocated less memory than the free lists define, free list exhaustion does not occur (although queue congestion can still occur)

The discard priority of the cell or frame and the congestion level of the target emission queue determines if the node enqueues the cell.

Priority and resource interactions

Interaction occurs between the following:

- emission priorities
- discard priorities
- connections
- queues and queue types
- free lists
- weighted random early detection (WRED)

Copyright © 2005, Nortel
Nortel Confidential

The level of interaction depends on the configuration for mapping connections to queues, and the types of queues required.

Discard priority applies to an emission priority independently of the queue states under other emission priorities. That is, queues under one emission priority may discard cells while the queues under another priority do not. These usage patterns occur when there is no free list congestion (that is, when less than 75% of free list is in use) and for both per-VC and common queuing. Further, with per-VC queuing some queues within a single emission priority may be congested (discard cells) while others are not.

Discard priority can apply across emission priorities when there is free list congestion (75 percent or more free list is in use). This level of interaction occurs because the free lists are not defined in terms of emission priority (as queues are). That is, when the free list is congested, discards occur for traffic of all emission priorities regardless of the congestion state of individual queues.

For example, there may be many per-VC queues, each at less than 35% capacity (all queues operating under congestion control level 3), and all traffic can be accepted into the queues (no discards). This pattern continues as long as free list usage is below 75 percent. However, a sufficient number of these queues may use more than 75 percent the free list (the free list is operating under congestion control level 2 while individual queues are at congestion control level 3). Due to free list congestion, the interface discards all traffic with discard priority 3, regardless of emission priority.

When shaping is not implemented, traffic from different service categories can be mapped to separate emission priorities, and therefore to separate queues. In this configuration, the following characteristics are true within a service category:

- if per-VC queuing is configured: the queue controller or queue manager (depending on the FP type) discards independently for each queue and does not make discard decisions across multiple per-VC queues.
- if common queuing is configured: there is interaction between connections. That is, two or more connections with the same emission priority contend for queue resources.

When shaping is implemented, two or more service categories may map to a single emission priority. As a result, a single queue may hold cells from difference service categories. Discard is based on the discard priority of cells for each connection.

Copyright © 2005, Nortel Nortel Confidential

Example of the discard priority process

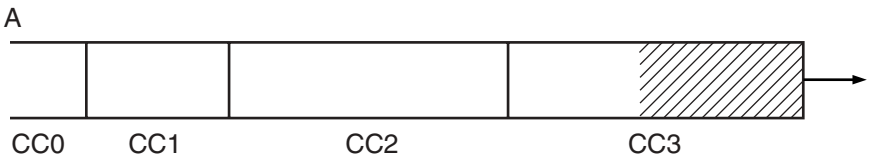
The figure Example of discard priority process for queue congestion control (page 27) shows a series of examples that illustrate how discard priority works. The examples are based on the discard priorities for a common queue and apply in similar fashion for free lists.

In each example, the node preserves the cells that it has already accepted into one of the queues and discards only those new incoming cells that have a discard priority that is below the threshold for the target queue. In example A, the queue level is below the threshold, so the queue accepts cells for all discard priorities. As the queue fills up, as in example B, the queue level equals the first threshold (35%). As long as the queue level is greater than or equal to 35%, the node discards all new incoming cells with a discard priority of 3. This process continues as the queue level rises and falls. Example D illustrates severe congestion in which the node accepts only network control cells into the queue.

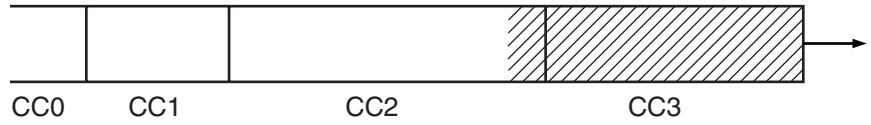
The result is that even though congestion increases, subscriber traffic with a high discard priority (1) continues to receive service. At points where there is severe congestion, the node continues to support network control traffic.

Copyright © 2005, Nortel
Nortel Confidential

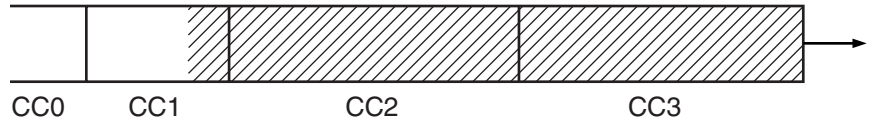
Example of discard priority process for queue congestion control



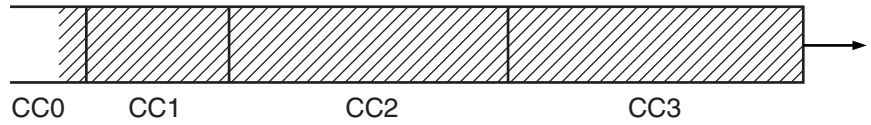
Example A: the congestion level is below the first threshold. The queue accepts traffic for all discard priorities for this queue.



Example B: the congestion level is just past the first threshold. The emission queue now accepts traffic for discard priorities 2, 1, and 0. The node discards traffic with priorities 3 only for this queue.



Example C: the congestion level is just past the second threshold. The emission queue now accepts traffic for discard priorities 1 and 0. The node discards priorities 3 and 2.



Example D: the congestion level is just passed the third threshold. The emission queue now accepts traffic for discard priority 0 only (network control traffic). The node discards all subscriber traffic for this queue.

PPT 2855 012 AA

Traffic mapping to internal discard and emission priority

Traffic mapping depends on the traffic flow direction (from the link or toward the link) and the traffic type (frame or cell relay traffic). For this reason, traffic (frame and cell relay) mapping to internal emission and discard priorities is described separately for traffic flow in both the receive direction (traffic flow traversing an ATM FP from the link toward the backplane or the CPU) and the transmit direction (traffic flow traversing a FP from the backplane toward the link).

Copyright © 2005, Nortel
Nortel Confidential

Receive mapping - cell relay

For traffic received from an ATM link, the CLP value from the ATM cell header and the ATM service category value for the connection are used to determine the internal discard priorities of each cell. Once a cell is received from the link, the cell's internal discard priority is compared with the memory congestion state. If the discard priority is greater than the memory congestion status, the cell is discarded; otherwise, memory is allocated for the cell.

Next, the cell is placed in the appropriate backplane queue. Before queuing the cell, the cell's internal discard priority value is compared with the backplane congestion state, and the cell is discarded if its discard priority value is greater than the backplane congestion state. See the table Receive mapping: ATM service category and CLP to discard priorities (page 29) for a summary of receive mapping to internal discard priorities.

Receive mapping - frame

For traffic received from an ATM link, the cell loss priority (CLP) value from the ATM cell header and the ATM service category (of the ATM connection that the cell is received through) are used to determine the internal discard priorities of each cell.

As with receive mapping for cell relay, once a cell is received from the link, the cell's internal discard priority value is compared with the value for the memory congestion state. If the discard priority is greater than the memory congestion status, the cell is discarded along with any other cells belonging to the same frame.

Currently, some frame traffic carried over ATM (frame relay using the Logical Trunk feature) is encapsulated through ATM adaptation layer 5 (AAL5) and requires processor intervention in the receive direction (HTDS traffic is hardware forwarded). Once the frame has been reassembled from ATM cells, the frame is queued directly onto the CPU queue. The frame's discard priority is derived from the transporting cell's ATM service category and CLP. Before enqueuing the frame, the frame's internal discard priority value is compared with the processor congestion state. The frame is discarded if its discard priority is greater than the CPU queue congestion state.

Software overrides

Once the frame is processed, the software can override the discard and emission priorities of the frame before it is sent to the queue. When a cell arrives (receive direction), the software determines the discard and emission priorities according to the table Receive mapping: ATM service category and CLP to discard priorities (page 29). Applications such as FR-ATM and PORS can change these software determined values. For more information about the FR-ATM application, see NN10600-920 *Nortel Multiservice Switch 7400/*

Copyright © 2005, Nortel Nortel Confidential

15000/20000 Operations: Frame Relay to ATM Interworking. For more information about PORS, see NN10600-435 *Nortel Multiservice Switch 7400/ 15000/20000 Operations: Path-Oriented Routing System*.

See the table Receive mapping: ATM service category and CLP to discard priorities (page 29) for a summary of receive mapping to internal emission and discard priorities. For AAL5, the frame CLP (internal) is set to 1 if the CLP in any of the constituent ATM cells was equal to 1.

Receive mapping: ATM service category and CLP to discard priorities

Service category	CLP	Discard priority
CBR	0	Very important (1)
	1	Least important (3)
RT-VBR	0	Very important (1)
	1	Least important (3)
NRT-VBR	0	Important (2)
	1	Least important (3)
UBR	0	Least important (3)
	1	Least important (3)

Transmit mapping - cell relay

For transmitted cells, emission priority is used to select one of the link transmit queues. The emission priority is derived from the ATM service category as configured or derived from signaled broadband bearer capability information element (BBC-IE) for a particular ATM connection.

For ATM cell relay traffic, the discard priority is derived from the cell's CLP and connection ATM service category. See the table Transmit mapping: CLP and ATM service category to emission and discard priority (for cell relay traffic) (page 30) for a summary of cell relay discard priority transmit mapping. The discard priority is used to implement the discard policy when the cell demands resources that are congested (link and memory).

The CLP value is unchanged for ATM cell relay traffic in the transmit direction. Traffic in the transmit direction is not tagged.

Copyright © 2005, Nortel Nortel Confidential

Transmit mapping: CLP and ATM service category to emission and discard priority (for cell relay traffic)

Service Category	CLP	Emission priority	Discard priority
CBR	0	high	Very important (1)
	1		Least important (3)
RT-VBR	0		Very important (1)
	1		Least important (3)
NRT-VBR	0		Important (2)
	1		Least important (3)
UBR	0		Least important (3)
	1	low	Least important (3)

The mapping of a service category to a specific emission priority depends on the FP type. See the following sections:

- CQC emission priorities (page 33)
- ATM IP emission priorities (page 51)
- Schedulers for GQM-based FPs (page 96)

Transmit mapping - frame

For frames in the transmit direction, the emission priority is used to select one of the three link transmit queues. The emission priority is derived from the ATM service category for the connection. The service category is configured or signaled for the ATM connection to which the frame is forwarded for processing.

The internal discard priority is already set when the frame is received from the backplane and is used to implement the discard policy when demand for the resources (link and memory) exceeds capacity. This priority is based on the frame's priority. The discard priority of the frame and the configured ATM service category determine the CLP value of the cells comprising the AAL5 frame. See the table Transmit mapping: discard priority and ATM service category to CLP (for AAL5 frames) (page 31). Overview of queuing and traffic scheduling (page 8)

Copyright © 2005, Nortel
Nortel Confidential

Transmit mapping: discard priority and ATM service category to CLP (for AAL5 frames)

Service category	DP = 0	DP = 1	DP = 2	DP = 3
CBR	0	0	0	1
RT-VBR	0	0	0	1
NRT-VBR	0	0	0	1
UBR	0	1	1	1
DP = discard priority.				

Copyright © 2005, Nortel
Nortel Confidential

Queuing and scheduling on CQC-based FPs

This chapter describes queue management techniques on Nortel Multiservice Switch node function processors that are based on the cell queue controller (CQC) ASIC.

Navigation

- Overview to CQC queuing and scheduling (page 32)
- CQC emission priorities (page 33)
- Traffic scheduling for CQC-based function processors (page 35)
- CQC discard priority (page 38)
- Interaction between emission and discard priorities (page 39)
- Free list and queue configurations (page 42)
- CQC queue limits and congestion thresholds (page 43)
- CQC queuing and scheduling for VP termination (page 49)

Overview to CQC queuing and scheduling

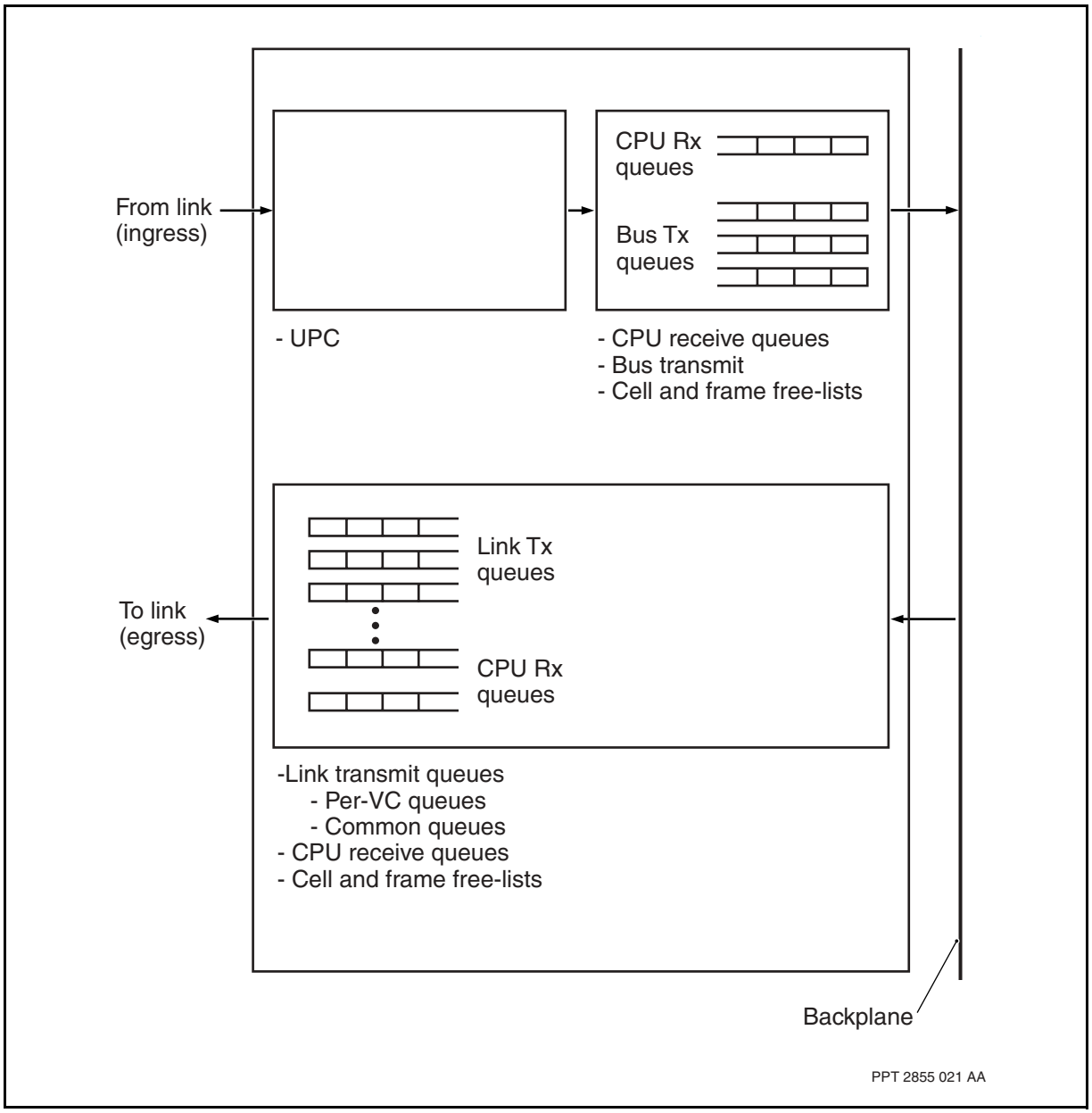
Emission priorities apply at ingress (for free lists, CPU queues, and bus queues) and at egress (for free lists and link transmit queues, both common and per-VC).

The figure Cell queue memory on CQC-based function processors (page 33) shows where queues occur on the CQC-based function processor.

Copyright © 2005, Nortel

Nortel Confidential

Cell queue memory on CQC-based function processors



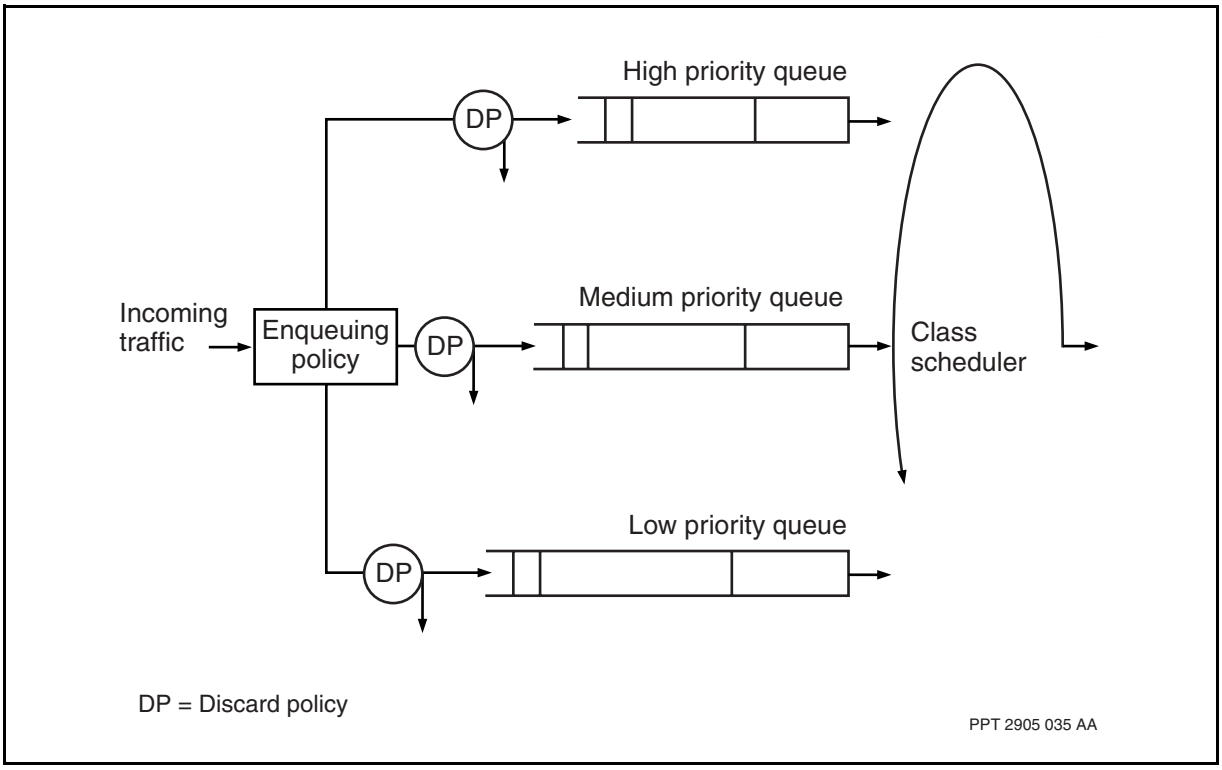
CQC emission priorities

CQC-based function processors have three emission priorities: high, medium, and low. The node puts cells in the queues which are in turn served by the traffic scheduler.

The figure CQC emission priorities and scheduler (page 34) shows the basic structure of PQC queuing and the relationship to the traffic scheduler.

Copyright © 2005, Nortel
Nortel Confidential

CQC emission priorities and scheduler



For each ATM service category, the CQC emission priorities are fixed values and cannot be changed. See the table Fixed emission priority values for CQC-based function processors (page 34).

Fixed emission priority values for CQC-based function processors

ATM service category	Fixed emission priority value
CBR	high
RT-VBR	medium
NRT-VBR	low
UBR	low

Per-VC queues and common queues

CQC-based function processors have two types of link transmit queues:

- common queues
- per-VC queues

Copyright © 2005, Nortel

On common queues, the node supports high, medium, and low emission priorities. All common queues hold traffic that belongs to multiple connections. On per-VC queues, the node supports medium and low emission priorities.

See Traffic scheduling for CQC-based function processors (page 35) for specifics on how the traffic scheduler services these queues.

Traffic scheduling for CQC-based function processors

The node services the common and per-VC queues through a priority scheme in the traffic scheduler.

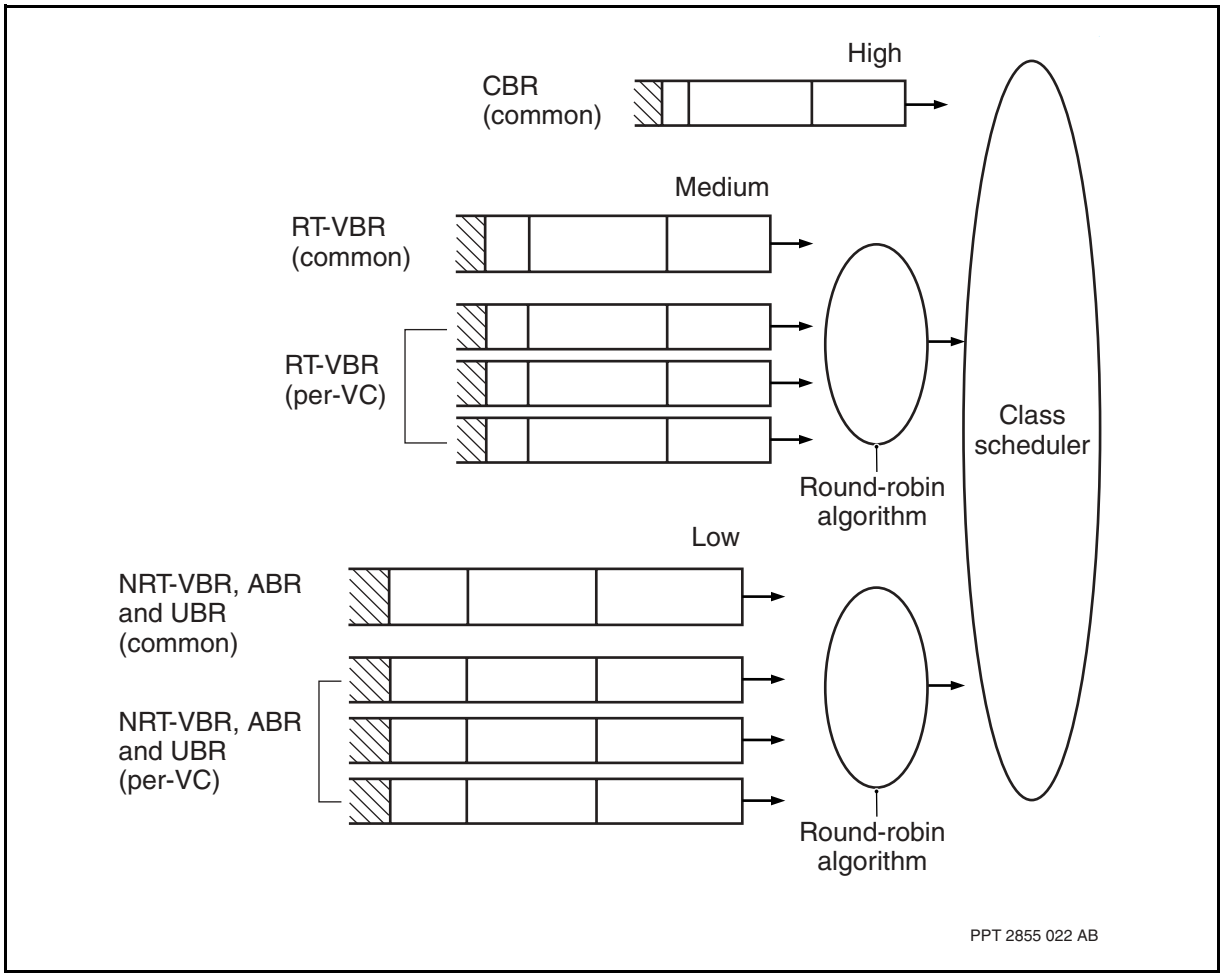
The traffic scheduler services the high-priority common queue first. When the high-priority common queue is empty, the scheduler services the medium priority queues and, when it is empty, services the low priority queues.

Within the medium and low priority queues, the scheduler serves per-VC and common queues through a round-robin algorithm. Each per-VC queue and the common queue have equal opportunity for service through the scheduler. If traffic arriving at a higher priority queue interrupts the scheduler, it suspends round-robin polling of the queue at this priority. When the scheduler returned to this priority, the round-robin algorithm resumes where is left off.

The figure CQC queuing and traffic scheduling (page 36) shows how CQC traffic scheduling works.

Copyright © 2005, Nortel Nortel Confidential

CQC queuing and traffic scheduling



Traffic scheduling and shaping stacks

All per-VC queues have a shaping stack. There are 24 shaping stacks at the medium priority and 24 at low priority. Through a round-robin algorithm, the scheduler serves multiple eligible stacks within the same priority level. Common queues do not use traffic shaping.

When per-VC queuing is enabled, unshaped connections are treated with the same fairness as shaped connections in the round-robin algorithm. However, per-VC queues have a maximum rate of 58 962 cells/s (25 Mbits/s) when traffic shaping is allocated to two or three ports under ATM resource control for the function processor.

The scheduler is also responsible for servicing the three common queues. Hence, a per-VC queue receives service (emits traffic) only when the common and per-VC queues at the higher priority are empty.

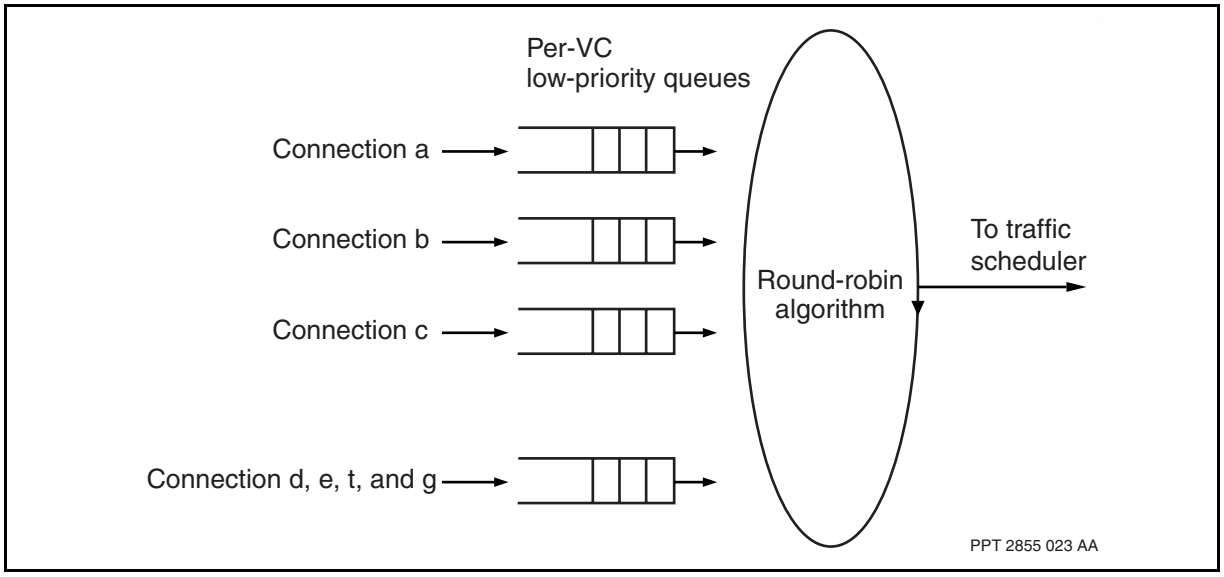
Copyright © 2005, Nortel
Nortel Confidential

Per-VC queuing is only available if traffic shapers have been allocated to this link at the level of ATM resource control. This condition must be met because per-VC queuing requires shapers allocated to the link. The maximum shaping rate available is dependent on the number of ports being shaped on the ATM-FP, as well as on the global scaling factor. For more information, see the chapter on section on traffic shaping for CQC function processors in NN10600-706 *Nortel Multiservice Switch 7400/15000/20000 ATM Traffic Shaping and Policing Fundamentals*. The configurations shown in that section apply even when traffic shaping is not used.

Mixed queues with a single priority

Per-VC queue traffic (shaped) and common queue traffic (unshaped) over the same priority onto a link can cause scheduling unfairness to common queue traffic. The figure Example of traffic over per-VC and common low-priority queues (page 37) shows this mix of queue types on the same ATM interface and its result.

Example of traffic over per-VC and common low-priority queues



In the figure, observe the following characteristics:

- connections a, b, and c each use a dedicated per-VC queue (shaped)
- connections d, e, f, and g use the same common queue

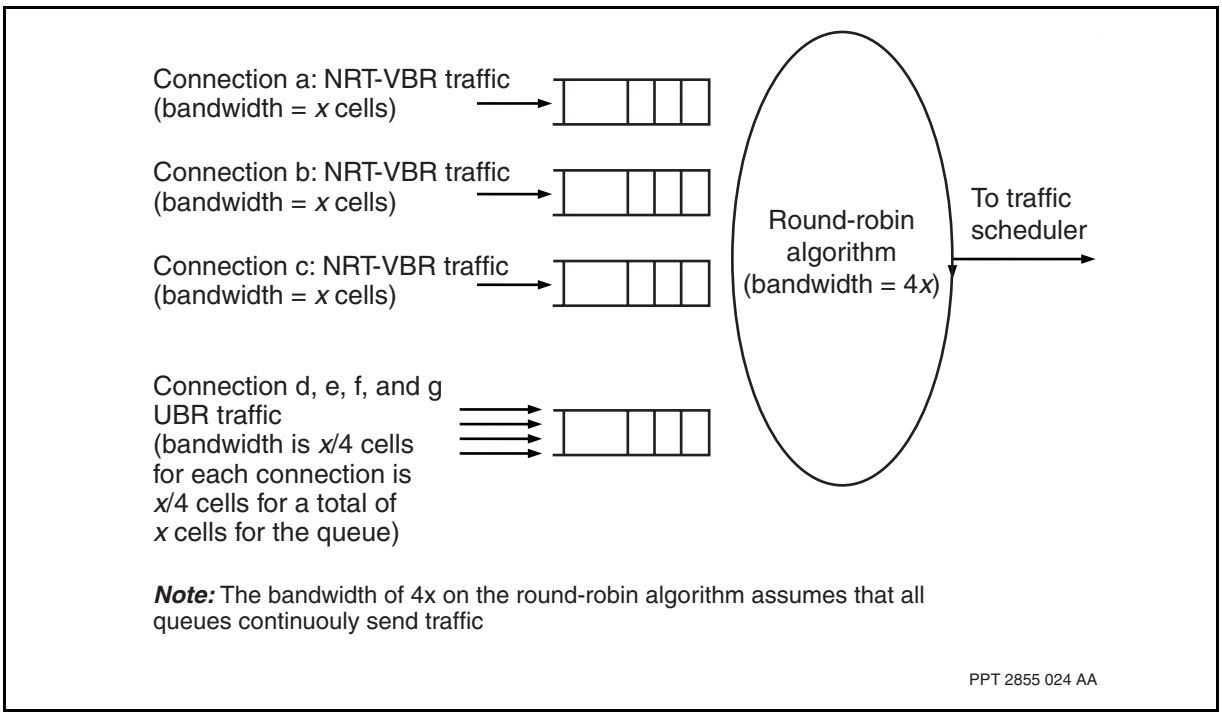
The node serves each of the four queues in a round robin fashion. This means that traffic for each connection in the per-VC queues is always guaranteed service within each round robin. On the other hand, traffic for each connection in the common queue must contend for resources both with each connection in the common queues and with connections in the per-VC queues. Emission from the common queue is based on first-in first-out without ensuring fair

Copyright © 2005, Nortel
Nortel Confidential

scheduling on a cell basis between connections. Unfairness is particularly present if the per-VC traffic is consuming enough resources to cause congestion of the common queue and if there are enough connections sending bursty traffic over the common queue.

None the less, this configuration does have sound application on the lowest priority where NRT-VBR and UBR traffic is queued. The figure Example of NRT-VBR and UBR traffic over per-VC and common queues (page 38) shows how NRT-VBR connections are guaranteed a higher proportion of bandwidth compared to the UBR connections

Example of NRT-VBR and UBR traffic over per-VC and common queues



CQC discard priority

Discard priority is identical for all function processor types. For information on how discard priorities apply to queues and free lists, see Discard priority: all function processors (page 21).

However, interaction between discard priority and emission priority is dependent on the function processor type. This interaction is described in the next section, Interaction between emission and discard priorities (page 39).

Copyright © 2005, Nortel
Nortel Confidential

Interaction between emission and discard priorities

The overview of queuing and traffic scheduling in Overview of queuing and traffic scheduling (page 8) describes the interaction between emission and discard priorities. The interactions described in that chapter are common to all ATM function processors.

The following sections provide specifics on emission and discard priority interactions. For CQC-based function processors, this interaction is known as service category mapping and port aggregation.

Service category mapping priorities for CQC

For CQC-based function processors, Nortel Multiservice Switch nodes use the same mapping for service category to emission and discard priority as used for the PQC on ATM IP function processors.

CQC uses service category mapping to three emission priorities and four discard priorities.

The figure Default service category mapping to priorities: CQC-based function processors (page 40) summarizes the default mappings. For information on service category mapping, see the section on ATM IP PQC service category mapping in NN10600-705 *Nortel Multiservice Switch 7400/15000/20000 ATM Traffic Management Fundamentals*.

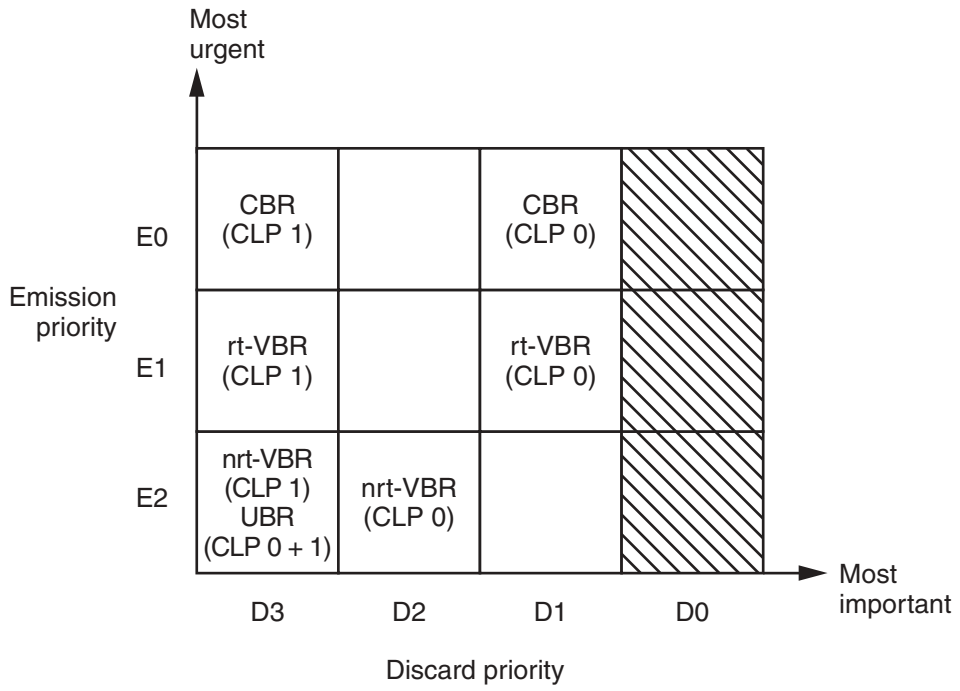
Where traffic from NRT-VBR and UBR service categories map to emission priority 3, the values for the following parameters must be the same for each service category:

- transmit queue limit
- minimum per-VC queue limit
- reference rate

These values are subject to a semantic check, in which the values for the NRT-VBR parameters must apply to both service categories.

Copyright © 2005, Nortel Nortel Confidential

Default service category mapping to priorities: CQC-based function processors



Note: Emission priority is not affected by the CLP (emission priority is determined by the connection's atmServiceCategory attribute).

Reserved for control traffic

PPT 2638 001 AA

CQC port aggregation

CQC-based function processors have an interaction between emission and discard priorities associated with a port based on the aggregate congestion state of the free list. Carriers use port aggregation to meet service level agreements under congestion conditions (peak bandwidth periods). You can disable or enable port aggregation on CQC-based function processors. Port aggregation is enabled by default.

This technique uses the congestion state of the most congested common queue for the transmit link and applies that state as the congestion state of the port. The node calculates an overall congestion state for the port resource by taking the maximum congestion value of all of the individual queues associated with the resource.

Copyright © 2005, Nortel
Nortel Confidential

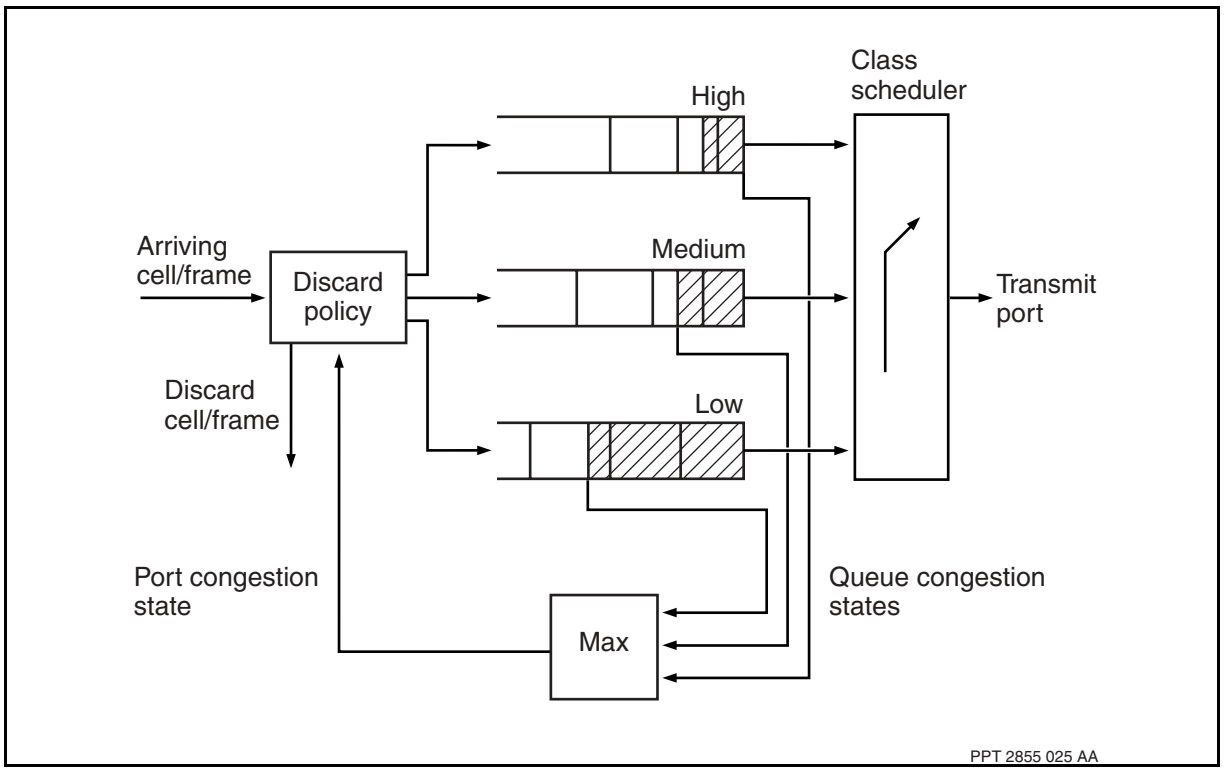
Using the aggregate congestion state, the discard policy proceeds as follows:

- As the function processor processes the cell for the port resource, it checks the aggregated congestion state of the free list.
- If the discard priority of the arriving cell is higher than the aggregated congestion state (for example, cell discard priority is 2 and the aggregated congestion state is congestion control level 3 [CC3]), then the node admits the cell to the queues.
- If the discard priority of the arriving cell is equal to or lower than the aggregated congestion state (for example, cell discard priority is 2 and the aggregated congestion is at congestion control level 2 (CC2), then the node discards the cell.

Before admitting a frame or cell to any of the queues belonging to a port resource, its discard priority must be higher (more important) than the overall link congestion state as defined by the free list.

The figure Port aggregation on CQC-based function processors (page 41) illustrates how the Nortel Multiservice Switch node uses the aggregated congestion state to determine discard policy.

Port aggregation on CQC-based function processors



Copyright © 2005, Nortel
Nortel Confidential

As cells and frames arrive to be queued, the node checks the discard priority against the congestion state of the transmit link. The node discards cells or frames with a discard priority that is greater than the congestion state of the link. See the table Acceptance and discard by link congestion state (page 42).

Acceptance and discard by link congestion state

	Discard priority of cell/frame			
Link congestion state	**Internal control traffic (discard priority = 0)**	**Most important traffic (discard priority = 1)**	**Important traffic (discard priority = 2)**	**Least important traffic (discard priority = 3)**
0	accept (see [1])	discard	discard	discard
1	accept	accept	discard	discard
2	accept	accept	accept	discard
3	accept	accept	accept	accept
[1] The node accepts traffic as long as it does not exceed the lengths of the low and medium queues.				

The node verifies the free list congestion state in a similar manner when you allocate memory to buffer cell or frames that the queues receive from either the bus or the link.

Free list and queue configurations

For link transmit queues, the Nortel Multiservice Switch node discards less important traffic under the following conditions:

- resources for processor, memory, bus, or link are severely congested
- a function processor or link failure has occurred

Nortel Multiservice Switch node CQC-based function processors apply techniques that estimate link transmit queue length and the discard priority on per-VC queues. The following subsections provide details on these techniques as applied to link transmit queues.

Free list lengths and congestion thresholds

The application of congestion thresholds on free list lengths is identical to the general application described in Free lists (page 11) and Free list discard priorities (page 23).

Copyright © 2005, Nortel Nortel Confidential

CQC queue limits and congestion thresholds

Queue limits and congestion thresholds apply to the CPU receive queue, bus transmit queue, and link transmit queues. The following tables summarize limits and thresholds:

- Default common queue thresholds (in cell blocks) (page 44) summarizes default common queue thresholds for each of the queue cell limits defined in the table Default common queue limit (in cell blocks) (page 44).
- Default medium priority per-VC queue limits and thresholds (in cell blocks) for low-speed function processors (page 45) summarizes default medium priority per-VC queue limits and thresholds (in cell blocks) for low-speed function processors (DS1/E1/IMA).
- Default medium priority per-VC queue limits and thresholds (in cell blocks) for high-speed function processors (page 46) summarizes default medium priority per-VC queue limits and thresholds (in cell blocks) for high-speed function processors (OC-3/DS3/E3/JT2).
- Default low priority per-VC queue limit and thresholds (in cell blocks) for low-speed function processors (page 47) summarizes default low priority per-VC queue limits and thresholds (in cell blocks) for low-speed function processors.
- Default low priority per-VC queue limit and thresholds (in cell blocks) for high-speed function processors (page 48) summarizes default low priority per-VC queue limits and thresholds (in cell blocks) for high-speed function processors.

For transmit queues only, the common and per-VC queue limits are configurable. For configuration details, see NN10600-710 *Nortel Multiservice Switch 7400/15000/20000 ATM Configuration Management*.

The queue limit represents the maximum number of cell blocks that the function processor can hold in each queue. When the queue reaches its limit, further cells received are discarded until the queue level falls below the limit.

Each threshold value represents the number of blocks queued before causing the queue enters a more severe congestion state. The first congestion indication is when the queue is 35 percent full, the second indication when it is 75 percent full and the third when it is 90 percent full.

Copyright © 2005, Nortel Nortel Confidential

Default common queue limit (in cell blocks)

Queue type	Low priority	Medium priority	High priority
CPU receive	2240	480	96
bus transmit	1600	480	96
link transmit - high-speed function processors (OC-3/DS3/E3/JT2)	2304	480	96
link transmit - low-speed function processors (DS1/E1/IMA)	1792	288	96

Default common queue thresholds (in cell blocks)

Queue limit	CC0 (~90 percent)	CC1 (~75 percent)	CC2 (~35 percent)
2304	2304	1792	960
2240	2048	1792	960
1792	1664	1408	768
1600	1536	1280	704
480	448	384	208
288	288	224	120
96	88	72	40
CC = congestion control level			

The following tables define the queue limits and thresholds for the per-VC queues for low- and high-speed function processors.

- Default medium priority per-VC queue limits and thresholds (in cell blocks) for low-speed function processors (page 45)
- Default medium priority per-VC queue limits and thresholds (in cell blocks) for high-speed function processors (page 46)
- Default low priority per-VC queue limit and thresholds (in cell blocks) for low-speed function processors (page 47)
- Default low priority per-VC queue limit and thresholds (in cell blocks) for high-speed function processors (page 48)

Copyright © 2005, Nortel Nortel Confidential

These tables identify queues according to the corresponding shaping rate. For shaping stack rates, see the table on the defined traffic shaping rates for CQC-based function processors (available rates per shaped port) in NN10600-706 *Nortel Multiservice Switch 7400/15000/20000 ATM Traffic Shaping and Policing Fundamentals*.

Default medium priority per-VC queue limits and thresholds (in cell blocks) for low-speed function processors

Cell rate (cells/s)	Queue limit	CC0 (~90%)	CC1(~75%)	CC2 (~35%)
943 396	288	256	224	104
666 667	288	256	224	104
471 698	288	256	224	104
333 333	288	256	224	104
235 849	288	256	224	104
166 667	288	256	224	104
117 924	288	256	224	104
83 333	288	256	224	104
58 962	288	256	224	104
41 667	288	256	224	104
29 481	288	256	224	104
20 833	288	256	224	104
14 740	288	256	224	104
10 416	288	256	224	104
7370	288	256	224	104
5208	288	256	224	104
3685	288	256	224	104
2604	208	192	160	72
1842	144	128	112	52
1302	104	96	72	36
921	88	80	64	30
651	88	80	64	30
460	88	80	64	30
325	88	80	64	30
230	88	80	64	30

(1 of 2)

Copyright © 2005, Nortel
Nortel Confidential

Default medium priority per-VC queue limits and thresholds (in cell blocks) for low-speed function processors (continued)

Cell rate (cells/s)	Queue limit	CC0 (~90%)	CC1(~75%)	CC2 (~35%)
163	88	80	64	30
115	88	80	64	30
82	88	80	64	30
CC = congestion control level				
(2 of 2)				

Default medium priority per-VC queue limits and thresholds (in cell blocks) for high-speed function processors

Cell rate (cells/s)	Queue limit	CC0 (~90%)	CC1 (~75%)	CC2 (~35%)
943 396	480	448	384	176
666 667	480	448	384	176
471 698	480	448	384	176
333 333	480	448	384	176
235 849	480	448	384	176
166 667	480	448	384	176
117 924	480	448	384	176
83 333	480	448	384	176
58 962	480	448	384	176
41 667	480	448	384	176
29 481	480	448	384	176
20 833	480	448	384	176
14 740	480	448	384	176
10 416	352	320	256	120
7370	240	224	192	88
5208	176	160	128	60
3685	120	112	96	44
2604	88	80	64	30
1842	88	80	64	30
1302	88	80	64	30
(1 of 2)				

Copyright © 2005, Nortel
Nortel Confidential

Default medium priority per-VC queue limits and thresholds (in cell blocks) for high-speed function processors (continued)

Cell rate (cells/s)	Queue limit	CC0 (~90%)	CC1 (~75%)	CC2 (~35%)
921	88	80	64	30
651	88	80	64	30
460	88	80	64	30
325	88	80	64	30
230	88	80	64	30
163	88	80	64	30
115	88	80	64	30
82	88	80	64	30
CC = congestion control level				
(2 of 2)				

Default low priority per-VC queue limit and thresholds (in cell blocks) for low-speed function processors

Cell rate (cells/s)	Queue limit	CC0 (~90%)	CC1 (~75%)	CC2 (~35%)
943 396	1792	1664	1408	640
666 667	1792	1664	1408	640
471 698	1792	1664	1408	640
333 333	1792	1664	1408	640
235 849	1792	1664	1408	640
166 667	1792	1664	1408	640
117 924	1792	1664	1408	640
83 333	1792	1664	1408	640
58 962	1792	1664	1408	640
41 667	1792	1664	1408	640
29 481	1792	1664	1408	640
20 833	1792	1664	1408	640
14 740	1792	1664	1408	640
10 416	1792	1664	1408	640
7370	1792	1664	1408	640
(1 of 2)				

Copyright © 2005, Nortel
Nortel Confidential

Default low priority per-VC queue limit and thresholds (in cell blocks) for low-speed function processors (continued)

Cell rate (cells/s)	Queue limit	CC0 (~90%)	CC1 (~75%)	CC2 (~35%)
5208	1792	1664	1408	640
3685	1792	1664	1408	640
2604	1280	1152	960	448
1842	896	832	704	320
1302	640	576	480	224
921	448	416	352	160
651	320	288	240	112
460	224	208	176	80
325	160	144	120	56
230	112	104	88	40
163	88	80	64	30
115	88	80	64	30
82	88	80	64	30
CC = congestion control level				
(2 of 2)				

Default low priority per-VC queue limit and thresholds (in cell blocks) for high-speed function processors

Cell rate (cell/s)	Queue limit	CC0 (~90%)	CC1 (~75%)	CC2 (~35%)
943396	2304	2048	1792	832
666667	2304	2048	1792	832
471 698	2304	2048	1792	832
333 333	2304	2048	1792	832
235 849	2304	2048	1792	832
166 667	2304	2048	1792	832
117 924	2304	2048	1792	832
83 333	2304	2048	1792	832
58 962	2304	2048	1792	832
41 667	2304	2048	1792	832
(1 of 2)				

Copyright © 2005, Nortel
Nortel Confidential

Default low priority per-VC queue limit and thresholds (in cell blocks) for high-speed function processors (continued)

Cell rate (cell/s)	Queue limit	CC0 (~90%)	CC1 (~75%)	CC2 (~35%)
29 481	2304	2048	1792	832
20 833	2304	2048	1792	832
14 740	2304	2048	1792	832
10 416	1664	1536	1280	576
7370	1152	1024	896	416
5208	832	768	640	288
3685	576	512	448	208
2604	416	384	320	144
1842	288	256	224	104
1302	208	192	160	72
921	144	128	112	52
651	104	96	80	36
460	88	80	64	30
325	88	80	64	30
230	88	80	64	30
163	88	80	64	30
115	88	80	64	30
82	88	80	64	30
CC = congestion control level				
(2 of 2)				

CQC queuing and scheduling for VP termination

CQC-based function processors support basic VPTs only. A basic VPT VCC has the same queuing and scheduling requirements as an independent VCC.

Copyright © 2005, Nortel Nortel Confidential

Queuing and scheduling on ATM IP FPs

This chapter describes the Nortel Multiservice Switch implementation for ATM IP function processors:

Navigation

- Overview of ATM IP queuing and scheduling (page 50)
- ATM IP emission priorities (page 51)
- Traffic scheduling on the AQM (page 53)
- Class scheduling (page 54)
- ATM IP connection scheduling (page 62)
- Discard priorities on ATM IP function processors (page 66)
- Interaction between emission and discard priorities (page 66)
- Priority interactions and minimum bandwidth guarantee (page 69)
- Port aggregation on ATM IP function processors (page 71)
- Emulation of port aggregation congestion management (page 75)
- ATM IP queue limits and discard thresholds (page 76)
- ATM IP queuing and scheduling for basic VP termination (page 79)
- ATM IP queuing and scheduling for standard VP termination (page 79)

Overview of ATM IP queuing and scheduling

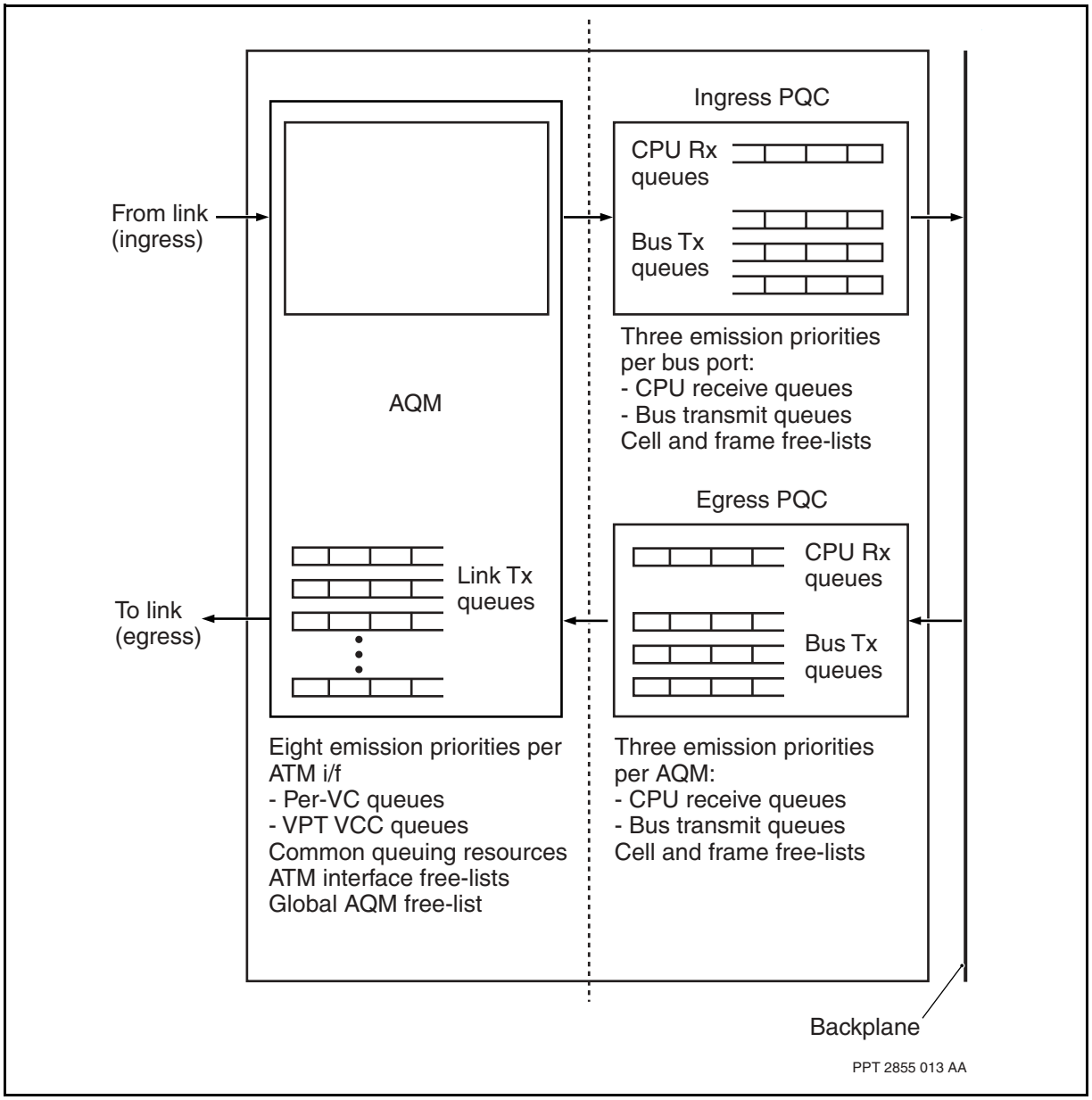
The figure Queuing and scheduling resources on ATM IP function processors (page 51) summarizes the resources required on the Nortel Multiservice Switch node's queue controller (PQC) and the ATM queue manager (AQM).

For information on virtual path terminators (VPT) and associated VCCs, see ATM IP queuing and scheduling for standard VP termination (page 79).

Copyright © 2005, Nortel

Nortel Confidential

Queuing and scheduling resources on ATM IP function processors



ATM IP emission priorities

On ATM IP function processors, the number of emission priorities available depends on the point at which queuing occurs. The PQC queuing architecture differs from the AQM.

PQC emission priorities

PQC emission priorities have the following characteristics:

- three emission priorities: high, medium, and low

Copyright © 2005, Nortel
Nortel Confidential

- all priorities have common queuing only
- free lists for both cell and frame buffers

The basic structure of PQC emission priorities is identical to the structure of CQC emission priorities in the CQC-based function processor. The difference between the two queue controllers is in the application of the discard policy in the context of emission priority. Discard policy for PQC is discussed later in this chapter.

For information on the CQC emission priorities, see CQC emission priorities (page 33). For information on CQC discard policy, see CQC discard priority (page 38).

On a properly engineered node, congestion on the PQC is not expected. When discussing ATM IP queuing and scheduling in this document, the emphasis is on the AQM, which supports the link transmit queues.

AQM emission priorities

AQM emission priorities have the following characteristics:

- eight emission priorities, numbered 0 through 7
- there are two premium emission priorities (priorities 0 and 1) that have absolute priority with minimum delay and cell delay variance (CDV)
- there are six regular emission priorities that have an optional minimum bandwidth guarantee (MBG) for starvation avoidance
- any two emission priorities out of the eight available can support traffic shaping
- up to six emission priorities configured for weighted fair queuing (WFQ)
- per-VC queuing is available on all emission priorities and may be shaped
- common queuing is available on all emission priorities and is unshaped

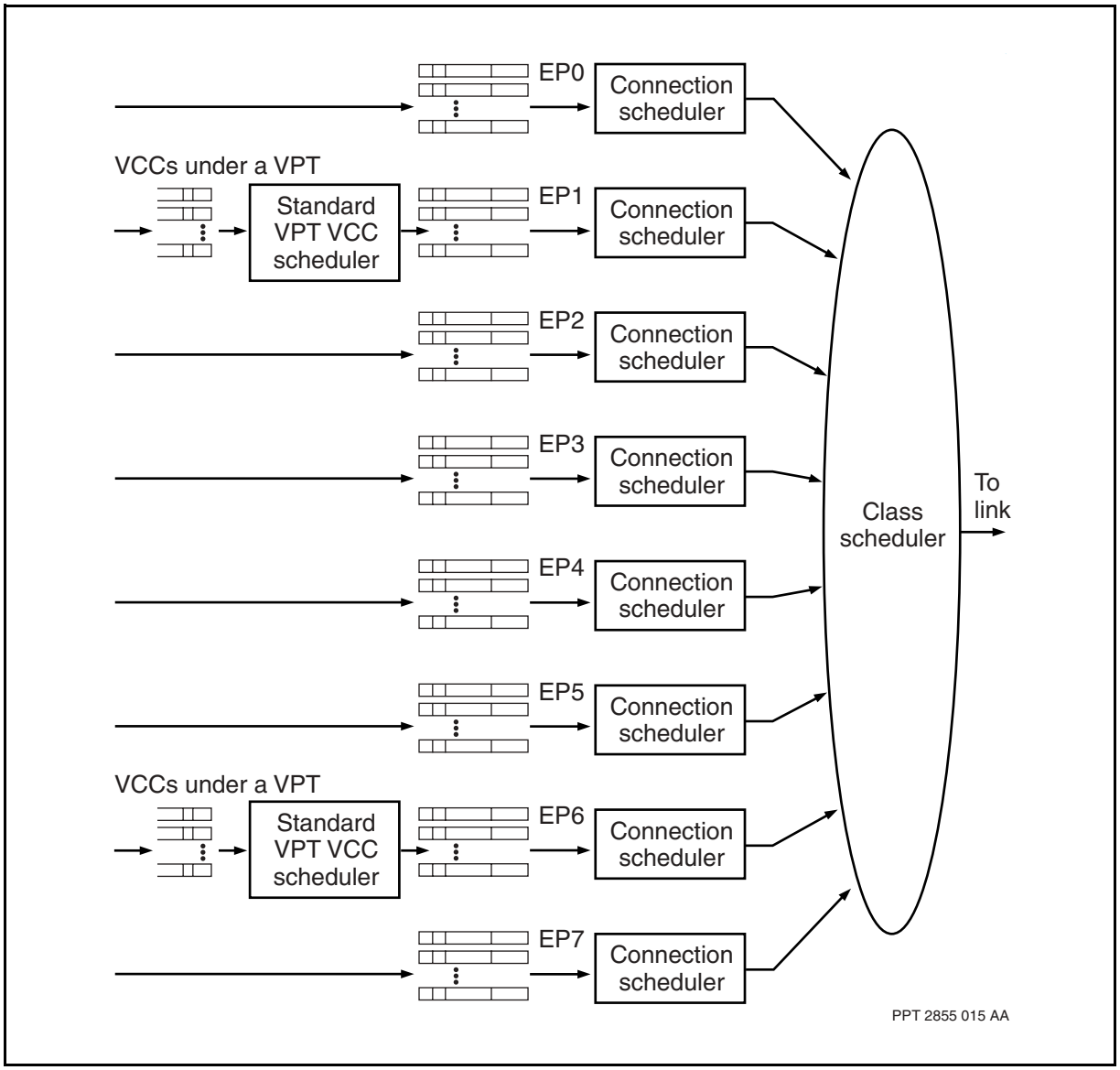
The figure AQM emission priorities and schedulers on ATM IP function processors (page 53) summarizes how the emission priorities work on ATM IP function processors.

Each link has eight emission priorities. Through configuration, you associate a service category with an emission priority as a way of controlling the emission priority of that service category relative to other categories. Therefore, the traffic from all connections defined for a specific service category is directed to a single emission priority. Also, one emission priority may support two or more service categories.

This larger number of available levels provides more flexibility in terms of which service category maps to which emission priority.

Copyright © 2005, Nortel Nortel Confidential

AQM emission priorities and schedulers on ATM IP function processors



Traffic scheduling on the AQM

The AQM applies scheduling algorithms at three levels:

- class scheduling, which coordinates cell transmit opportunities within an emission priority to a link and distributes these opportunities among all emission priorities that share the link
- connection scheduling, which applies to the individual virtual circuit or virtual path connections within an emission priority
- standard VPT VCC scheduling, which applies to the VCCs under a standard VPT

Copyright © 2005, Nortel
Nortel Confidential

The figure AQM emission priorities and schedulers on ATM IP function processors (page 53) shows the relationships between these scheduling levels.

The following sections describe the class and connection schedulers. Note that the standard VPT VCC scheduler is a special case that applies to connections over standard virtual path (VP) terminations. For information on VP terminations and standard VPT VCC scheduling, see ATM IP queuing and scheduling for standard VP termination (page 79).

Nortel Multiservice Switch nodes apply scheduling and traffic shaping mechanisms on transmit queues. On ATM IP function processors, these mechanisms have the following operating characteristics:

- The class scheduler operates according to the MBG discipline which, when configured, provides MBGs for low priority traffic.
- The connection scheduler operates on principal connections within an emission priority according to a weighted fair queuing (WFQ) method. See Weighted fair queuing for ATM IP FPs (page 62).
- The traffic shaping controls connection traffic using a shaped fair queue (SFQ) method. See the section on traffic shaping for ATM IP function processors in NN10600-706 *Nortel Multiservice Switch 7400/15000/20000 ATM Traffic Shaping and Policing Fundamentals*.
- The standard VPT VCC scheduler operates on VCCs under a VPT according to the WFQ method.

ATM IP function processors can use either linear shaping or VBR shaping to the connections within an emission priority.

The following sections provide information on priority and connection scheduling. For information on standard VPT VCC scheduling, see ATM IP queuing and scheduling for standard VP termination (page 79).

Class scheduling

The class scheduler arbitrates service opportunities among emission priorities. At each cell opportunity, the class scheduler evaluates the eligibility of all priorities using a set of prioritizing rules and selects which priority to serve. See ATM IP emission priorities (page 51) for a description of emission priorities on ATM IP function processors.

Configurable characteristics of the class scheduler are:

- ATM service category to emission priority mapping (levels 0 to 7)
- shaping option (linear shaping, VBR shaping, or non-shaping) for each category

Copyright © 2005, Nortel Nortel Confidential

- MBG for each of emission priorities 2 to 7
- choice of per-VC or common queuing

Two emission priorities can have either linear or VBR shaping. The scheduler serves emission priorities 2-7 according to the minimum bandwidth guarantee discipline, with a static priority order from 2 to 7.

Minimum bandwidth guarantee

Minimum bandwidth guarantee (MBG) assures that the node assigns a minimal share of link bandwidth to lower priority service categories such as UBR or NRT-VBR, even when transmit congestion occurs. For ATM IP function processors, MBG applies to AQM queues only.

MBG temporarily raises the emission priority of a lower priority queue so that it is above all other queues in the MBG range (emission priorities 2 through 7). In this way, the node guarantees a minimum service to queues with lower emission priorities. From a service category perspective, given that you map each service category to a separate queue, MBG permits the node to temporarily service traffic from a lower service category over traffic from a higher category.

You can configure MBG for each of the non-premium emission priorities (2 through 7), although typically it is set for only one or two of these priorities. The MBG range is 0 to 50 per cent of the link bandwidth after the absolute priorities are factored in, and the total of all MBG values cannot exceed 100 percent of this remaining link bandwidth.

The MBG is the percentage of the bandwidth that remains after all traffic on emission priority 0 and 1 are served. For example, if traffic on ep 0 and ep 1 consume 20% of link bandwidth, the MBG is applied to the remaining 80% of the link bandwidth.

The actual MBG value can differ from the provisioned value as a result of the MBG calculation. The following factors are included in the calculation:

- The guaranteed bandwidth is calculated from resources that are not being used for ep 0 and ep1. If ep 0 and ep1 use 50% of the available bandwidth and the MBG is set to 20% then approximately 20% of the remaining 50% would be guaranteed.
- The MBG uses a series of predetermined values based on a denominator of 64. The actual value being used is the multiple of 1/64 closest to and greater than the provisioned MBG value. For examples of provisioned MBG values and the corresponding actual values, see the table Examples of provisioned and actual MBG values (page 56).

Copyright © 2005, Nortel
Nortel Confidential

Examples of provisioned and actual MBG values

Provisioned MBG value	Actual MBG value as a multiple of 1/64	Actual MBG value
1%	1/64	1.56%
2% or 3%	2/64	3.13%
4%	3/64	4.69%
5% or 6%	4/64	6.25%
7%	5/64	7.18%

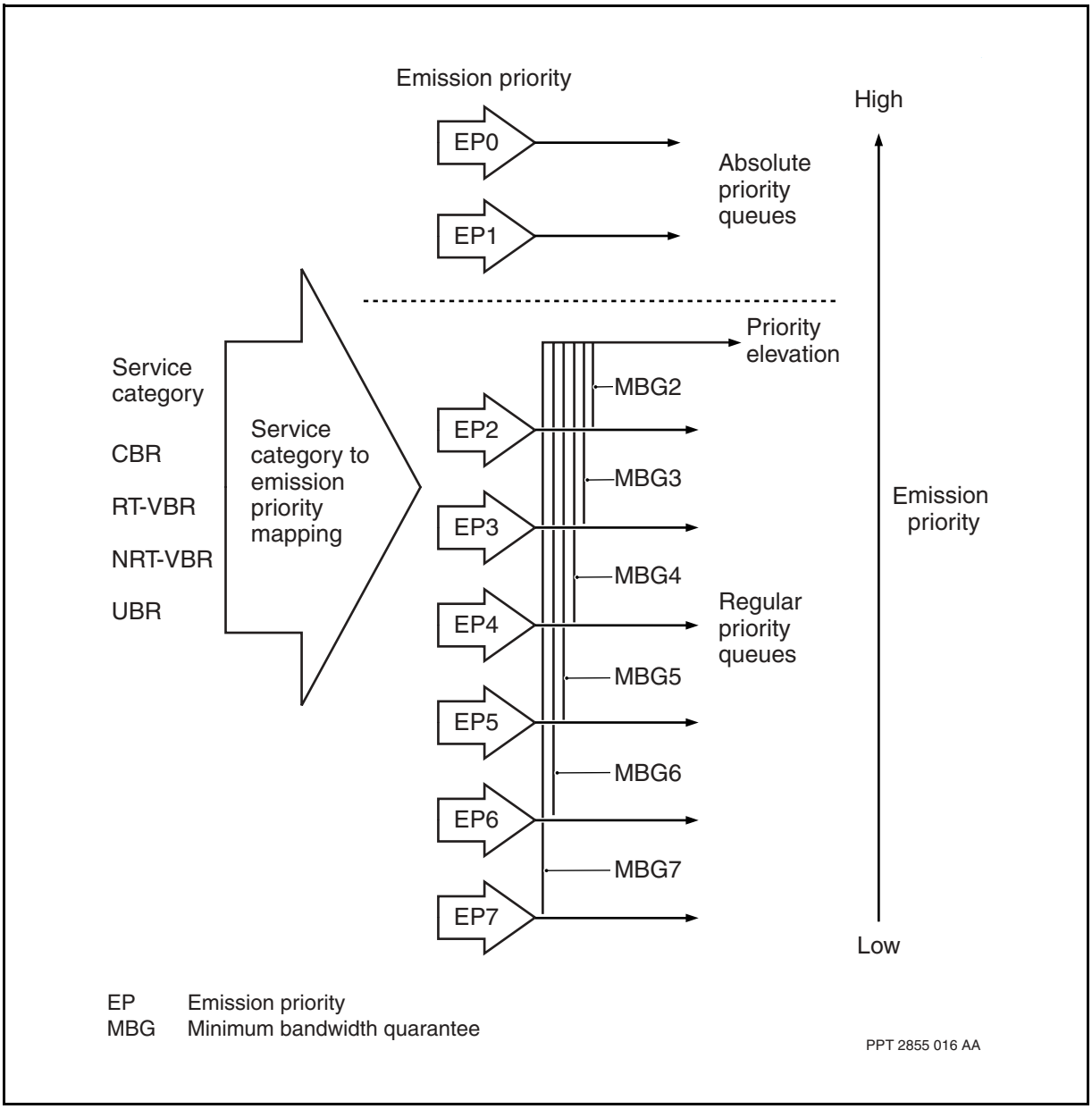
MBG has the following operating characteristics:

- Emission priorities 0 and 1 are always served ahead of any emission priority with MBG.
- If MBG is not configured, all emission priorities are served in absolute priority within the MBG range of emission priorities.
- If an emission priority has a non-zero MBG but no traffic, the cell opportunities revert back to the other priority queues.

The figure Minimum bandwidth guarantee on ATM IP function processors (page 57) illustrates how MBG applies.

Copyright © 2005, Nortel Nortel Confidential

Minimum bandwidth guarantee on ATM IP function processors



Minimum bandwidth guarantee compared to bandwidth pools

It is important to differentiate between guaranteed bandwidth allowed by the scheduler and the pool capacity for bandwidth reservation.

Pool capacity is a bandwidth management and reservation tool that limits the maximum amount of port bandwidth that the node admits for each priority. In CAC, bandwidth pools ensure that the sum of the ECRs of connections belonging to a particular service category does not exceed pool capacity,

Copyright © 2005, Nortel
Nortel Confidential

adjusted by the configured overbooking factor. For example, a small bandwidth pool may be needed for the CBR service category to limit the amount of video calls in a network.

Giving a larger minimum bandwidth guarantee to a class (as compared to its bandwidth pool reservation), offers advantages in cell delay and cell delay variation.

For more information on bandwidth pools, see the bandwidth pool management section in NN10600-708 *Nortel Multiservice Switch 7400/ 15000/20000 ATM CAC and Bandwidth Fundamentals*.

Typical configurations for class scheduling

This section shows a set of basic configuration examples. Two examples describe how to configure the ATM IP function processor to operate in CQC compatible emission priorities (shaping and non-shaping). Also, two examples describe MBG applications using minimum bandwidth guarantee; there are two examples for FR-ATM compatibility, and one example for PORS compatibility.

The table Example of CQC-compatible service category mapping (page 58) shows CQC compatible examples. These examples apply to cases when migrating from CQC to ATM IP function processors.

Example of CQC-compatible service category mapping

ATM service category	Non-shaping (default)	Shaping (example)	CQC emission priority (as a reference point only)
CBR	EP0	EP0 (not shaped)	high
RT-VBR	EP1	EP1 (VBR shaper)	medium
NRT-VBR	EP4	EP5 (VBR shaper)	low
UBR	EP7	EP5 (VBR shaper)	low
The default MBGs for emission priorities in the MBG range is “priority” which indicates that there is no bandwidth guarantee other than that provided by its emission priority.			

The table Minimum bandwidth guarantee mapping example (page 59) shows a configuration example for starvation-free, MBG operations. This example assumes that

- CBR and RT-VBR service categories are well-engineered for the link speed (link bandwidth and speed are sufficient for expected traffic load)

Copyright © 2005, Nortel Nortel Confidential

- the objective of the MBG is to ensure starvation-free service to Nortel Multiservice Switch trunks over ATM which support the NRT-VBR service category.

Multiservice Switch node software guarantees starvation-free service even if CBR and RT-VBR traffic temporarily (or permanently) exceeds the predicted bandwidth.

Minimum bandwidth guarantee mapping example

ATM service category	Non-shaping (example)		Shaping (example)	
CBR	EP2	MBG2 = priority	EP2	MBG2 = priority
RT-VBR	EP3	MBG3 = priority		
NRT-VBR	EP4	MBG4 = 2%	EP4	MBG4 = 2%
UBR	EP7	MBG7 = priority		
Priority indicates that there is no bandwidth guarantee other than that provided by its emission priority.				

The key points illustrated in the table Example of CQC-compatible service category mapping (page 58) and Minimum bandwidth guarantee mapping example (page 59):

- Each service category has an associated emission priority level
- Each service category may be optionally shaped (even CBR)
- two emission priorities can be shaped; shaped service categories must map to one of these two emission priorities
- Each of emission priorities 2 through 7 has an associated MBG, but only the lower emission priorities need to specify a bandwidth guarantee.

For interworking scenarios where either frame relay to ATM, or Multiservice Switch Path-Oriented Routing System (PORS) to ATM conversions are required, frame relay transport priorities and PORS emission priorities are mapped to ATM service categories. After the mapping, the service categories are assigned to the desired emission priority. The table FR-ATM transport priority to ATM service category default mapping (page 60) presents the default frame relay transport priorities to ATM service category mapping.

Copyright © 2005, Nortel Nortel Confidential

FR-ATM transport priority to ATM service category default mapping

Frame relay transport priority	ATM service category	
	without CBR (see Note)	with CBR
TP11	RT-VBR	CBR
TP9	RT-VBR	RT-VBR
TP6	NRT-VBR	NRT-VBR
TP0	UBR	UBR
Some service providers do not offer CBR in the network as a matter of network policy.		

For more details about FR-ATM interworking, see NN10600-920 *Nortel Multiservice Switch 7400/15000/20000 Operations: Frame Relay to ATM Interworking*.

The table PORS emission priority to ATM service category default mapping (page 60) shows the mapping between PORS mapped mode emission priorities and ATM service categories.

PORS emission priority to ATM service category default mapping

PORS emission priority	ATM service category	CQC emission priority (for reference only)
0 (high)	CBR	high
1 (medium)	RT-VBR	medium
2 (low)	NRT-VBR	low

PORS PLC configuration selects a PORS emission priority (0 through 2) which maps to an ATM service category as defined in the table PORS emission priority to ATM service category default mapping (page 60). The ATM interface connection administration then defines the mapping of ATM service category to ATM emission priority (0 through 7). The ATM interface connection administrator also allows other parameters to be configured, such as common or per-VC queuing, and queue limits.

When selecting the PORS emission priority, consider the other traffic which may be mapped to that service category (for example, ATM bearer service, ATM MPE, logical trunks, or FR-ATM).

Copyright © 2005, Nortel Nortel Confidential

Customized configuration for class scheduling

If the examples in the tables Example of CQC-compatible service category mapping (page 58), FR-ATM transport priority to ATM service category default mapping (page 60), and PORS emission priority to ATM service category default mapping (page 60) do not match specific needs, operators can customize the scheduler's configuration. For example, it is possible to shape only some selected service categories. Also, it is possible to assign CBR to an absolute priority, but assign RT-VBR to the MBG range.

The following rules are to be observed during the configuring process, and they are enforced by semantic checks. These rules ensure QOS requirements are met and hardware configuration guidelines are followed:

- permitted emission priority order: CBR ≤RT-VBR ≤NRT-VBR ≤UBR
- at most two emission priorities can be shaped
- CBR and RT-VBR may share the VBR shaper
- MBG applies to emission priority 2 through 7
- the sum of MBGs for emission priority 2 through 7 cannot exceed 100 percent

The following are general configuration guidelines:

- High emission priority levels should be assigned to services which require minimum delay and low loss. For example, CBR and RT-VBR should use emission priorities 0 through 3. Lower emission priority levels should be assigned to services with lower QOS requirement (for example, UBR should use emission priority 7).
- MBG specifies only the absolute minimum required for starvation avoidance. As such, the MBGs typically total to a small percentage of the available link bandwidth.
- MBGs are for real time allocation of bandwidth. This is distinct from the bandwidth allocation performed by connection admission control (CAC).
- MBG is only required for lower emission priorities. For example, a guarantee for emission priority 2 does not perform any useful function. Emission priority 2 is normally above all other emission priorities in the MBG range and can never take precedence over emission priority 0 or 1.
- Whenever possible, unassigned emission priority levels should leave room for future expansion of new service categories. For example, even though ABR is not currently deployed, leave room for it in the emission priority hierarchy so that it can be added without disrupting NRT-VBR and UBR services.

Copyright © 2005, Nortel
Nortel Confidential

ATM IP connection scheduling

Connection scheduling arbitrates service opportunities among connections within an emission priority. Once a service opportunity is available for an emission priority, the connection scheduler evaluates the service eligibility for each connection within that priority according to a set of rules, and serves the connection that is eligible for the opportunity. Nortel Multiservice Switch ATM node IP function processors use one of two different connection scheduling techniques:

- weighted fair queuing (WFQ), for unshaped emission priorities
- shaped fair queuing (SFQ) for shaped emission priorities

The remainder of this section describes WFQ. For a description of SFQ, see the section on traffic shaping for ATM IP function processors in NN10600-706 *Nortel Multiservice Switch 7400/15000/20000 ATM Traffic Shaping and Policing Fundamentals*.

Weighted fair queuing for ATM IP FPs

The WFQ connection scheduler provides service to connections with weights proportional to their bandwidth demands. It schedules departure order among connections so that when a connection does not have a cell to send to a cell slot, the opportunity is offered to other connections (work-conserving).

The weighted fair queuing (WFQ) connection scheduler is based on a set of per-VC queues that the node services through a single link-class queue. ATM IP function processors offer per-VC queuing for all links and service categories (including CBR). The node serves each per-VC queue according to its WFQ weight, and enqueues cells into the link-class queue for that emission priority.

The WFQ connection scheduler has the following characteristics:

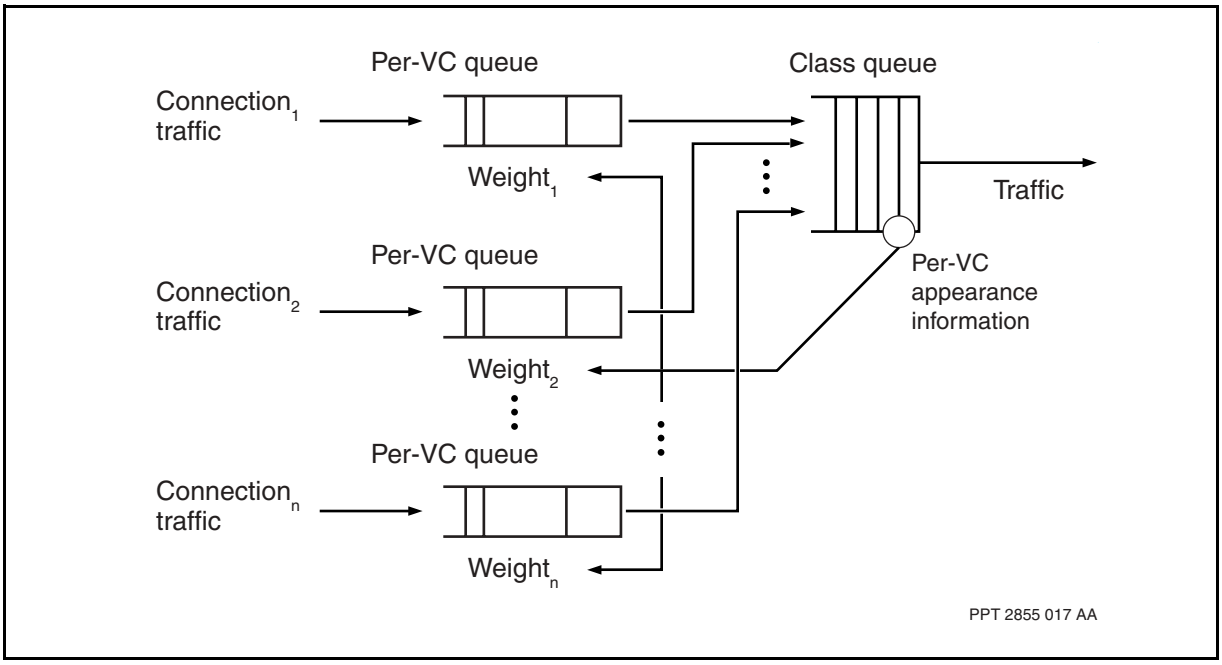
- weighted fairness of bandwidth among connections, optionally based on one of the following:
  - — equivalent cell rate (ECR) derived from traffic contract through CAC
  - — peak cell rate (PCR)
  - — sustained cell rate (SCR)
- per-VC queuing to isolate well-behaved connections from the effects of misbehaving connections
- lower average delay and CDV for well-behaved connections than a weighted round robin (WRR) scheme
- work-conserving dynamic allocation of bandwidth, where bandwidth unused by one connection is assigned to other connections

Copyright © 2005, Nortel Nortel Confidential

The weight of a WFQ connection determines how many cells in the per-VC queue that the link-class queue can service. If the number of cells eligible for service is less than the number of cells in the per-VC queue, the remaining cells are left in the per-VC queue. As a result, the relative weight among all connections can impose some fairness between these connections.

The figure Multiservice Switch node weighted fair queuing connection scheduler (page 63) illustrates how the WFQ connection scheduler works.

Multiservice Switch node weighted fair queuing connection scheduler



For each service category, the service provider can choose the basis for connection weight for each connection within a service category is based on ECR, PCR, or SCR. The defaults are

- ECR for the CBR, RT-VBR, UBR, and NRT-VBR service categories.

When selecting fairness based on ECR, the weight for a connection is proportional to the ECR that is calculated by Nortel Multiservice Switch node CAC. When selecting PCR- or SCR-based fairness, the weight for a connection is proportional to the PCR or SCR traffic parameter.

The default weight policy for UBR connections is set to the equivalent cell rate (ECR) which is equal to the bandwidth allocated to the UBR connection. Therefore, the fairness weight for unshaped UBR connections is directly proportional to the bandwidth allocated to the connection. The weight determines transmit opportunities given to a UBR connection by the weighted

Copyright © 2005, Nortel
Nortel Confidential

fair queueing (WFQ) scheduler on ATM IP FPs which provide more transmit opportunities to UBR with MDCR connections than regular UBR connections which have a default weight of 1.

There is also the option to override the weight on a per-VC basis by configuring at the VCC or VPC level. The weight is in the range of 1 through 4095. The table ECR to weight mapping (ATM IP OC3 function processors) (page 64) gives some sample ECR to weight mappings. The values in this table are calculated using the following formula:

$$CW = \frac{\text{weight cell rate} * \text{Min}(4095, \text{per-VC queue}}{\text{link rate}}$$

where

CW	is the connection weight in cells
weight cell rate	is a value in cells/s for the connection, based on the weight policy option (that is, ECR, PCR, SCR) that is applicable for the connection weight cell rate
per-VC queue length	is the per-VC queue length in cells, based on the ATM service category for the connection

ECR to weight mapping (ATM IP OC3 function processors)

ECR (cell/s)	WFQ weight
353 207	4095
176 603	2047
96 000	1113
80 000	927
58 962	683
23 584	273
11 792	136
4,711	54
3685	42

Copyright © 2005, Nortel
Nortel Confidential

Weighted fair queuing and common queuing

Weighted fair queuing (WFQ) simulates common queuing by setting an infinite weight for each per-VC queue. If common queuing is required, all connections must have infinite weight. That is, per-VC queuing is not available for that emission priority. WFQ supports common queuing for CBR and RT-VBR connections and first-in first-out (FIFO) queuing for NRT-VBR and UBR connections.

Common queuing for CBR and RT-VBR traffic has these characteristics:

- uses a cell buffer resource exclusively available for real-time common queuing
- queue limit is set at 96 cells for CBR and 480 for RT-VBR (defined when the queue limit is set to autoconfigure through configuration)
- connection weight is infinite
- CBR connections map emission priorities 0 or 2 and RT-VBR connection map to emission priorities 1 or 3
- the operational transmit queue length at the connection displays the length of the common queue for this service category

FIFO queuing for NRT-VBR and UBR connections has these characteristics:

- uses the ATM interface free list (no dedicated resource)
- queue limit is set to the free list size (defined by software and cannot be redefined through configuration)
- connection weight is infinite (defined through configuration as up to the queue limit)
- all connections map to emission priority 4 or lower
- the operational transmit queue length at the connection displays the number of cells enqueued for this connection

The implications for servicing NRT-VBR connections in a FIFO queuing configuration are:

- because connection weight is set to infinite, the incoming cells in each connection are immediately queued in the class queue for the link (since weight is infinite, there are no limits)
- over time, a connection may use up a significant portion of the class queue for the link (there is no fairness imposed by the relative connection weights that determine the maximum cell buffer usage for a per-VC queue that is serviced by the class queue)

Copyright © 2005, Nortel
Nortel Confidential

Per-VC and common queuing in non-port aggregation configurations

RT-VBR traffic does not draw from the free list if the connections are configured for common queuing. That is, each RT-VBR connection has a dedicated buffer and is not influenced by congestion in the free list.

Attention: CBR traffic is not affected by the absence or implementation of port aggregation.

However, if RT-VBR connections are configured for per-VC queuing, and NRT-VBR and UBR connections are configured for FIFO, then all connections are drawing from the free list. In this scenario, the node can discard NRT-VBR or UBR traffic in the presence of RT-VBR CLP1 traffic. Enabling port aggregation prevents the enqueuing of RT-VBR CLP1 traffic and allows some bandwidth to NRT-VBR or UBR traffic connections in preference to low priority RT-VBR traffic.

For information on port aggregation and emulation of port aggregation, see Port aggregation on ATM IP function processors (page 71) and Emulation of port aggregation congestion management (page 75).

Discard priorities on ATM IP function processors

Discard thresholds are set by software and cannot be re-configured. For information on how discard priorities apply to queues and free lists, see Overview of queuing and traffic scheduling (page 8).

Interaction between emission and discard priorities

In ATM IP function processors, emission priority levels and congestion control levels apply for each service category.

In general, interactions as described in Overview of queuing and traffic scheduling (page 8) apply to ATM IP function processors. However, ATM IP function processors also offer a MBG feature for starvation avoidance. The following sections provide information on this feature.

For ATM IP function processors, Nortel Multiservice Switch nodes use the ATM service category of the connection to determine the virtual connection (VC) traffic default mapping for discard and emission priorities. The PQC and the AQM have different mapping requirements, as described in Service category mapping to priorities: ATM IP PQC (page 67). For general information on service category mapping to priorities, see Overview of queuing and traffic scheduling (page 8).

Copyright © 2005, Nortel Nortel Confidential

Service category mapping to priorities: ATM IP PQC

For PQC-based function processors, Nortel Multiservice Switch nodes use the same mapping for service category to emission and discard priority as used for the CQC. See Interaction between emission and discard priorities (page 39).

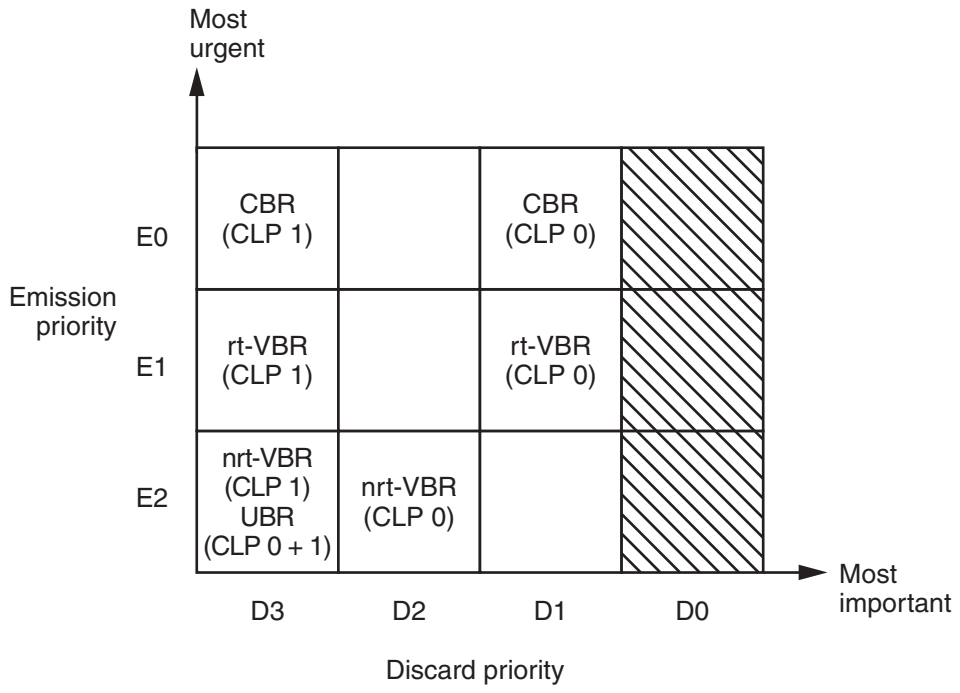
Service category mapping to priorities: ATM IP AQM

The AQM uses service category mapping to any of eight emission priorities and four discard priorities. The figure Default service category mapping to priorities: AQM on ATM IP function processors (page 69) shows the AQM mapping relationship between the ATM service categories and the Nortel Multiservice Switch system of emission and discard priorities. This figure shows the default mappings. As a result, the queue limits specified for each of NRT-VBR, or UBR can be independent of the queue limits for other priorities. Overrides can be configured at the connection level.

Note that the AQM has more emission priorities than currently available service categories. The extra emission priorities can accommodate new service categories that standards bodies may develop in future.

Copyright © 2005, Nortel
Nortel Confidential

Default service category mapping to priorities: PQC on ATM IP function processors



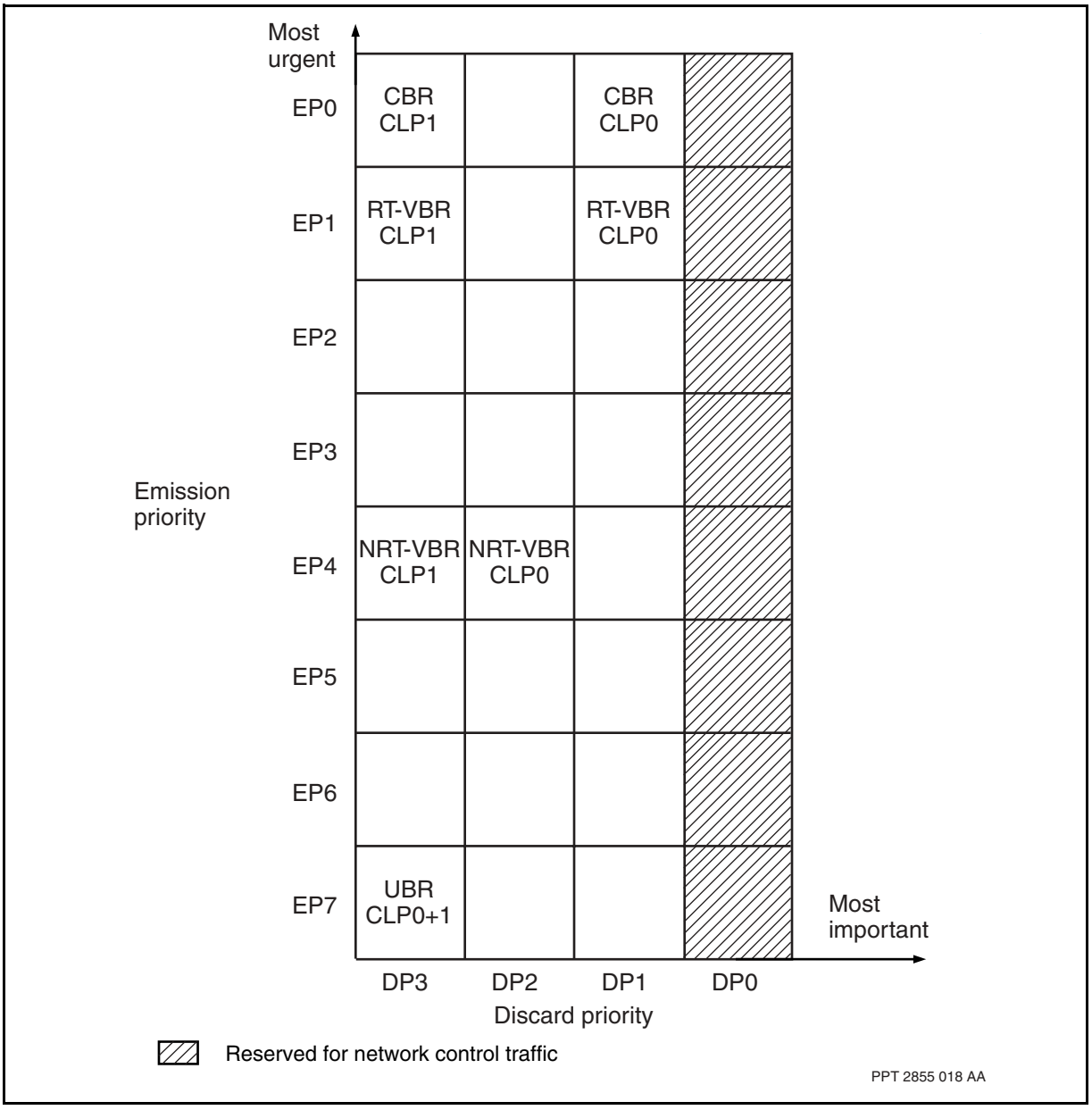
Note: Emission priority is not affected by the CLP (emission priority is determined by the connection's atmServiceCategory attribute).

Reserved for control traffic

PPT 2638 001 AA

Copyright © 2005, Nortel
Nortel Confidential

Default service category mapping to priorities: AQM on ATM IP function processors



Priority interactions and minimum bandwidth guarantee

Elevating the emission priority of a lower priority queue through MBG indirectly affects how discards occur across all queues. Because MBG increases the service rate of an emission priority, it reduces the number of discards on connections that use the emission priority.

Copyright © 2005, Nortel
Nortel Confidential

When evaluating the extent of discards, there is a close interaction between two factors:

- the presence or absence of MBG functionality
- the presence or absence of cell free list congestion

The following sections describe the effects on discards for the following scenarios:

- Priority interaction without MBG and no free list congestion (page 70)
- Priority interaction under free list congestion (page 70)
- Priority interaction with MBG and no free list congestion (page 70)

Priority interaction without MBG and no free list congestion

If none of the emission priorities have MBG active and there is no free list congestion, the node transmits cells from each service category according to defined emission priorities. Since there is no free list congestion, there is no discard in the free list (although there may still be discard on one or more queues).

The amount of traffic for a given service category is limited by the following traffic characteristics on the function processor:

- queue length
- arrival pattern of incoming cells
- transmit opportunities remaining from higher emission priority traffic

There is no interaction between queues when any queue is congested (greater than 35% full).

Priority interaction under free list congestion

Free list congestion is independent of MBG. As the cell free list reaches each congestion control level in turn, the node begins to discard all traffic for that level. The discarded cells therefore never reach the transmit queues. The discard priority applies regardless of the emission priority (is independent of the emission priority). Free list congestion forces discard priority to take precedence.

Priority interaction with MBG and no free list congestion

With the application of MBGs, the node can override the strict segregation of emission priorities.

Under a non-MBG priority configuration, the node transmits CBR traffic before traffic for lower emission priorities. For example, if a five percent MBG is active for a lower emission priority and the port or link is fully saturated with CBR

Copyright © 2005, Nortel

traffic at emission priority 2, the node services the lower priority traffic at a higher emission priority than CBR for five percent of cell transmit opportunities. In this way, MBG elevates emission priority.

In this example, CBR is configured in the minimum bandwidth guarantee range (emission priorities 2 to 7). As a result, CBR traffic may experience congestion and CDV impact.

Port aggregation on ATM IP function processors

Port aggregation is a congestion response control. Port aggregation is a technique through which NRT-VBR CLP0 and UBR traffic can accumulate in the buffers until the node begins to discard RT-VBR CLP1 traffic. The objectives are:

- to cause free list congestion at the CC3 level or higher
- to configure RT-VBR traffic to use the free list for cell buffering, thereby resulting in RT-VBR CLP1 cell discard

The free list is either the global free list on OC3 ATM IP function processors cards or the link pool free list on DS3 and E3 ATM IP function processors.

For port aggregation to have the required results, free list congestion must be possible through a single NRT-VBR or UBR connection. For this reason, the per-VC queue limit for any connection must be equal to the free list size. This queue limit is automatically set by software when port aggregation is enabled. The queue limit cannot be changed except by changing the size of the free list. Also, software sets the thresholds for the queue congestion control levels for NRT-VBR and UBR connections to those used for the free list (75%, 80%, and 90%).

Enabling port aggregation on ATM IP function processors involves a critical change through configuration. Where critical change is undesirable for a node, emulation of the port aggregation feature is a preferred alternative. See Emulation of port aggregation congestion management (page 75) for information.

To enable port aggregation on ATM IP function processors, the service provider configures the characteristics described in the following sections:

- Per-VC queuing for RT-VBR traffic (page 73)
- Traffic allocation to service categories (page 74)
- Queue configuration for NRT-VBR and UBR (page 74)
- NRT-VBR and UBR allocated to unshaped queues (page 74)

Copyright © 2005, Nortel Nortel Confidential

For more information on congestion control thresholds and packet-wise discard levels, see Characteristics of ATM IP packet-wise discard (page 138) and Interaction between congestion control and packet-wise discard levels (page 72).

Interaction between congestion control and packet-wise discard levels

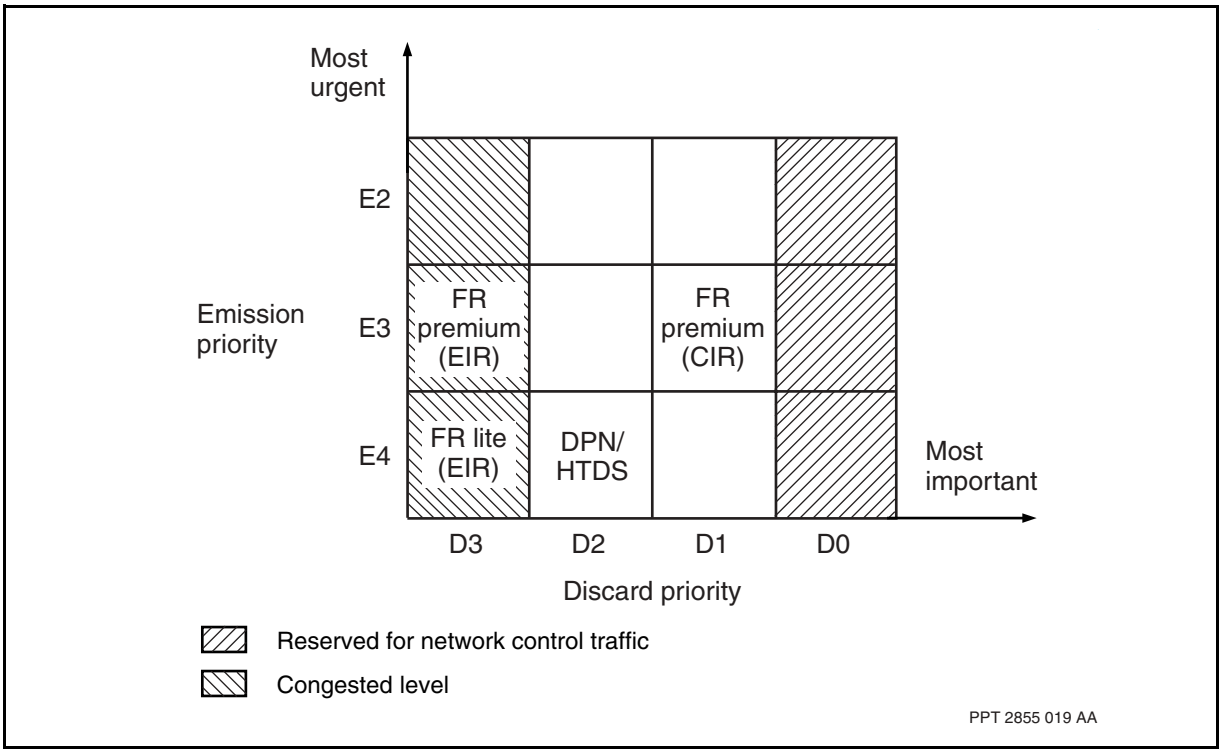
The thresholds for connection queues and the free list interact to ensure that cell discard occurs in the most effective manner possible. Under port aggregation configurations, cell discard occurs with the following characteristics:

- Cell discard due to EPD at congestion control (CC) level 2 can occur even though there is no congestion at CC level 3. For example, in the figure Congestion management example (page 73), EPD discards of DPN/HTDS traffic occur before complete discard of frame relay excess information rate (EIR) traffic.
- Cell discard due to EPD at congestion control (CC) level 1 can occur even though there is no congestion at CC level 2. For example, in the figure Congestion management example (page 73), EPD discards of frame relay committed information rate (CIR) traffic occur before complete discard of DPN/HTDS traffic.
- RT-VBR (CLP0) traffic discards can occur due to EPD even though congestion has not yet reached the CC level 2 threshold.
- Per-VC CBR traffic discards can occur due to EPD even though congestion has not yet reached the CC level 1 threshold.

The table Summary of free list congestion and discard thresholds as percentages of queue length (page 73) lists the percentages for the congestion, PPD, and EPD thresholds.

Copyright © 2005, Nortel
Nortel Confidential

Congestion management example



Summary of free list congestion and discard thresholds as percentages of queue length

Congestion control level	CC level threshold as a percentage of free list length	Free list PPD as a percentage of free list length	Free list EPD as a percentage of free list length
CC0	100	99.0	95.0
CC1	90	89.1	85.5
CC2	80	79.2	76.0
CC3	75	74.3	71.25
Percentages are approximate values.			

Per-VC queuing for RT-VBR traffic

RT-VBR traffic must be configured for per-VC queuing (not common queuing). Software enforces this configuration through semantic checks. If RT-VBR traffic is configured for common queuing, the node allocates a dedicated buffer in the AQM for queuing all RT-VBR cells. As a result, the free list does not affect this dedicated buffer and RT-VBR is then immune to free list congestion.

Copyright © 2005, Nortel
Nortel Confidential

Traffic allocation to service categories

Service category traffic must be allocated to emission priorities as summarized in the table Port aggregation: allowable mapping of service categories to emission priorities (page 74).

NRT-VBR connections must have expanded queue limits and adjusted thresholds. The congestion control level thresholds apply to a given emission priority across all links on the AQM. For example, on a DS3 or E3 ATM IP function processor, emission priority 4 on all links has thresholds 75%, 80%, and 90%. Any connections at emission priorities 4 through 7 have an expanded queue limit and adjusted thresholds. Any connections at emission priorities 0 through 3 use the queue limit and thresholds defined for the service category. This condition ensures that CBR and RT-VBR connections conform to the requirements for a bounded queue delay.

Port aggregation: allowable mapping of service categories to emission priorities

Service Categories	Emission priorities							
	0	1	2	3	4	5	6	7
CBR	Yes	No	Yes	No	No	No	No	No
RT-VBR	No	Yes	No	Yes	No	No	No	No
NRT-VBR	No	No	No	No	Yes	Yes	Yes	Yes
UBR	No	No	No	No	Yes	Yes	Yes	Yes

Queue configuration for NRT-VBR and UBR

Connections under the NRT-VBR and UBR service categories can be configured for either per-VC or common queuing. If configured for common queuing, each connection must have an infinite weight. If configured for per-VC queuing, the node uses the WFQ algorithm to calculate the weight for each connection. In both queuing configurations, each connection has a queue limit that is equal to the free list size and uses the same thresholds as the free list (75%, 80%, and 90%).

NRT-VBR and UBR allocated to unshaped queues

Connections under the NRT-VBR and UBR service categories cannot be shaped. However, note that semantic checks do not enforce this criteria. As a result, if these connections are shaped, unwanted behavior may occur. If a connection has shaping enabled, port aggregation adjustments are not applied to that connection. This condition means that the connection has the queue limit and thresholds as determined by the SFQ queue limit adjustment algorithm.

Copyright © 2005, Nortel
Nortel Confidential

Configuring port aggregation

Port aggregation is configured at the function processor level, under the ATM resource control for AQM 0. When port aggregation is enabled, software automatically configures the queue limit size and congestion control levels for all NRT-VBR and UBR connections on the function processor. These settings are described in the section titled Port aggregation on ATM IP function processors (page 71).

When port aggregation is disabled, software sets the queue limit according to the common or per-VC queuing configuration, and the thresholds are set at the normal queue thresholds (35%, 75%, and 90%).

Lastly, regardless of the setting for port aggregation, the queue limits and congestion control levels are consistent with settings on CQC-based function processors.

Emulation of port aggregation congestion management

Although port aggregation is a preferred method of congestion control, it does involve a critical change to the ATM IP function processor. If service disruption due to critical change is not desirable, ATM IP function processors can emulate port aggregation.

Through port aggregation emulation, you can manage congestion through either MBG or the free list congestion state. This discussion refers to the figure Congestion management example (page 73) to show how this occurs. In this example, the objective is to ensure that DPN/HTDS traffic is not discarded before frame relay premium excess information rate (EIR) traffic.

Congestion management through MBG

You can apply MBG as a way to manage congestion on the function processor. To use this approach, you configure an appropriate value for MBG relative to peak bandwidth conditions under which congestion occurs. In this example, MBG ensures that DPN and HTDS traffic (at emission priority 4) receives appropriate access to the traffic scheduler.

This approach does not provide a sweeping reduction in congestion, but does ensure that the node services some of the traffic on lower priority queues. As a result, these lower priority queues can accept more cells, thereby reducing discards.

Congestion management using the free list state

This approach uses the overall congestion level of the free list to determine discard.

Copyright © 2005, Nortel
Nortel Confidential

By configuring the queue for NRT-VBR or UBR to be equal to the ATM interface free list, you achieve traffic differentiation between discard priority 3 and discard priorities 0, 1, and 2 traffic independent of emission priority. Under peak bandwidth conditions, buffered EIR traffic causes cell free list depletion. Congestion control level 3 and 2 cell free list state transition results in discard of discard priority 3 cells and provides serving capacity to higher discard priority traffic.

ATM IP queue limits and discard thresholds

On ATM IP function processors, there are different configurations for queues for the PQC and AQM.

PQC queue limits and thresholds

The PQC supports common queuing only. This characteristic is due to the operations that the PQC and the AQM undertake, and the relationship between them. Normally, there is only minimal depletion of the PQC-based queues. There is no traffic management configuration required for the PQC CQM.

The queue limit for all queues on the PQC (CPU, bus transmit, and link transmit) is 3072 cells. The congestion control levels are summarized in the table PQC congestion control levels (page 76).

PQC congestion control levels

Congestion control level	Threshold (in cells)
CC0	3072
CC1	2048
CC2	1792
CC3	960
These thresholds apply to all queues on the PQC: CPU receive, bus transmit, and link transmit.	

The ingress PQC supports CPU and bus transmit queuing. The egress PQC performs AAL5 frame segmentation and cell forwarding. Nominal queuing is expected for cells and frames during both processing and transmission due to the high speed of the data path that connects the egress PQC to the AQM. Instead, queuing of data is expected inside the AQM where larger cell buffers are available.

Copyright © 2005, Nortel
Nortel Confidential

AQM queue limits and thresholds

ATM IP function processors have the following enhancements over CQC-based function processors:

- can configure different queue limits between NRT-VBR and UBR, since they are served on different emission priorities
- NRT-VBR and UBR default (function processor dependent) queue limits are 10 240 cells (see Expanded default queue limits for NRT-VBR and UBR (page 77))
- CBR has transmit queue limits, since traffic can be shaped or served on per-VC queues
- per-VC queue limits are not tied to the shaping rate (defaults can be overridden through configuration or the auto-configure value)
- total buffer space is split between services using common queues and those using per-VC queues

Where traffic from two or more service categories is mapped to a single emission priority, values for the following parameters do not need to be the same:

- transmit queue limit
- minimum per-VC queue limit
- reference rate

You can set these parameters to suit the requirements of each service category.

For general characteristics of queue limits, see Overview of queuing and traffic scheduling (page 8). For more details on queue limits related to VCCs within a VPT, see ATM IP queuing and scheduling for standard VP termination (page 79).

Expanded default queue limits for NRT-VBR and UBR

The ATM IP function processors have over four times the cell buffer capacity of the CQC-based function processors. For non real-time service categories (NRT-VBR and UBR) the default queue limit has been expanded to take advantage of the extra buffers. This expansion applies to common and per-VC queues.

For ATM IP function processors (OC3, DS3 and E3), the default transmit queue limits for CBR, and RT-VBR service categories are identical to those on CQC-based ATM OC3, DS3 and E3 function processors. This is to ensure consistent default delay characteristics when using ATM IP or CQC-based

Copyright © 2005, Nortel Nortel Confidential

function processors. The table Default queue limits and thresholds (cells) (page 78) shows default transmit queue limit and thresholds for ATM IP function processors.

Default queue limits and thresholds (cells)

Service category	Transmit queue limit (See [1])	Congestion control 0 threshold (~90%)	Congestion control 1 threshold (~75%)	Congestion control 2 threshold (~35%)	Reference rate (cell/s)
CBR	96 cells	86	72	33	65 511
RT-VBR	480 cells	433	360	168	14 740
NRT-VBR	10 240 cells	9240	7680	3600	65 511
UBR	10 240 cells	9216	7680	3584	65 511
NRT-VBR (high-speed CQC-based FP - see [2])	2304 cells	2048	1792	960	14 740
[1] These default values are set when the queue limit is set to autoconfigure for each service category. [2] These values are provided for reference only. Note that the actual operational values for each congestion threshold for CQC-based function processors vary due to hardware granularity.					

When considering the queue limits in the table Default queue limits and thresholds (cells) (page 78), remember that any specific service category can have two relevant congestion control levels for cell relay: one for CLP0 cells and one for CLP0+1 cells. Frame SAR requires four congestion control levels. Note that some applications such as FR-ATM modify the default discard priority setting for transmit traffic.

Common queuing is used for a service category that is configured with the parameter for unshaped transmit queuing set to common. The transmit queue limit is the limit that applies to the common queue for a service category. For NRT-VBR and UBR, the default value for transmit queue limit is set to the link free list size.

For per-VC queuing, 10 240 is the default maximum queue length. To achieve the same delay characteristics, the per-VC queue limit reference rate is 65 511 cell/s. ATM IP function processors offer per-VC queuing for all links and all service categories (including CBR if desired) with no affect on performance or features.

Per-VC queuing is enabled in the following cases:

- SFQ service categories (required)

Copyright © 2005, Nortel Nortel Confidential

- unshaped non-ABR service categories with per-VC queuing enabled (default)

For purposes of calculating the queue limit, unshaped connections are assumed to have a connection cell rate equal to the link rate. This assumption is valid because the shaping rate for those connections is the link rate.

For the relationships between delay, cell rate (shaping rate or link rate), transmit queue limit, minimum per-VC queue limit, and reference rate, see Overview of queuing and traffic scheduling (page 8). The maximum transmit queue length is expanded for ATM IP function processors to take advantage of the larger cell buffers while guaranteeing a constant delay over a larger range of per-VC shaping rates.

ATM IP queuing and scheduling for basic VP termination

The basic VPT VCC has the same queuing and scheduling capabilities as an independent VCC.

For more information on basic VPTs, see NN10600-702 *Nortel Multiservice Switch 7400/15000/20000 ATM Routing and Signalling Fundamentals*.

ATM IP queuing and scheduling for standard VP termination

ATM IP function processors provide advanced queuing and scheduling features that support VPTs. VPTs on ATM IP function processors use hardware scheduling at the VP and VC layers.

For VPTs, this scheduling level is referred to as standard VPT VCC scheduling. See Traffic scheduling on the AQM (page 53) and the figure AQM emission priorities and schedulers on ATM IP function processors (page 53).

For introductory information on standard VPTs, see NN10600-702 *Nortel Multiservice Switch 7400/15000/20000 ATM Routing and Signalling Fundamentals*.

At the VP layer for a VPT, the connection scheduler services cells at the VPT connection point like any other relay-point VPC or independent VCC connection. When the VPT is not shaped, the node applies WFQ to schedule cells. When the VPT is shaped, the node applies SFQ.

Class scheduling handles cell transmit opportunities to a link and distributes the opportunities among the service categories that share the link. The node associates the standard VPT with a service category and consequently with one of the eight available emission priorities. Through configuration, you can associate a VPT with any service category. However, the recommended

Copyright © 2005, Nortel Nortel Confidential

approach is to associate the VPT with the service category that offers at least the QOS that the associated VCCs require. For example, a CBR VCC assigned to a UBR VPT could not deliver proper service to the end user.

Standard VPT VCC queuing and scheduling

Standard VPT VCC queuing and scheduling provides two levels of emission priority:

- real time, which is used for VPT VCCs that have a service category of CBR or RT-VBR
- non-real-time, which is used for VPT VCCs that have a service category of NRT-VBR or UBR

All standard VPTs on a function processor have the following queue structure:

- a variable number of per-VC queues, one for each VPT VCC
- a single per-VPT queue
- a VPT WFQ queue, which contains all cells connections to the sum of the weights for all VPT VCCs

The per-VC queue has a queue limit and thresholds according to the service category of the VCC. The per-VPT queue has a queue limit and thresholds according to the NRT-VBR service category. This limit is reduced by an amount that is proportional to the shaping rate of the VPT.

The per-VPT queue limit must include sufficient space for all cells that are queued for any active VCCs. Cell queuing conforms to threshold criteria in the following order:

1 to the thresholds of the per-VC queue

2 to the thresholds of the per-VPT queue

When the length of either queue reaches a defined threshold, the cell is discarded according to rules for packet-wise discard.

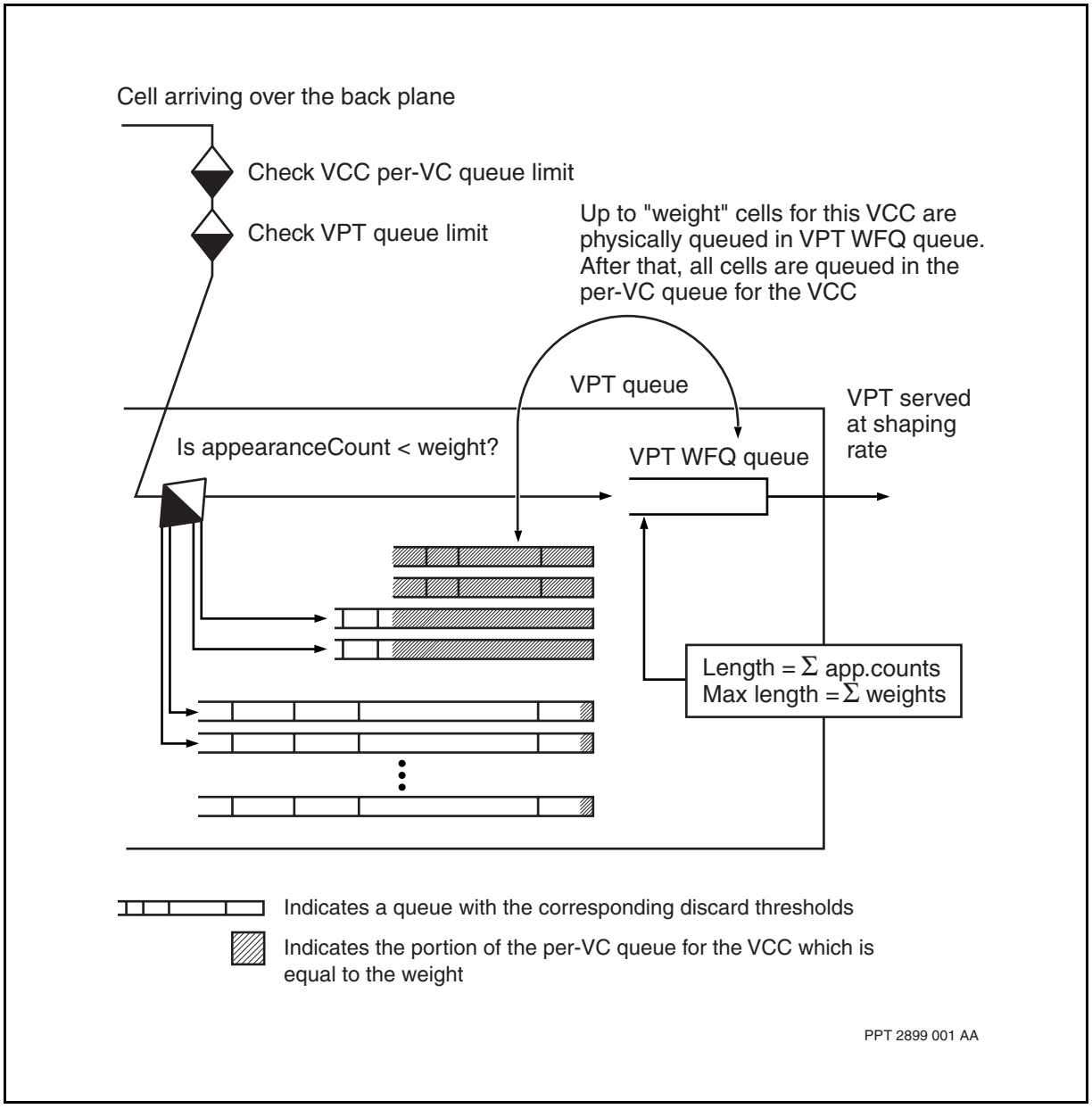
There are no specific queue limits or thresholds for the VPT WFQ queue, either configured or set by software. The number of cells in the VPT WFQ queue never exceeds the sum of the weights for all VCCs under the VPT. The queue length may be less than the sum of weights if some connections are not busy. For each connection, the number of cells in the VPT WFQ queue is referred to as the appearance count. The appearance count is always less than or equal to the connection weight. If the weight for a VCC is equal to the per-VC queue limit, all cells that are eligible for queuing are stored in the VPT WFQ queue (that is, no cells are queued in the per-VC queue). If the weight for a VCC is less than the per-VC queue limit, the node queues eligible

Copyright © 2005, Nortel Nortel Confidential

cells in the VPT WFQ up to a number equal to the connection weight. Additional eligible cells for that connection are stored in the per-VC queue until the appearance count goes down.

The figure Queuing and scheduling in a standard VPT (page 81) shows how the node processes incoming cells in the per-VPT queue.

Queuing and scheduling in a standard VPT



Copyright © 2005, Nortel
Nortel Confidential

When a cell arrives over the back plane for a VPT VCC, processing proceeds as follows:

1 Determine if the cell can be queued:

If VptVccQueueLength + 1 > VptVccThreshold(Clp, serviceCategory) then discard the cell

If VptQueueLength + 1 > VptThreshold(Clp, serviceCategory) then discard the cell

2 Determine where the cell is to be queued:

If the appearance count < connection weight then queue the cell directly to the VPT WFQ queue

Otherwise queue the cell in the per-VC queue for the VCC.

3 Reduce by one the appearance count for the corresponding connection when the scheduler transmits a cell from the front of the VPT WFQ queue. The VPT WFQ queue operates in a FIFO manner; the first cell queued is the first transmitted.

4 Move a cell from the per-VC queue to the VPT WFQ queue (if there are cells in the per-VC queue) and increase the appearance count by one.

Weighting characteristics for VPT VCCs

The node queues cells from all connections directly to the per-VPT queue according to weight. To support the two emission priorities (real-time and non-real-time), standard VPT VCC scheduling uses WFQ. Software automatically configures the weights for the standard VPT VCCs under each service category. The table Default service category weights for per-VPT cell queuing (page 82) shows that real-time VCCs have priority over non-real-time VCCs. The difference in weight for real-time compared to non-real-time connections minimizes the effects of CDV on the real-time VCCs.

Default service category weights for per-VPT cell queuing

Service category	Default weight (see Note)	Default per-VC queue length
CBR	96 cells	96 cells
RT-VBR	96 cells	480 cells
NRT-VBR	1 cell	10 240 cells
UBR	1 cell	10 240 cells
These weights are not configurable at the service category level. Weights can be changed for each connection at the connection level only.		

Copyright © 2005, Nortel Nortel Confidential

Changing these weights for each connection allows you to customize connection priority. For example, if the VPT VCCs are all NRT-VBR, you can configure the weights of the individual VCCs such that preference is given to connections with higher bandwidth requirements. The capability to override the weight for each VCC provides considerable flexibility. This level of configuration requires careful engineering to ensure that each VCC has the correct weight relative to the traffic it supports and the traffic supported by other VCCs.

Queue limits for VPTs

The section ATM IP queue limits and discard thresholds (page 76) provides information on the general characteristics for queue limits on ATM IP function processors. VPTs have additional characteristics, which are described in the following paragraphs.

The per-VPT queue itself operates like any other relay-point VPC or independent VCC connection queue. The per-VPT queue limits the total number of cell buffers allocated to all of the associated VCCs. In this way, the VPT queue functions as a pool, and protects other connections on the link (including other VPTs) from cell free list starvation.

Further, each VCC under the VPT has a per-VC queue limit, regardless of whether or not per-VC queuing applies to the VCC. Each VCC consumes space from the VPT pool up to the queue limit for that VCC. For each real-time VCC, this operating characteristic limits the number of cells that the per-VPT queue can hold for that VCC. For each non-real-time VCC, the queue limit places a bound on the total number of cells that can be present in its per-VC queue and its associated VPT queue. That is, for non-real-time VCCs, the queue limit represents that total number of cells that can be enqueued regardless of queue they are in. The queue limits for each VCC provides protection and fairness among each of the VCCs under the VPT. Lastly, the sum of the limits for all connection queue limits under a VPT can exceed the per-VPT queue limit, thereby permitting overbooking of the VPT pool.

The default limit for the per-VPT queue is based on the queue limit configuration for the NRT-VBR service category for the ATM interface that is associated with the VPT. This characteristic applies to VPTs under per-VC queuing only. If the VPT is served under FIFO queuing, the default limit is the same as the limit for the virtual common queue.

Using the NRT-VBR service category configuration covers most VPT applications, including the following:

- voice and data integration
- multiplexing data VCs only

Copyright © 2005, Nortel
Nortel Confidential

As a result, the service category of the VPT itself does not affect its configured queue limit when using per-VC queuing.

VPT queue limit configuration

Per-VPT queues have the following configuration characteristics:

- The per-VPT queue limit is influenced by the shaping configuration. This influence means that the per-VPT queue limit follows the same queuing delay as opposed to cell rate curve (that is, maximum delay curve) as other connections. The figure General principles of interaction between parameters for defining per-VC queue limits (page 19) shows this curve and the relationships between delay, cell rate (shaping rate), transmit queue limit, minimum per-VC queue limit, and reference rate.
- When the per-VPT queue is not shaped, the cell rate is derived as follows
  - for per-VC queuing, the cell rate is derived from the NRT-VBR transmit queue limit
  - for common queuing (including FIFO), the node derives the cell rate from the service category
- The per-VPT queue can be configured with an absolute limit under the interface on a per-VPT basis for per-VC queuing only.

The per-VC queue limit for standard VPT VCCs have the following configuration characteristics:

- The queue limit is based on the service category queue limit configuration for the VCC service category under the ATM interface. That is, the queue limit is based on values for minimum queue limit, transmit queue limit, and the per-VC queue limit reference rate. This approach results in the per-VC queue limit configuration curve that is shown in the figure General principles of interaction between per-VC queuing parameters: delay over cell rate (page 20).
- For standard VPT VCCs, the cell rate used for calculating the queue limit depends on the application of shaping at the VPT level. When the per-VPT queue is shaped, the VCC cell rate equals the per-VPT queue shaping rate. When the per-VPT queue is not shaped, the VCC queue limit is derived from the service category of the VCC. These characteristics are identical to those for independent VCCs under the ATM interface.
- Queue limits can also be configured with absolute values.
- The per-VC queue limit is capped off at the per-VPT queue limit whenever the calculated or configured queue limit for the VCC is greater than that of the VPT.
- The operator can override the queue limit value.
- Common queuing (including FIFO) does not affect the queue limit for the VCC.

Copyright © 2005, Nortel
Nortel Confidential

Queuing and scheduling on APC/PQC-based FPs

This chapter describes Nortel Multiservice Switch systems implementation for APC/PQC-based function processors (FPs).

Navigation

- Buffer management (page 85)
- Per-VC queuing on APC/PQC-based FPs (page 87)
- Overview of APC schedulers (page 89)
- APC class scheduling (page 90)
- Connection scheduling (page 92)

Buffer management

Conforming cells are monitored by the buffer management mechanism. The cells are buffered with thresholds which are pre-defined according to the connection's ATM service category type and the cells' cell loss priority type.

The goal of buffer (memory) management is to maximize the buffer efficiency through buffer sharing while providing proper isolation to protect individual traffic streams. It involves

- defining the allocation of buffer domains
- using threshold and limit constraints to partition the buffer domains to control the usage of buffer space

On APC/PQC-based function processors, the total buffer space is statically divided into two portions:

- one at the ingress for input-buffering cells from the link towards the backplane
- one at the egress for output-buffering cells from the backplane towards the link

Copyright © 2005, Nortel

Nortel Confidential

For more information, see

- Buffer pool and threshold allocation (page 86)
- User-configurable buffer limits for EPs (page 86)

Buffer pool and threshold allocation

Most of the total buffer space is used for output buffering. For output buffering at an APC device, the buffer space is allocated and managed as three buffer pools and one buffer threshold:

- global buffer pool

  The global buffer pool is used for egress buffering.

- class buffer limits

  The class buffer pool is used for buffer management for the five emission priority classes. It consists of the total buffer space allowed for an emission priority class across all APC devices. The class buffer limits for each APC device is configured through the *Lp Eng Arc Apc Ov bufferLimitPerEP* attribute.

  Class Buffer Limits are per EP per APC. On 4pOC12SmlrAtm there is one APC per link and therefore a link-EP class has the complete access of the EP class buffer limit. On 16pOC3SmlrAtm there is on APC per 4 links and therefore a link-EP class has access to the 1/4 of the EP class buffer limit.

- link-class buffer pool

  The link-class buffer pool is used for buffer management of emission classes within each APC device which constitutes a channel. For each APC device class within the link-class buffer pool, there is a minimum buffer guarantee and a maximum buffer limit.

- per-VC buffer threshold

  The per-VC buffer threshold is used for buffer management for connections within each emission priority class through the per-VC queue CLP1, CLP0+1, EPD and EFCI thresholds.

User-configurable buffer limits for EPs

The class and per-VC buffer limits are user configurable. Users can use the *bufferLimitPerEP* attribute to override the class buffer limits default values based on their QoS requirements.

The table Class buffer limits default values (page 87) shows the default values for APC/PQC-based function processors. Users can use the *bufferLimitPerEP* attribute to override the class buffer limits per EP default values based on their QoS requirements.

Copyright © 2005, Nortel Nortel Confidential

Class buffer limits default values

APC/PQC-based function processor	EP0	EP2	EP3	EP4	EP7
4-port OC12/STM4, 16-port OC3/STM1	39168	39168	130559	39168	13055

For information on the user configurable per-VC buffer pool limits, see Per-VC queue limits (page 87).

Per-VC queuing on APC/PQC-based FPs

APC/PQC-based function processors support only per-VC queuing, also known as per-connection queuing, to serve traffic. Per-VC queuing is a strategy that is used to ensure fairness among all connections associated with a particular ATM service category by providing an individual queue for each shaped and unshaped connection.

For more information, see

- Per-VC queue limits (page 87)
- Per-VC queue thresholds (page 88)

Per-VC queue limits

The per-VC queue limit, *txQueueLimit*, defines the maximum number of cells that can be buffered for a single connection. The default limit is defined on a per ATM service category basis. Alternately, you can configure a *txQueueLimit*, for each VC. See the table Default value of sameAsCa for txQueueLimit (page 87) for the 4-port OC12/STM4 ATM and 16-port OC3/STM1 ATM function processors.

Default value of sameAsCa for txQueueLimit

Function processor	CBR	rt-VBR	nrt-VBR	UBR
4-port OC12/STM4	96	480	10 240	10 240
16-port OC3/STM1	96	480	10 240	10 240

The calculation of the per-VC queue cell loss priority (*CLP0, CLP1*) and early packet discard (EPD) thresholds differs, depending on the value of the *txQueueLimit* attribute. In the case of user-configured *txQueueLimits*, the configured value is used to directly calculate the thresholds. If the default

Copyright © 2005, Nortel Nortel Confidential

value, sameAsCa, is used for *txQueueLimit*, it is then used to calculate the cell loss priority and early packet discard thresholds. These thresholds are functions of the following parameters:

- ATM service category txQueueLimit
- shaping rate of the VC
- minimum per-VC queue limit (*minPerVcQueueLimit*)
- scheduling rate (if not connection-shaped) or shaping rate (if connection-shaped) of the VC
- function processor dependent reference rate (*perVcQueueLimitReferenceRate*)

See the table Queue limit reference rates (page 88) for rates applicable to APC/PQC-based function processors.

Queue limit reference rates

Function processor/port type	**Queue limit reference rate, *perVcQueueLimitReferenceRate* (cells/s)**
4-port OC12/STM4 FP: channelized OC3, DS3, STS1 ports	65 511
4-port OC12/STM4 FP	262 044
16-port OC3/STM FP	65 511

Per-VC queue thresholds

Per-VC thresholds divide the per-VC buffer space into different queue length levels beyond which certain cells are dropped. When congestion occurs and buffer space is being used up, the per-VC thresholds achieve two objectives:

- define a CLP1 threshold that is smaller than a CLP0 threshold to protect the higher priority CLP0 and control cells
- use an Early Packet Discard (EPD) threshold to protect cells belonging to packets already buffered against corruption by the cells of new packets. For information on EPD, see Packet-wise discard for APC-based FPs (page 146).

APC/PQC-based function processors provide static per-VC CLP0, CLP1, and EPD thresholds for different ATM service category classes. For VCs belonging to a particular ATM service category, the thresholds for CLP0 and CLP1 cells are computed by multiplying the per-VC queue limit with the ATM service category-specific congestion control (CC) level. The congestion control levels

Copyright © 2005, Nortel Nortel Confidential

define the congestion state of the queue in term of percentage of the queue that is filled. See the table CLP1 and CLP0 per-Vc thresholds for APC/PQC-based function processors (page 89).

CLP1 and CLP0 per-Vc thresholds for APC/PQC-based function processors

ATM service category	CLP1	CLP0
CBR	38% (CC3)	90% (CC1)
rt-VBR	32% (CC3)	75% (CC1)
nrt_VBR	32% (CC3)	75% (CC2)
UBR	15% (CC3)	35% (CC3)
Thesholds are expressed as a percentage of the queue that is filled. Calculated threshold values may or may not exactly match the same values available in the EP threshold tables. If they do not match, they are approximated internally to the values in the tables that are nearest them.		

Overview of APC schedulers

APC scheduling prioritizes and arbitrates a cell transmission opportunity from the link level at the top of the hierarchy to the class level, and then down to the connection level where the opportunity is assigned to a single eligible connection.

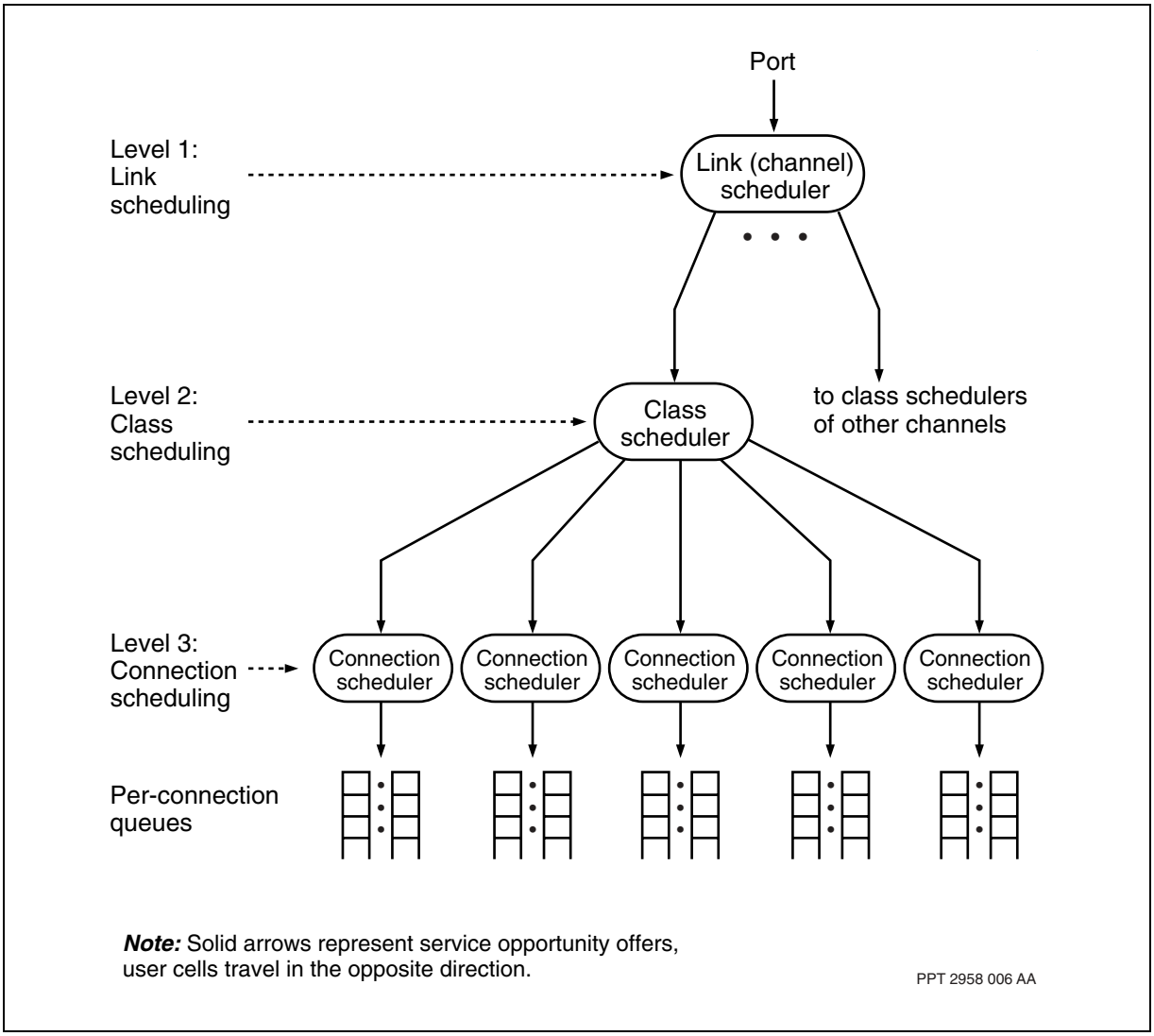
APC/PQC-based function processors three levels of schedulers:

- link scheduler
- class scheduler
- connection scheduler

The figure APC scheduling hierarchy (page 90) shows the relationship between these levels. As the scheduling process continues, each connection scheduler is also monitoring to see whether any of its connections have cells to be served. If there are cells waiting in the per-connection queues, the connection scheduler notifies its next higher-level scheduler, the class scheduler, of its service demand. Similarly, the class scheduler in turn informs its next higher-level scheduler, the sub-port scheduler, of its service demand. In this way, as a service opportunity becomes available, decisions can be made at a high level scheduler as to which lower level service demanding scheduler can be serviced.

Copyright © 2005, Nortel Nortel Confidential

APC scheduling hierarchy



APC class scheduling

In APC/PQC-based function processors, the class (emission) scheduler evaluates the eligibility of all emission priorities (EPs) belonging to a sub-port (ATM interface) using a set of prioritizing rules and selects one EP to serve. It permits flexible mapping from ATM service categories to EPs.

In APC/PQC-based function processors, the class scheduler has five emission priorities (EP0, EP2, EP3, EP4, and EP7). The APC/PQC-based emission priorities have the following characteristics:

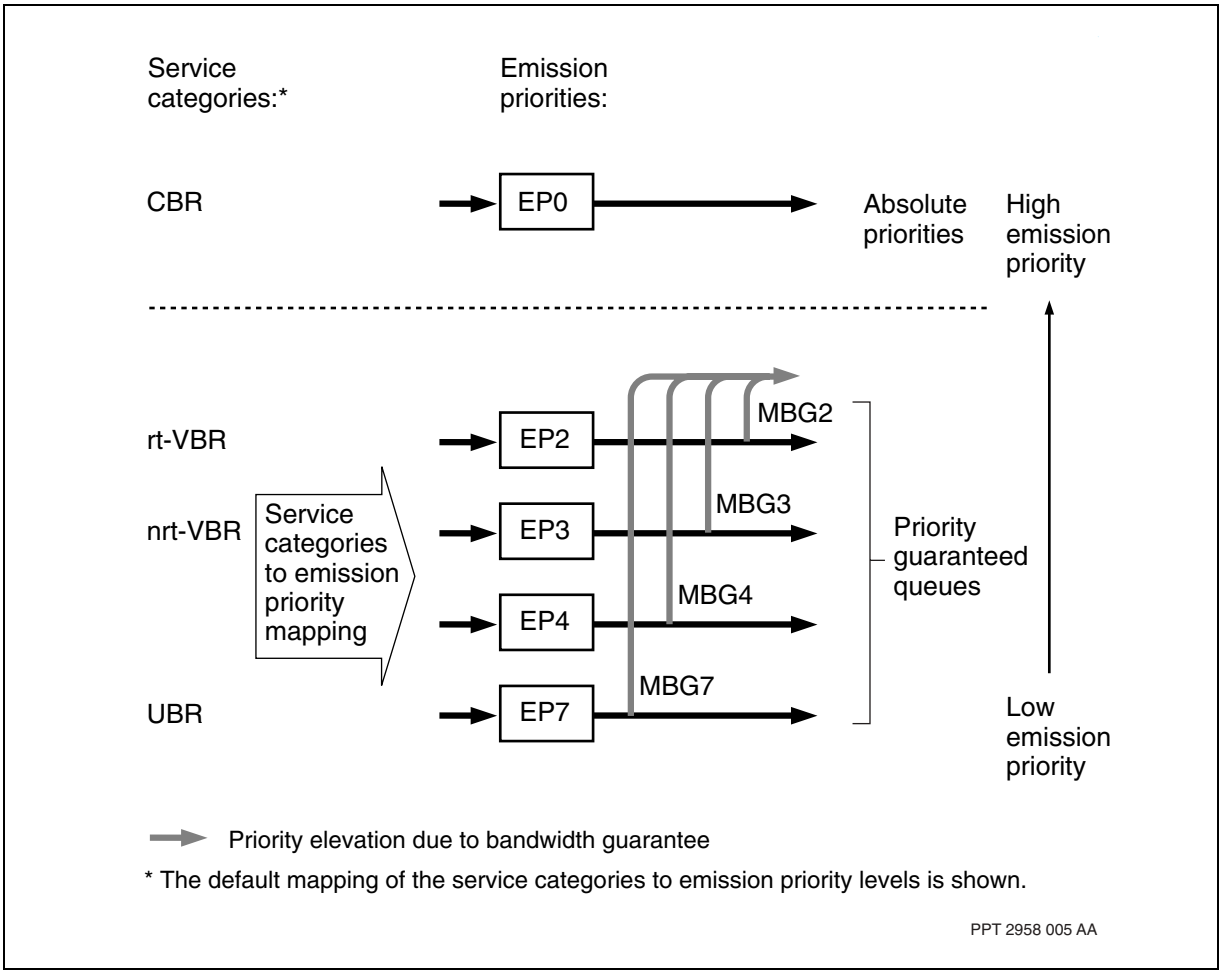
- one premium emission priority (EP0) that has absolute priority with minimum delay and cell delay variance (CDV)

Copyright © 2005, Nortel
Nortel Confidential

- 31 programmable shaping rates for connections that use the premium emission priority (EP0) at each sub-port (ATM interface)
- four regular unshaped emission priorities (EP2, EP3, EP4, and EP7 where EP7 is the lowest priority) that have an optional minimum bandwidth guarantee (MBG) for starvation avoidance

See the figure APC class scheduling (page 91) for a high level view of APC class scheduling.

APC class scheduling



The above figure shows the default mapping of the ATM service categories to the EPs. Nortel Multiservice Switch systems support four ATM service categories, CBR (default mapping of EP0), rt-VBR (default mapping of EP2), nrt-VBR (default mapping of EP3), and UBR (default mapping of EP7). Note that two or more service categories can be configured to map to the same EP level. When needed, multiple ATM service categories are mapped to EP0 so the classes can be shaped.

Copyright © 2005, Nortel
Nortel Confidential

The class scheduler distributes the service opportunities left over from the premium emission priority (EP0) among the regular emission priorities (EP2, EP3, EP4, and EP7) with minimum bandwidth guarantees (MBGs) available for each EP. These MBGs can be configured to zero or non-zero percentage values. A zero percentage value assigned to a class means that no bandwidth guarantee is provided for that class. A non-zero MGB value is the amount of bandwidth guarantee that is desired in terms of the percentage of the link rate.

Minimum bandwidth guarantee

Minimum bandwidth guarantee (MBG) for class schedulers under APC/PQC function processors is similar to that for ATM IP function processors, with some differences. APC/PQC-based function processors have these differences:

- APC/PQC-based function processors allow MBG on emission priorities 0, 2, 3, 4, and 7 only
- the configurable MBG range is 0 to 100 per cent of the link bandwidth after the absolute priorities are factored in

Since the MBGs for the regular emission priories is affected by the bandwidth consumption of EP0, the premium emission priority, the desired MBGs may not be achieved. This would be the case when EP0 consumes a bandwidth of more than the 100% minus the sum of the four MBGs of EP2, EP3, EP4, and EP7, minus the bandwidth consumed by EP0.

Minimum bandwidth guarantee compared to bandwidth pools

The relationship between MBG and bandwidth pools is the same as that for class schedulers under ATM IP function processors. See Minimum bandwidth guarantee compared to bandwidth pools (page 92)

Connection scheduling

In APC/PQC-based function processors, the connection scheduler for each emission priority arbitrates the cell transmission opportunities. For each sub-port (ATM interface), there are five connection schedulers, one for each ATM service category class. Within each connection scheduler for each ATM service category, there are two connection scheduler options:

- rate connection scheduler
- weighted fair queue (WFQ) connection scheduler.

Rate connection scheduler

The rate connection scheduler is a traffic shaper which spaces cell departures from per-Vc queues according to their transmission rates. The rate connection scheduler schedules departure time and as such any unused cell transmission opportunities on a connection is not offered to other connections (non work-conserving).

Copyright © 2005, Nortel Nortel Confidential

Weighted fair queuing

The implementation of WFQ is similar to the implementation on ATM IP function processors. See Weighted fair queuing for ATM IP FPs (page 62).

The principles illustrated by the example shown in the table ECR to weight mapping (ATM IP OC3 function processors) (page 64) are descriptive of sample ECR to weight mappings for the ATM IP OC-3 function processor. These principles also apply to the WFQ implementation on AQC/PQC-based function processors, using the same formula.

Copyright © 2005, Nortel Nortel Confidential

Queuing and scheduling on GQM-based FPs

This chapter describes Nortel Multiservice Switch systems implementation of GQM-based function processors (FPs).

Navigation

- Buffer management on GQM-based FPs (page 94)
- Queuing on GQM-based FPs (page 95)
- Schedulers for GQM-based FPs (page 96)

Buffer management on GQM-based FPs

Conforming cells are monitored by the buffer management mechanism. The cells are buffered with thresholds which are pre-defined according to the connection's ATM service category type and the cells' cell loss priority type.

The goal of buffer (memory) management is to maximize the buffer efficiency through buffer sharing while providing proper isolation to protect individual traffic streams. It involves

- defining the allocation of buffer domains
- using threshold and limit constraints to partition the buffer domains to control the usage of buffer space

The domains used for buffer management and congestion control for GQM-based FPs include:

- global domain
  - — global common buffer pool of 1M cells with thresholds
- link domain for each link
  - — provisionable per-link limits on the usage of the global buffer pool with thresholds
  - — per-link limits are provisioned through the ATM interface attribute *txCellMemory*

Copyright © 2005, Nortel

Nortel Confidential

- class domain
  - — link-class reserved buffer pool to prevent buffer starvation avoidance with no discard thresholds. For the 16pOC3PosAtm FP the cell buffer pool for each link-class is 1 024 K; for the 4pMRPosAtm FP in 4-port OC12 mode, the cell buffer pool for each link-class is 3 456 K; and for the 4pMRPosAtm FP in 1-port OC48 mode, the cell buffer pool for each link-class is 13 824 K.
  - — link-class buffer limit to prevent buffer exhaustion with discard thresholds. For the 16pPC3PosAtm FP, the cell buffer limit for each link-class is 256 K, and for the 4pMRPosAtm FP in both 4-port OC12 and 1-port OC48 the cell buffer limit for each link-class is 884 K.
- queue domain for common queued and per-VC queued connections within each emission priority
  - — provisionable queue limit with discard thresholds
  - — queue limit is provisioned through the attribute *txQueueLimit*

Each domain attempts to use its own buffer pool first. If the pool space runs out, more space may be borrowed from the buffer pool at the next higher aggregation. For example, if a link-class buffer pool runs out, it can borrow buffer space from the global pool provided the link-class limit and per-link limit on the global pool have not been exceeded. This approach prevents any link-class from experiencing complete buffer starvation while still ensuring that the higher priority link-classes get preference for buffers. Since each link-class has a reserved buffer pool, no link-class can ever be completely starved.

Queuing on GQM-based FPs

GQM-based FPs support common and per-VC queuing, also known as per-connection queuing. Choosing a queuing method depends on the kind of scheduler used for an emission priority (EP). When an EP is configured to do shaping or weighted fair queue (WFQ) scheduling, then per-VC queues are used.

Queue thresholds on GQM-based FPs

GQM-based FPs support queue thresholds that are coded different from other ASICs. As a result, the default value of the *txQueueThresholds* attribute may have a value different from other FPs.

Common queuing on GQM-based FPs

Common queuing places cells from multiple connections into a single queue according to a first-in first-out sequence. With the exception of the 4pMRPosAtm FP in 1-port OC48 mode, GQM-based FPs can be provisioned for common queuing for up to 45,000 connections in any ATM service category. When operating in 1-port OC48 mode, a 4pMRPosAtm FP allows up to 16 000 connections configured for common queuing in a service category.

Copyright © 2005, Nortel Nortel Confidential

This is due to the restriction that up to 16 000 connections can be provisioned on an ATM interface, and there can be only one such interface in 1-port OC48 mode.

Per-VC queuing on GQM-based FPs

Per-VC queuing is a strategy that is used to ensure fairness among all connections associated with a particular ATM service category by providing an individual queue for each shaped and unshaped connection.

The per-VC queue limit, *txQueueLimit*, defines the maximum number of cells that can be buffered for a single connection. The default limit is defined per ATM service category. Alternatively, you can configure a *txQueueLimit* for each VC. See the table Configurable parameters for per-VC queue default limits (page 20) for the GQM-based FPs.

On 16pOC3PosAtm FP, the default *TxQueueLimit* values of Nrt-VBR and UBR for a total of 10K should be sufficient for most applications. The *txQueueLimit* and *minPerVcQueueLimit* must be in the range of 32 to 261 120 cells and the *perVcQueueLimitReferenceRate* must not exceed 353 207 cells/s (OC3 line rate).

On 4pMRPosAtm FP, the default *TxQueueLimit* values of Nrt-VBR and UBR for a total of 40K should be sufficient for most applications. The *txQueueLimit* and *minPerVcQueueLimit* on this FP must be in the range of 32 to 1011840 cells. When operating in 4-port OC12 mode, the *perVcQueueLimitReferenceRate* on this FP must not exceed 1 412 830 cells/s (OC12 line rate). When operating in 1-port OC48 mode, the *perVcQueueLimitReferenceRate* on this FP must not exceed 5 651 320 cells/s (OC48 line rate).

Within the 45 000 possible connections of a GQM-based FP, up to 32 000 of them can have per-VC queues. When 90% of the 32 000 limit is occupied, alarm 7060 1201 is generated as a warning. Any attempt to set up a per-VC queue beyond the 32 000 limit will fail with alarm 7039 2001 generating against the component *Lp Eng Arc* for resource exhaustion. You can display the status of per-VC queues, as described in NN10600-715 *Nortel Multiservice Switch 7400/15000/20000 ATM Fault and Performance Management*. Again, because only up to 16 000 connections can be configured for an ATM interface, a 4pMRPosAtm FP in 1-port OC48 mode can only have up to 16 000 connections using per-VC queuing.

Schedulers for GQM-based FPs

GQM scheduling prioritizes and arbitrates cell transmission opportunities among the ATM service category classes within each link (a port on the FP), and among the connections within each service class. A scheduler determines the order in which cells are dequeued from buffer memory.

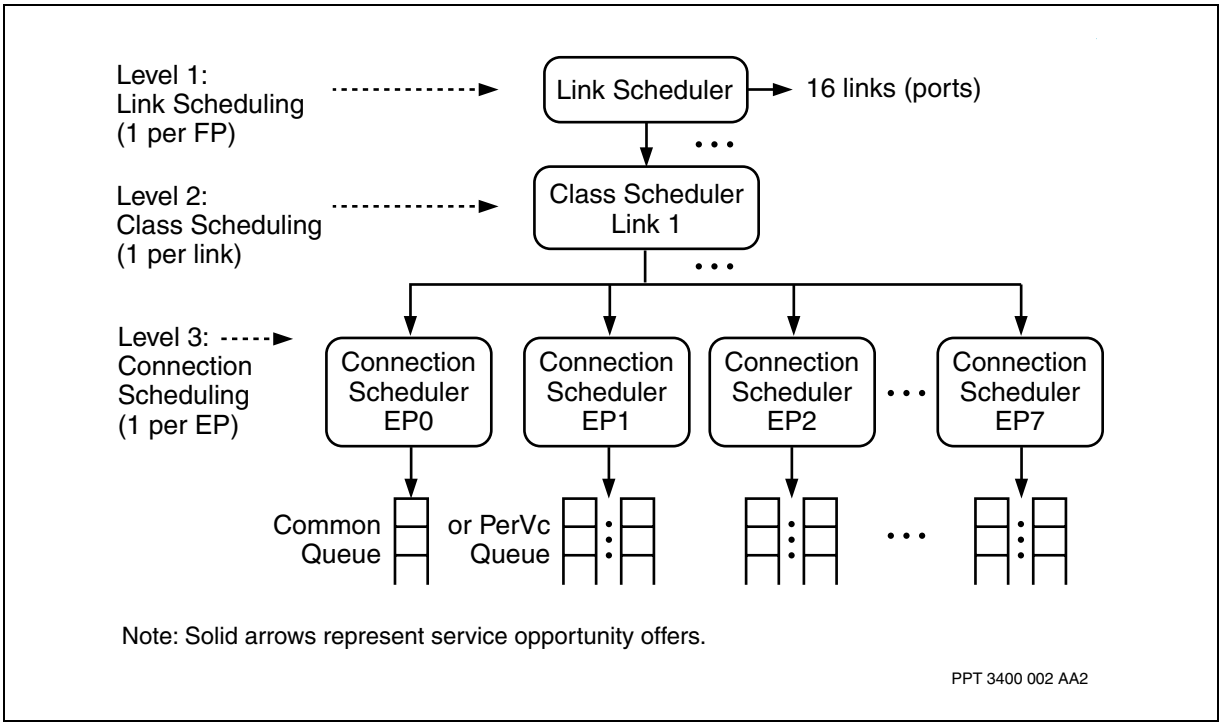
Copyright © 2005, Nortel Nortel Confidential

GQM-based FPs share the same types of schedulers as described in Overview of APC schedulers (page 89). In addition, see

- GQM link scheduling (page 97)
- GQM class scheduling (page 97)
- GQM connection scheduling (page 99)
- Configuring MBG values for GQM-based FPs (page 100)

The GQM scheduling hierarchy is different by allowing common queuing below a connection scheduler, as shown in the figure GQM scheduling hierarchy (page 97).

GQM scheduling hierarchy



GQM link scheduling

In GQM-based FPs, the link scheduler can arbitrate cell transmission opportunities on all ports of the FP in a round-robin manner. The round-robin is controlled by the hardware PHY device.

GQM class scheduling

The class scheduler evaluates the eligibility of all emission priorities (EPs) on a single link in order to control the relative delay and loss priorities of one service category relative to the other service categories.

Copyright © 2005, Nortel
Nortel Confidential

GQM-based FPs support all eight ATM EPs numbered 0 to 7. There is absolute priority for EP0 and EP1 with minimum delay and cell delay variation (CDV). EPs 2 to 7 use the conventional weighted fair queue (WFQ) discipline with service preferences over each other determined by minimum bandwidth guarantees (MBGs) for starvation avoidance. Bandwidth is first allocated to EP0 and EP1, then the remaining bandwidth is divided between EP2 to EP7 in proportion to the MBGs. The MBG represents the service weight that is associated with an EP served by the WFQ scheduler, which is work-conserving. (A work-conserving scheduler does not waste a cell opportunity if any of its children have data to send.) For special configuration requirements for GQM-based FPs, see Configuring MBG values for GQM-based FPs (page 100).

When an EP has no traffic to send, its bandwidth share is re-distributed to the other active EPs. The bandwidth allocation algorithm is as follows.

EPn = (the bandwidth left over from EP0 + EP1) times MBG divided by (the sum of the MBGs for active EPs)

where the sum of the MBGs is equal to or less than 100%

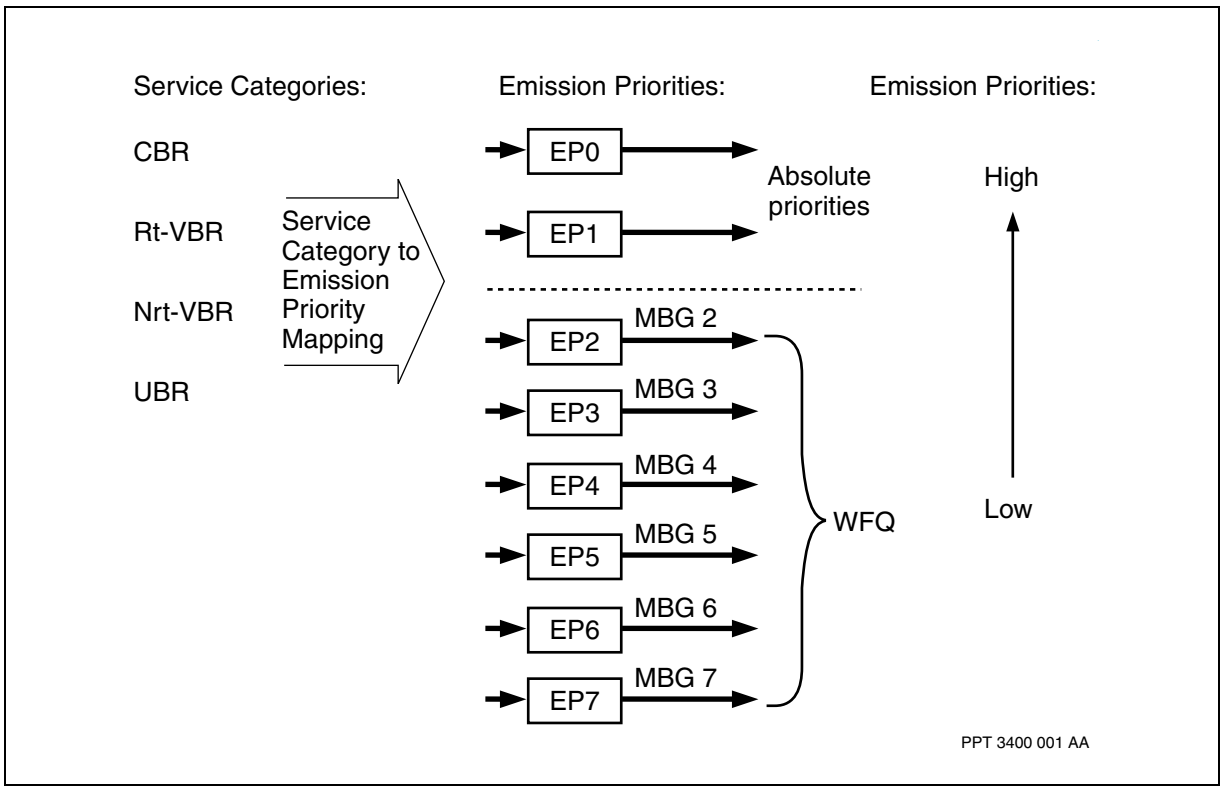
Each ATM service category can be provisioned with the EP it is to use, or use the defaults. Since there are only four supported categories, a maximum of four of the eight available EPs can be used. An ATM service category can be provisioned to use any EP provided:

- CBR has an equal or higher EP than Rt-VBR
- Rt-VBR has an equal or higher EP than Nrt-VBR
- Nrt-VBR has an equal or higher EP than UBR or UBR with MDCR
- no two ATM service categories share an EP unless they are both shaped

The relationships of EPs for class scheduling are shown in the figure GQM class scheduling (page 97).

Copyright © 2005, Nortel Nortel Confidential

GQM class scheduling



When an ATM service category is using an EP, you can use the default value of the MBG or you can provision one. Although the MBGs are provisioned for GQM-based FPs the same as for any other Nortel Multiservice Switch node ATM FP, there are differences in the GQM scheduling behavior due to how the residual bandwidth left over from EP 0 and EP1 is allocated. Choose MBG values carefully, as described in Configuring MBG values for GQM-based FPs (page 100).

GQM connection scheduling

In GQM-based FPs, the connection scheduler for each emission priority (EP) arbitrates the cell transmission opportunities among connections. The versions of connection schedulers are:

- Rate scheduler (page 99)
- Weighted fair queue (WFQ) scheduler (page 100)

Rate scheduler

The rate scheduler spaces cell departures according to the connection's transmission rate, and is also known as a traffic shaper. A shaper schedules departure time and is not work-conserving. (A non-work-conserving scheduler means that the cell opportunity which is unused by a child will not be offered to other children.)A shaper is typically used prior to a network

Copyright © 2005, Nortel
Nortel Confidential

interface where the connection traffic stream can be policed against limits on rate and delay variation. The individual connection weights do not apply to a rate scheduler. A rate scheduler needs per-VC queuing.

Weighted fair queue (WFQ) scheduler

The WFQ scheduler serves each connection in an inter-leaved round-robin manner with service rounds proportional to the connection's weight. WFQ schedules departure sequence and is work-conserving. It is used for a class that needs fairness (and protection from other misbehaving connections) and less stringent CDV. For connections using a WFQ scheduler, per-VC queuing is required to achieve weighted fairness in bandwidth allocation among the connections.

The WFQ can also be used for a class that desires bounded CDV but does not need fairness among connections. In this case, a WFQ serves a single queue exclusively which is used as a common queue for all connections of the class. In this common application, connection weights have no effect.

Configuring MBG values for GQM-based FPs

When a GQM-based FP uses a lower emission priority (EPs) 2 to 7 for an ATM service category, the minimum bandwidth guarantee (MBG) value can be configured. The attribute *minimumBandwidthGuarantee* of the component *Ep* ranges from 1 to 100 with an additional value *priority*. The value *priority* indicates the EP is scheduling according to a strict priority scheme without providing any additional bandwidth guarantee to the EP. Since the component *Ep* is optional, not configuring it means the attribute defaults to *priority* for the EP.

Configuring the MBGs for the lower EPs is the same for GQM-based FPs as other Nortel Multiservice Switch node ATM FPs, but the scheduling behavior for the GQM-based FPs differs because of the way residual bandwidth is allocated. The residual bandwidth is what remains after the absolute priorities of EP0 and EP1 are serviced. A GQM-based FP allocates the MBG value to each EP, and then allocates any remaining bandwidth amongst the EPs according to a strict priority such that a lower EP may not receive any extra bandwidth (that is, be starved) if the higher priority EPs use it all. This type of scheduling is called priority guaranteed queuing (PGQ).

A GQM-based FP divides the total residual bandwidth among the EPs proportional to the MBG of the EP. This type of scheduling is weighted fair queuing (WFQ) where the MBG represents the weight for the EP. The priority of the EP does not affect how much bandwidth it receives. See Weighted fair queue (WFQ) scheduler (page 100).

Copyright © 2005, Nortel Nortel Confidential

Both PGQ and WFQ provide identical behavior when one or two of the lower EPs are used. One EP would receive all of the bandwidth, while each of two would receive its MBG value. The unused bandwidth of one EP is available to the other EP.

Both PGQ and WFQ provide identical behavior when more than two of the lower EPs are used and all of the EPs are congested. Each EP receives its specified MBG value.

The PGQ and WFQ behavior differs when three or more of the lower EPs are used provided some EPs are not using their assigned MBG values. This behavior offers service differentiation. The higher EPs in the lower range (for example, EP2) can be routinely assigned a much higher MBG than the actual traffic planned for this EP since the delay and loss characteristics of an EP are both a function of the use of the EP.

The ATM data model allows for any number of the lower EPs (2 to 7) to specify a non-zero MBG while the remaining EPs specify priority. However, as an exception, a mix of priority and numerical MBG configuration is not supported for GQM-based FPs. If one of the lower EPs (2 to 7) is configured with a numerical MBG, then the lower EPs (2 to 7) with ATM service mapping must also be configured with numerical MBGs. This is to ensure that the operator clearly defines the intended bandwidth allocation among the EPs. GQM-based FPs support either specifying the same priority option for all EPs or the explicit configuration of the MBG value for all used EPs. A semantic check enforces using one or the other when either is configured.

MBG values for the EPs apply to individual ATM interfaces only. Numerical MBGs for the EPs can be configured on one ATM interface and priority MBGs for the EPs can be configured on another interface. By default, an ATM interface has EPs with priority MBGs.

The MBG significantly affects the realized quality of service (QoS) attributes of cell loss, delay, and jitter of each EP. Assuming that the sum of the MBGs of EPs 2 to 7 equals 100, you can approximate each EP as a virtual link with capacity determined by its MBG value.

The table, The EP default weight values for a priority MBG (page 102), describes the actual weights used as MBGs for the EPs in use, where the EPs have priority MBGs by default. The EPs use the default value or can be configured with an MBG value of priority. In the table, EPa, EPb, EPc, and EPd represent four EPs that have descending priority order. The subscripts a, b, c, and d can be any value between 2 and 7. If each of the ATM services, CBR, rtVBE, nrtVBR, and UBR, maps to a unique EP, then up to four EPs will be used. If there are two or more ATM services mapped to one EP, then fewer than four EPs will be used.

Copyright © 2005, Nortel
Nortel Confidential

The EP default weight values for a priority MBG

Number of lower EPs in use	Assigned EP weighted values			
	EPa	EPb	EPc	EPd
1	100	not applicable	not applicable	not applicable
2	99	1	not applicable	not applicable
3	90	9	1	not applicable
4	77	18	4	1

Copyright © 2005, Nortel
Nortel Confidential

Memory management

This chapter describes the memory management capabilities of Nortel Multiservice Switch function processors (FPs).

Navigation

- Overview of memory management (page 103)
- Memory management for APC-based FPs (page 103)
- Memory management for ATM IP FPs (page 105)
- Memory management for CQC-based FPs (page 116)

Overview of memory management

Nortel Multiservice Switch nodes use memory resources to store connection records, and data frames and cells awaiting transmission or processing. Memory resources also support internal operations. Memory management partitions the memory resources of a FP into sections to fulfill these needs.

Resource control mechanisms are in place to manage the queue and memory resources that monitor and control the link, processor, backplane, and memory resources. These mechanisms permit you to configure ATM queue management connection pools and frame connection resources for virtual channel connections (VCC) under a virtual path termination (VPT) and Multiservice Switch node logical network number (LNN) connections. FPs resources have two areas:

- ATM resources
- frame resources

For all ATM FPs, buffer space is divided into 64-byte cell blocks (for cell free lists) and 256-byte frame blocks (for frame free lists). The specific combination of blocks depends on the FP.

Memory management for APC-based FPs

The APC-based FPs include:

- 16pOC3SmIrAtm (NTHW21, NTHW24, and NTHW31)

Copyright © 2005, Nortel

Nortel Confidential

- 4pOC12SmIrAtm (NTHW86 and NTHW11)

Each APC-based card has four APC devices. The attribute *bufferLimitPerEP* replaces the attribute *classBufferPoolLimit* (cBPL) and changes the buffer sizes from a percentage to a number of cells. The attribute *bufferLimitPerEP* divides the total amount of the egress buffer on each APC device amongst the five emission priorities (EPs). The attribute specifies the maximum buffer size allowed for each EP on each APC so that one EP cannot use all the buffer space. (AQM-based FPs show the attribute, but it applies only to APC-based FPs.

There are four APCs on a 4pOC12SmIrAtm FP. On a 4pOC12SmIrAtm FP, since each APC handles one OC-12 link, the attribute bufferLimitPerEP allows the APC egress buffer to be among the five EPs on the link.

On a 16pOC3SmIrAtm FP, since each APC handles four OC-3 links, the attribute bufferLimitPerEP allows the APC's egress buffer to be divided into five blocks with one block per EP shared among the traffic of that EP across all four OC-3 links. On 16pOC3SmIrAtm FP, the memory block for the EP is further divided equally among the four OC-3 links. For example, if the bufferLimitPerEP value for EP2 is 39168 cells for an APC, then each of the EP2 priorities on the four links of the APC automatically has a buffer limit of 9792 (from 39168 divided by 4).

The attribute *bufferLimitPerEP* is a vector with eight values divided into numbers of cells. Since an APC supports only five of the eight values (EP0, EP2, EP3, EP4, and EP7), only these have non-zero default values. These five values can be user-configured buffer limits (maximum buffer sizes) for the five EPs supported on the APC.

The default egress buffer space settings have changed in PCR5.1.1 from 511K cells to 255K minus 2 cells (261118 cells). The change indicates the actual amount of buffer memory that is always used by these FPs. The change does not affect performance, the maximum number of ATM connections per FP and per port, or the maximum call setup rate. The table Default buffer sizes of attribute bufferLimitPerEP for APC-based FPs (page 105) identifies how much buffer space has been allotted to each EP, and whether it can be changed.

Copyright © 2005, Nortel
Nortel Confidential

Default buffer sizes of attribute bufferLimitPerEP for APC-based FPs

EP	Default value of cells prior to PCR5.1.1	Default value of cells in PCR5.1.1 and later	Percentage of total space	User-configured
0	78 490	39 168	15 %	yes
1	0	0	0	no
2	78 490	39 168	15 %	yes
3	261 632	13 0559	50 %	yes
4	78 490	39 168	15 %	yes
5	0	0	0	no
6	0	0	0	no
7	26162	13 055	5 %	yes

In PCR5.1 GA, a semantic check enforces that the sum of the values associated with *bufferLimitPerEP* under Lp Eng Arc Apc Ov does not exceed 261118 cells. To ensure proper isolation among EPs, the sum of attribute *bufferLimitPerEP* values for all EPs per APC must not exceed the total available APC egress buffer space, which is 261118 cells.

For migrations to PCR5.1.1, the default values do not need to be modified before or after the migration provided no subcomponent Ov is present and provided the older defaults have not been manually changed. When the default values are used, a semantic check in PCR5.1 GA ensures that the sum of each EP buffer does not exceed 261 118 cells. If you choose to reduce the sum below 261 118 cells, it is recommended that you first set to zero (0) those EP buffers that will not be needed, and then lower the remaining EP buffers.

Memory management for ATM IP FPs

This section provides information on memory management for ATM IP FPs.

This chapter provides information in the following sections:

- PQC buffer space on ATM IP FPs (page 106)
- Memory management on ATM IP FPs (page 108)
- PQC CQM memory management (page 110)
- AQM CQM memory management (page 113)

Copyright © 2005, Nortel Nortel Confidential

PQC buffer space on ATM IP FPs

On the Nortel Multiservice Switch node queue controller (PQC), use and allocation of buffer space is influenced by the following factors:

- Frame buffers are used for AAL5 segmentation and reassembly.
- Frames require much more processing. The complete frame must be buffered before it can be segmented for transmission and before it can be reassembled and passed to the higher layer frame application.
- Cell buffers are used for cell relay.
- If the link is not operating in a congested state, the number of cell buffers needed is minimized since arriving cells can be processed with maximum efficiency.

Configurable attributes are provided which specify the percentage of buffer memory that is allocated as frame memory, with the remaining buffer memory allocated as cell memory. There are separate attributes for receive and transmit cell queue memory (CQM). Memory is allocated to cell and frame free lists.

The ATM software on the FP requires a guaranteed minimum number of cell and frame blocks. If the value of either the receive or transmit frame memory allocation is smaller than the absolute minimum required, the ATM software will allocate the absolute minimum rather than the actual number requested. The actual amount of memory allocated can be displayed through the operational attributes under the ATM resource usage level.

The PQC buffer space has the following characteristics:

- The buffer space supports the free lists and common queuing.
- The buffer space is divided into 64-byte cell blocks and 256-byte frame blocks which make up the cell and frame free lists.
- One cell block is 64 bytes and its data portion is 48 bytes.
- One frame block is 256 bytes and its data portion is 240 bytes.
- The default partitioning of available CQM is 80 percent for frame blocks for ingress and 50 percent for frame blocks for egress, measured in cells.

As the number of connections increases, the FP uses more overhead memory to maintain those connections. As a result, less memory is available for cell and frame blocks.

The context space consists of a fixed amount of memory which is reserved for a Multiservice Switch node's internal system use.

Copyright © 2005, Nortel Nortel Confidential

AQM buffer space on ATM IP FPs

For the ATM queue manager (AQM) on ATM IP FPs, allocation of cell buffer space is influenced by the following factors:

- Within any emission priority, a large number of per-VC queues uses more buffer space than a common queue.
- If the link is not operating in a congested state, the number of cell buffers needed is minimized since arriving cells can be processed with maximum efficiency.

The node allocates buffer space to the queues on demand, depending on the traffic requirements.

The main advantage of a larger buffer space is to minimize CLR. A larger buffer space results in lower CLR for each connection. The CLR is very dependent on both the buffer allocation and the number of queues supported on the ATM FP. For more details on engineering considerations for configuring the frame and cell memory, see the *Nortel Multiservice Switch Release Notes*.

The AQM buffer space has the following characteristics:

- The buffer space supports the free lists, per-VC queues, and common queuing.
- The buffer space is divided into 64-byte cell blocks which make up the cell free list.
- One cell block is 64 bytes and its data portion is 48 bytes.

As the number of connections increases, the FPs uses more overhead memory to maintain those connections. As a result, less memory is available for cell and/or frame blocks.

The context space consists of a fixed amount of memory which is reserved for a Nortel Multiservice Switch node's internal system use.

The allocation of buffer space differs between ATM IP FPs:

- OC3 FPs have an AQM for each port. Therefore, each port is allocated the buffer space for an entire AQM.
- DS3 and E3 FPs have a single AQM on which buffer space is shared between all ports.

Copyright © 2005, Nortel Nortel Confidential

Memory management on ATM IP FPs

You configure these resources to refine the policies that determine the amount of resource that the node requires to support connections, multicast branches, connections under a VPT, and frame and cell blocks. Through configuration, you fine-tune the context and queuing memory (CQM) for two ASICs on the FP:

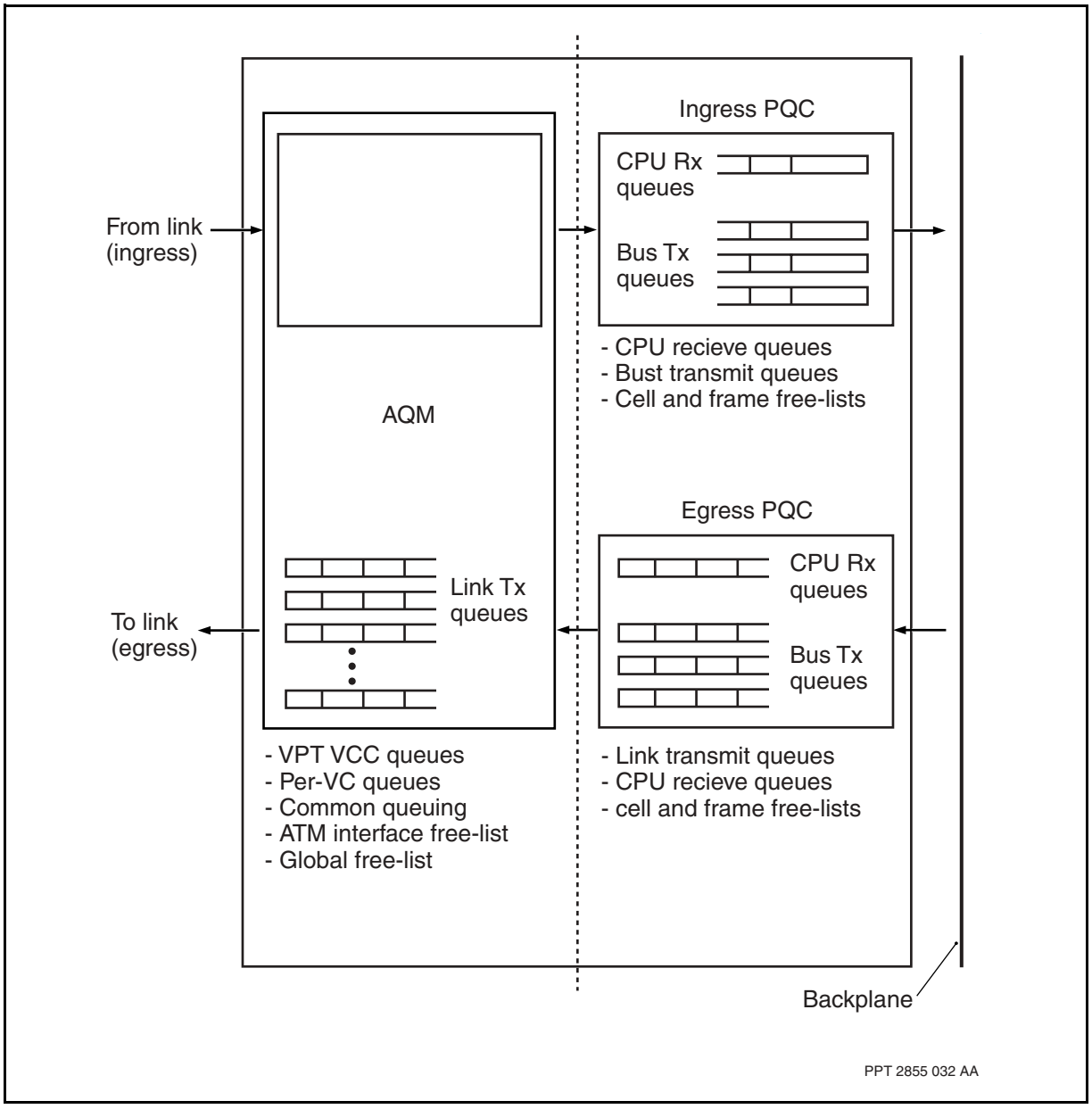
- PQC CQM
- AQM CQM

You configure frame resources through fine-tuning only the PQC CQM.

The figure PQC CQM and AQM CQM: ATM IP FPs (page 109) illustrates the PQC and the AQM, and the queues that each supports.

Copyright © 2005, Nortel Nortel Confidential

PQC CQM and AQM CQM: ATM IP FPs



The PQC CQM supports these base-layer functions:

- per-destination queuing towards the node's bus (three emission priorities)
- per-AQM queuing towards the links (three emission priorities)
- ATM cell forwarding
- frame forwarding (frame relay, IP)
- AAL5 segmentation and reassembly

Copyright © 2005, Nortel
Nortel Confidential

- Nortel Multiservice Switch node frame segmentation and reassembly

The AQM CQM supports these higher-layer functions:

- ATM connection identification
- per-VC queuing and VPT VCC queuing
- global and ATM interface pools

PQC CQM memory management

The PQC ASIC processes frames and cells between Nortel Multiservice Switch node FPs, and frame traffic to and from the link. The ATM IP FP has a fixed amount of PQC CQM memory. This memory is divided into three functional areas:

- connection space
- buffer space
- context space

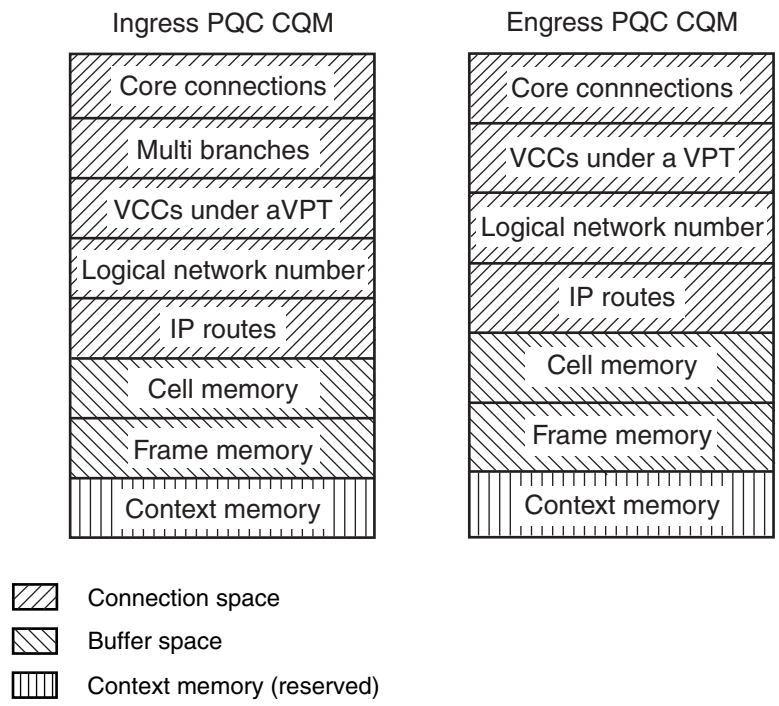
The figure Partitioning of PQC CQM (page 111) shows how Multiservice Switch nodes apply PQC CQM memory partitioning.

There are two PQCs on the FP. The connection and buffer spaces comprise the configurable memory space. The connection space consists of connection context records which the node uses to configure and monitor each connection. The buffer space takes up the remainder of the configurable PQC CQM memory and consists of cell and frame memory. The context space consists of a fixed amount of memory which is for internal use.

The default PQC memory configuration for an ATM IP FP supports the required number of connections. The node uses the remaining PQC memory for frame and cell buffering. The node undertakes cell and transmit queuing in the AQM CQM and uses the PQC CQM for frame reassembly operations. The default PQC configuration provides the maximum available memory for frame reassembly.

Copyright © 2005, Nortel
Nortel Confidential

Partitioning of PQC CQM



Note: IP routes is configurable in this release. This capability is available to support future implementation for IP routing over ATM.

PPT 2855 033 AA

Scope of configuration for these resources is:

- connection space resource is configurable at the ATM and frame resource level
- the amount of buffer space resource is the memory remaining after allocating connection space, and is automatically derived by the node
- core connections, multicast connections, VCCs under a VPT, logical network number (LNN), and IP routes resources are configurable
- cell memory is configurable as a percentage of the buffer memory available, and frame memory uses the remaining percentage of buffer memory

PQC CQM connection space

The PQC CQM connection space consists of connection context records, where each record configures and monitors a single connection. You configure the total connection pool capacity under the ATM resource control to define the maximum number of connections that the FP must support. This

Copyright © 2005, Nortel
Nortel Confidential

configured value defines the amount of memory that the node allocates for the connection space in this FP and the number of connections that the PQC can support.

For ATM cell applications, the connection space has allocations for core and multicast connections. Core connections include all point-to-point VCCs and VPCs. Multicast connections include point-to-multipoint SVCs. For frame applications, the node allocates memory resources for VCCs under a VPT. LNN space is configurable and supports future implementations for TCP/IP routing.

PQC CQM management has the following characteristics:

- The node supports multicast connections on the ingress PQC CQM. Therefore, the egress PQC CQM does not have a memory allocation for these connections. The egress PQC CQM distributes the unused multicast allocation to the cell and frame memory according to the configured percentages.
- You do not need to configure the minimum amount of multicast branch resources. When used, memory space for multicast connections space is shared with the core connection space.
- Frame relay applications use sub-connections to interwork with ATM. Sub-connections identify a specific data link connection identifier (DLCI) on the frame relay FP as a basis for interfacing ATM connections with frame relay connections.
- VCCs under a VPT use the core connection space.
- LNN resources maintain connectionless resources on frame-oriented services that interwork with ATM. This memory allocation supports the following connectionless services:
  - dynamic packet routing system (DPRS) one connection resource per service connection
  - logical trunks (three connection resources per service connection)

PQC CQM frame and cell blocks

The frame and cell blocks use the remainder of the configurable PQC CQM memory after connection space is allocated. Configurable attributes define the percentage of memory for cells, with the remainder of the buffer allocated to frames. This configuration approach applied to blocks on both the ingress and egress PQC CQM. This memory is allocated to cell and frame free lists, respectively.

In this way, you can fine tune the node to support the resource requirements of the whole network. If you expect frame traffic through a node to be high, you can reduce the amount of memory for cells thereby increasing the remainder for frames. You can also configure the buffer spaces across all PQCs to

Copyright © 2005, Nortel Nortel Confidential

minimize congestion. Egress congestion is more typical on the AQM (due to shaping and queuing congestion) than on the PQC (due to more cells than the AQM can handle).

The ATM software on the FP needs a guaranteed minimum number of cell and frame blocks. If the value of the ingress (or egress) frame memory allocation is smaller than the absolute minimum, the ATM software re-allocates resources to support the absolute minimum and ignores the configure values. You can review the actual amount of allocated memory through operational attributes.

AQM CQM memory management

The AQM ASIC performs specialized functions on ingress and egress traffic before passing this traffic to the backplane. The number of AQM ASICS present on the FP depends on the FP type. See the table Number of AQM ASICs by FP (page 113) for a summary of AQM ASIC configurations.

The AQM supports ATM services only.

The AQM ASIC uses a specific CQM for its internal processing. Like the PQC CQM, the AQM CQM has three functional areas: connection space, buffer space, and context memory. The figure Partitioning of AQM CQM (page 114) shows the AQM CQM allocation.

Number of AQM ASICs by FP

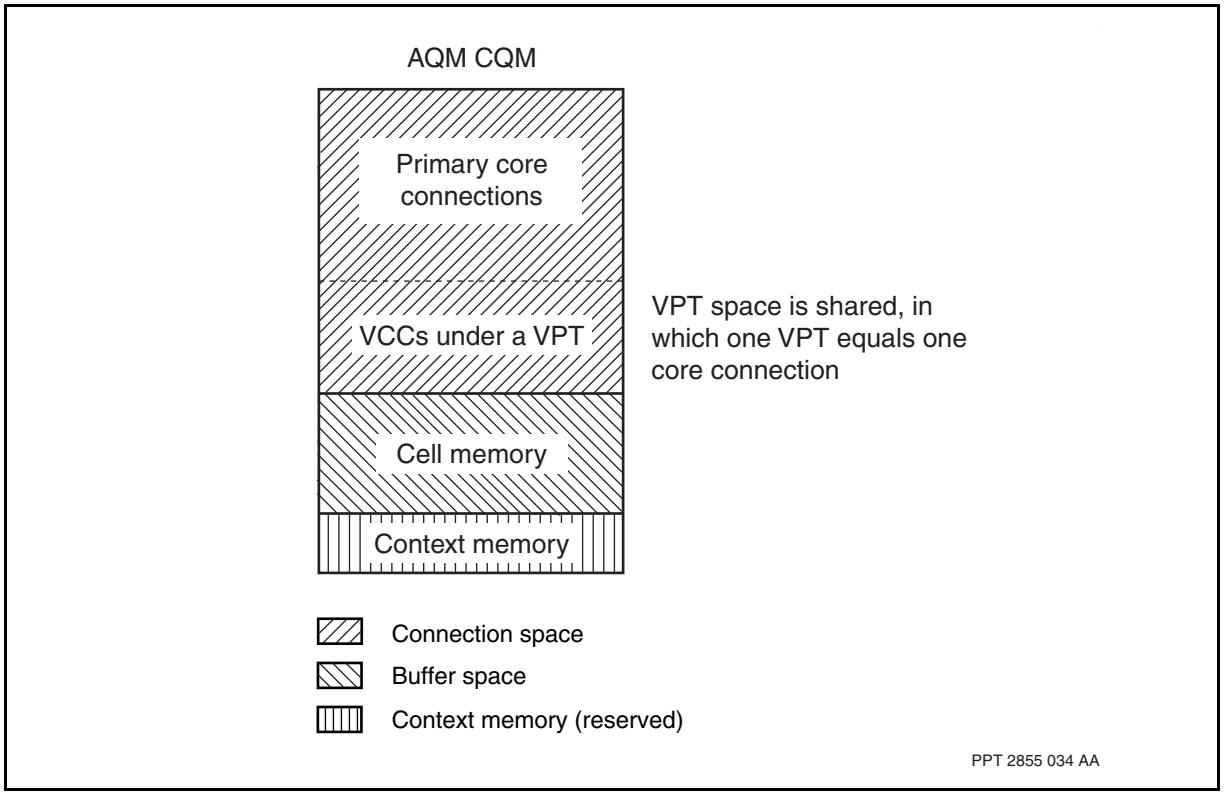
FP	Number of AQM ASICs
2-Port ATM IP OC-3 single mode FP	two, one per port
2-Port ATM IP OC-3 multi-mode FP	two, one per port
3-Port ATM IP DS3 FP	one, shared between all ports
3-Port ATM IP E3 ATM FP	one, shared between all ports
16-port OC3 ATM FP	one, shared between all ports
32-port DS1 MSA FP	one, shared between all ports
32-port E1 MSA FP	one, shared between all ports
32-port DS1 MSA FP with optical ports	two, one shared between all electrical ports, one for the optical port
(1 of 2)	

Copyright © 2005, Nortel
Nortel Confidential

Number of AQM ASICs by FP (continued)

FP	Number of AQM ASICs
32-port E1 MSA FP with optical ports	two, one shared between all electrical ports, one for the optical port
When SONET linear automatic protection switching (APS) is configured on OC-3 ATM IP FPs, the number of AQMs in service is equal to the number of APS components configured for the FP. For OC-3 ATM IP FPs, only one APS is possible. The AQM instance is directly related to the APS instance of that FP, therefore it allows configuration of both AQMs on the OC-3 ATM IP FP using a single AQM component.	
(2 of 2)	

Partitioning of AQM CQM



AQM CQM connection space

Like the PQC CQM, the AQM CQM connection space consists of connection context records, where each record configures and monitors a single connection. You configure the connection pool capacity to define the maximum number of connections that the FP must support. This configured value defines the amount of memory that the node allocates for the connection space.

Copyright © 2005, Nortel
Nortel Confidential

The AQM CQM connection space has allocations for core connections as well as virtual connections under VPT connections. Core connections include all VCCs and VPCs under actual interfaces. VPT connections include VCCs under virtual interfaces.

The AQM connection space is configurable. For each configured core connection, the number of possible VPT connections is reduced by one.

AQM CQM management has the following characteristics:

- Core connections represent the VCCs and VPCs for general ATM processing.
- VCCs under a VPT represent the connections used to support ATM VPT processing. They are used to distinguish VCCs under virtual interfaces from VCCs under actual interfaces.
- VPT VCC resources share some AQM CQM space with core connection resources. AQM CQM space required for one VPT and its associated VCCs is the same as the space required for one core connection.

It is possible to configure total AQM connection capacity to be greater than the PQC connection capacity. That is, the combined number of connections that all AQMs on the FP can support can be greater than the total connection pool capacity that you define through the ATM resource control. The total capacity is the actual limit for the FP. This configuration allows individual AQMs to support connections as long as space is available.

Lastly, the connection pool capacity for a single AQM must be less than or equal to the total capacity defined under ATM resource control.

AQM CQM cell blocks

The cell blocks use the remainder of the configurable AQM CQM memory after connection space is allocated. Nortel Multiservice Switch nodes use the AQM CQM cell blocks for:

- ATM interface free lists
- per-VC and common queuing

The four Mbyte AQM CQM offers a typical buffer space on the order of 56K cells, as shown in the table Example of buffer memory and corresponding number of connections: DS3/E3 ATM IP FPs (page 116). In the case of the ATM IP OC3 FP, this amount of buffer space is available per port. For ATM IP DS3 and E3 FPs, this buffer space is shared among all ports.

The ATM software requires a guaranteed minimum number of cell blocks on the AQM. The node automatically re-allocates unused space from the connection space for the cell buffers. The table Example of buffer memory and

Copyright © 2005, Nortel
Nortel Confidential

corresponding number of connections: DS3/E3 ATM IP FPs (page 116) shows buffer space allocations for a set of sample configurations by number of connections for the ATM IP DS3/E3 FPs.

Example of buffer memory and corresponding number of connections: DS3/E3 ATM IP FPs

Number of connections	Buffer memory (in cells)
100	49 152
500	47 104
1024	45 056
3072	38 912
The general formula for N connections is approximately 50 611 - (4 * N) cells. If N > 2048, deduct an additional 1024 cells from the result of the formula.	

ATM interface free lists

ATM interface free lists are configurable for the AQM on ATM IP FPs.

An ATM interface free list limits the amount of cell buffer memory that the FP allocates for connections destined for the interface. You can configure ATM interface pools for 100 per cent allocation or for overbooking.

Memory management for CQC-based FPs

This section provides information on memory management for CQC-based FPs. The following topics are discussed in this section:

- Buffer space for CQC-based FPs (page 116)
- Memory management on CQC-based FPs (page 117)

Buffer space for CQC-based FPs

Use of buffer space, and therefore its allocation, is influenced by the following factors:

- A large number of per-VC queues uses more buffer space than the common queues.
- Frame buffers are used for AAL5 segmentation and reassembly.
- Frames require much more processing. The complete frame must be buffered before it can be segmented for transmission and before it can be reassembled and passed to the higher layer frame application.
- Cell buffers are used for cell relay.

Copyright © 2005, Nortel Nortel Confidential

- If the link is not operating in a congested state, the number of cell buffers needed is minimized since arriving cells can be processed with maximum efficiency.

Configurable attributes are provided which specify the percentage of buffer memory that is allocated as frame memory, with the remaining buffer memory allocated as cell memory. There are separate attributes for receive and transmit CQM memory. Memory is allocated to cell and frame free lists respectively.

The main advantage of a larger buffer space is to minimize CLR. A larger buffer space results in lower CLR for each connection. The CLR is very dependent on both the buffer allocation and the number of queues supported on the ATM FP. For more details on engineering considerations for configuring the frame and cell memory, see *Nortel Multiservice Switch Release Notes*.

The ATM software on the FP requires a guaranteed minimum number of cell and frame blocks. If the value of either the receive or transmit frame memory allocation is smaller than the absolute minimum required, the ATM software will allocate the absolute minimum rather than the actual number requested. The actual amount of memory allocated can be displayed through the operational attributes under the ATM resource usage level.

The CQM buffer space has the following characteristics:

- The buffer space supports the free lists, per-VC queues, and common queuing.
- The buffer space is divided into 64-byte cell blocks and 256-byte frame blocks which make up the cell and frame free lists.
- One cell block is 64 bytes and its data portion is 48 bytes.
- One frame block is 256 bytes and its data portion is 240 bytes.
- The default partitioning of available CQM is 80 percent for frame blocks for ingress and 50 percent for frame blocks for egress, measured in cells.

As the number of connections increases, the FPs uses more overhead memory to maintain those connections. As a result, less memory is available for cell and/or frame blocks.

The context space consists of a fixed amount of CQM memory which is reserved for a Nortel Multiservice Switch node's internal system use.

Memory management on CQC-based FPs

The CQC supports these base-layer functions:

- per-destination queuing towards the node's bus
- queuing towards the links

Copyright © 2005, Nortel Nortel Confidential

- ATM cell forwarding
- frame forwarding (frame relay)
- AAL5 segmentation and reassembly
- Nortel Multiservice Switch node frame segmentation and reassembly

The figure CQM on the CQC-based FP (page 119) shows the CQM on a CQC-based FP and the queues that it supports.

CQC-based FPs have two resources that you must configure before allocating any ATM connections:

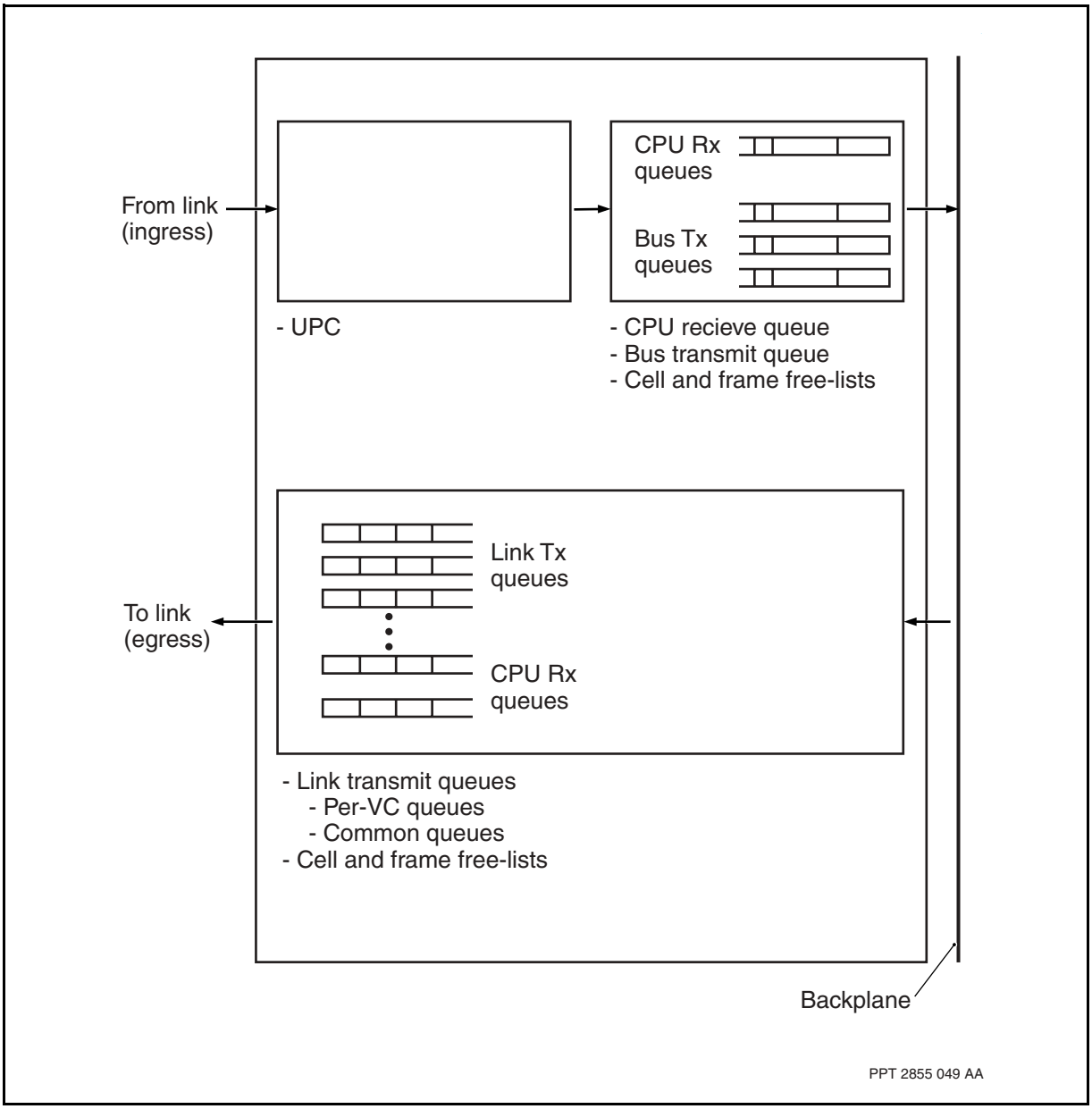
- cell queue memory (CQM) for both receive and transmit directions
- the division of the shaping stacks among the ATM FP's ports

The CQM on CQC-based FPs is divided into three spaces:

- connection space
- buffer space
- reserved space

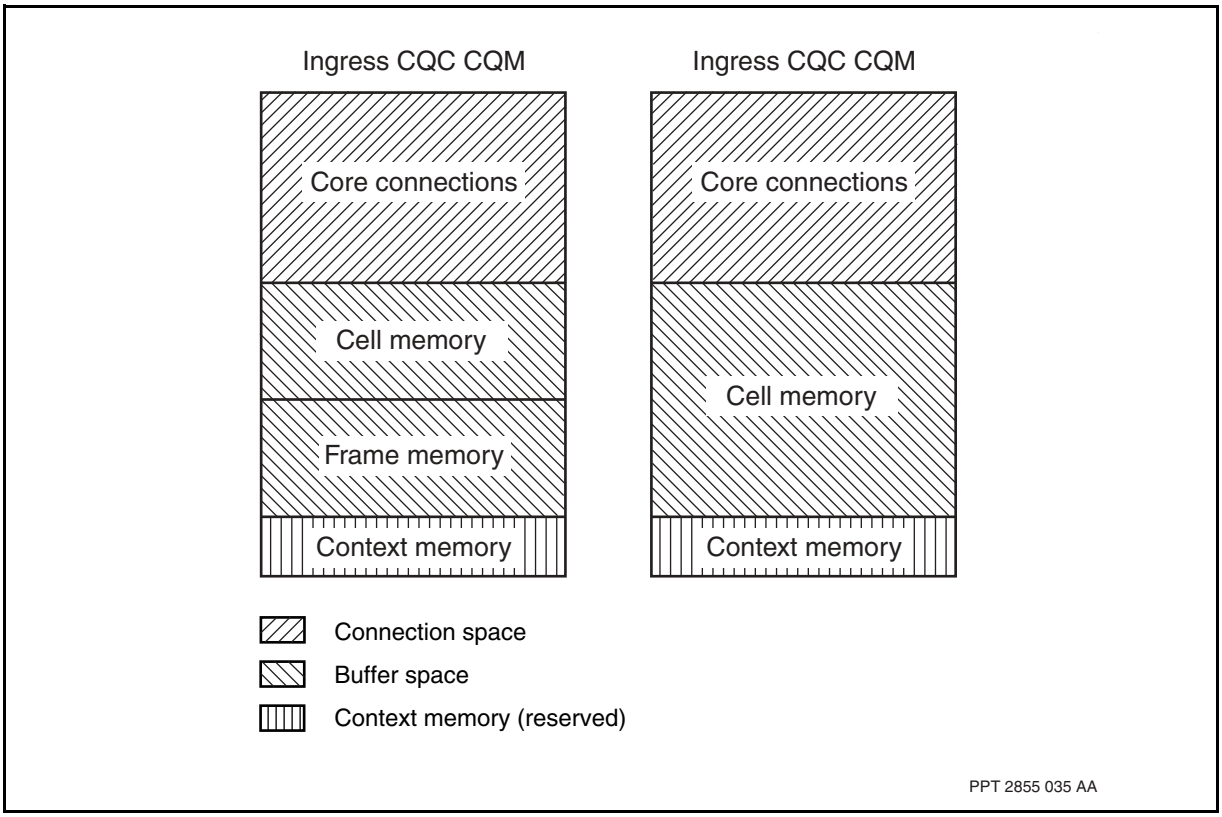
Copyright © 2005, Nortel
Nortel Confidential

CQM on the CQC-based FP



Copyright © 2005, Nortel
Nortel Confidential

Partitioning of CQC CQM



The following sections describe each of these memory spaces. For information on shaping stacks, see the section on traffic shaping on CQC-based FPs in NN10600-706 *Nortel Multiservice Switch 7400/15000/20000 ATM Traffic Shaping and Policing Fundamentals*.

CQC connection space

The connection space consists of connection context records which are used to configure and monitor a particular connection. You configure the total connection pool capacity under the ATM resource control to define the maximum number of connections that the FP must support. This configured value defines the amount of memory that the node allocates for the connection space in this FP and the number of connections that the PQC can support. You must establish a compromise between connection space and buffer space. The more connections a FP supports, the less memory is available for frame and cell buffers.

If the interface supports point-to-multipoint SVCs, you must give additional consideration to allocating connection space resources to multicast connections. This allocation affects how much memory resource is available

Copyright © 2005, Nortel
Nortel Confidential

for cell and frame memory. See NN10600-702 *Nortel Multiservice Switch 7400/15000/20000 ATM Routing and Signalling Fundamentals* for more information on point-to-multipoint connections.

Connection space is handled differently depending on the following factors:

- the ATM FP you are using, and
- the method you are using to assign connection pool capacity resources you are using

The table Guidelines for assigning connection space for CQC-based FPs (page 121) lists ATM FPs and indicates how connection space can be assigned for each.

Guidelines for assigning connection space for CQC-based FPs

FP	Number of ports	Assign connection space using
JT2 ATM	2	total connection pool capacity or per-port connection pool capacity
3-port DS1/E1 ATM	3	total connection pool capacity or per-port connection pool capacity
8-port DS1/E1 ATM	8	total connection pool capacity only
DS3/E3 ATM	3	total connection pool capacity or per-port connection pool capacity
OC-3 ATM	3	total connection pool capacity or per-port connection pool capacity

Connection space for 8-port DS1/E1 FPs

If you are using 8-port DS1/E1 ATM FPs, then connection space is allocated on the basis of the FP's total connection pool capacity.

This capacity specifies the maximum number of connections (VCC and VPC) that are available across all independent links and inverse multiplexing for ATM (IMA) virtual links on the FP. Total connection pool capacity must equal or exceed the sum of all connections configured on a FP. In this case, connection space is not allocated on a per-port basis and there is no dependence on the connection map. The total connection pool capacity for a FP can be set from 0 to 10 752, with a default of 3072.

Copyright © 2005, Nortel
Nortel Confidential

Connection space for other ATM FPs

If you are using any FPs other than the 8-port DS1/E1 ATM FPs (see the table Guidelines for assigning connection space for CQC-based FPs (page 121)), then you can assign connection space using

- total connection pool capacity
- per-port connection pool capacity

Use total connection pool capacity for non 8-port DS1/E1 ATM FPs unless you have to

- distribute connections unevenly across ports of a FP
- define more than 2560 connections on port 0 or port 1 of FP

For non 8-port DS1/E1 ATM FPs, the total connection pool capacity is divided evenly among all ports available. The number of connections supported on each port must exceed the range specified in the connection map for each ATM interface bound to a port on the LP. The table Total connection pool capacity limits for 2- and 3-port ATM FPs (page 122) shows the minimum, default, and maximum values for total connection pool capacity supported on all non 8-port DS1/E1 CQC-based FPs.

Total connection pool capacity limits for 2- and 3-port ATM FPs

Number of ports on FP	Minimum value	Default value	Maximum value	Value in multiples of
2	1024	3072	8192	512
3	1536	3072	7680	768

Use per-port connection pool capacity if you have to

- distribute connections unevenly across ports of a FP, or
- define more than 2560 connections on port 0 or port 1 of a FP

Using per-port connection pool capacity, the number of connections required per port is configurable. The minimum, maximum, and default values for per-port connection pool capacity are shown in the table Connection pool capacity limits for 2- and 3-port ATM FPs (page 123).

Copyright © 2005, Nortel Nortel Confidential

Connection pool capacity limits for 2- and 3-port ATM FPs

Port	Minimum value	Maximum value	Default value
0	512	4096	0
1	512	4096	0
2	512	2560	0
All ports can be simultaneously configured to maximum values. The “0” default for connection pool capacity implies that, by default, connection space is assigned using the total connection pool capacity.			

The number of connections defined per port must be greater than or equal to the number of connections for the connection mapping space. The value of the connection mapping space must in turn be greater than the combined values for maximum VCCs and maximum VPCs.

The number of connections under port 2 is smaller so that there is enough space allocation for:

- some reserved connections
- the minimum cell and frame buffers
- the multi-cast logical connection identifiers (LCI) for system use

The minimum number of connections permitted is 512. The number of connections is allocated as 256 connections reserved for VPCs and 256 connections reserved for the VPI zero VCC space. In this configuration there are no VCCs in the programmable VCC range. If it is necessary to define VCC with non- zero VCCs, additional connections can be allocated in multiples of 256 connections.

For additional technical description of connection mapping, see NN10600-702 *Nortel Multiservice Switch 7400/15000/20000 ATM Routing and Signalling Fundamentals*. For configuration information, see NN10600-710 *Nortel Multiservice Switch 7400/15000/20000 ATM Configuration Management*.

CQC buffer space configuration characteristics

The maximum number of connections is 4096 on ports 0 and 1, and 2560 on port 2. The default for each port is 1536. This restriction does not apply to 8-port DS1/E1 ATM FPs.

The default transmit allocation is 50 percent for frame memory and 50 percent for cell memory, measured in cells. The default receive allocation is 80 percent for frame memory and 20 percent for cell memory, again measured in cells.

Copyright © 2005, Nortel Nortel Confidential

The reason for the larger allocation on the receive direction is that cells from different frames may be interleaved, putting higher demands on receive buffering.

There is an absolute minimum amount of CQM that the system reserves on each FP for buffer space. For a configured value that is smaller than this minimum, the system allocates the minimum required over the configured amount.

A difference between the configured and the operational values may exist due to the requirement for a minimum number of frame buffers or rounding during computation within the FP.

Copyright © 2005, Nortel Nortel Confidential

Packet-wise discard

Nortel Multiservice Switch nodes apply discard and congestion control when incoming traffic is greater than the capacity of the outgoing link. The primary purpose of these controls is to ensure good throughput and delay performance while maintaining a fair allocation of network resources to each connection.

Navigation

- Overview of packet-wise discard (page 125)
- Overview of late packet discard (page 127)
- Overview of partial packet discard (page 128)
- Overview of early packet discard (page 130)
- High and low priority EPD offset (page 130)
- Overview of weighted random early detection (page 131)
- Applications for packet-wise discard (page 131)
- Congestion notification (page 134)
- Packet-wise discard for CQC-based FPs (page 135)
- Packet-wise discard for ATM IP FPs (page 138)
- Packet-wise discard for APC-based FPs (page 146)
- Packet-wise discard for GQM-based FPs (page 147)

Overview of packet-wise discard

There are four types of discard:

- late packet discard (LPD)
- partial packet discard (PPD)
- early packet discard (EPD)
- weighted random early detection (WRED)

Copyright © 2005, Nortel

Nortel Confidential

The table Application of packet-wise discard mechanisms (page 126) summarizes the application of packet-wise mechanisms by FP type and network point.

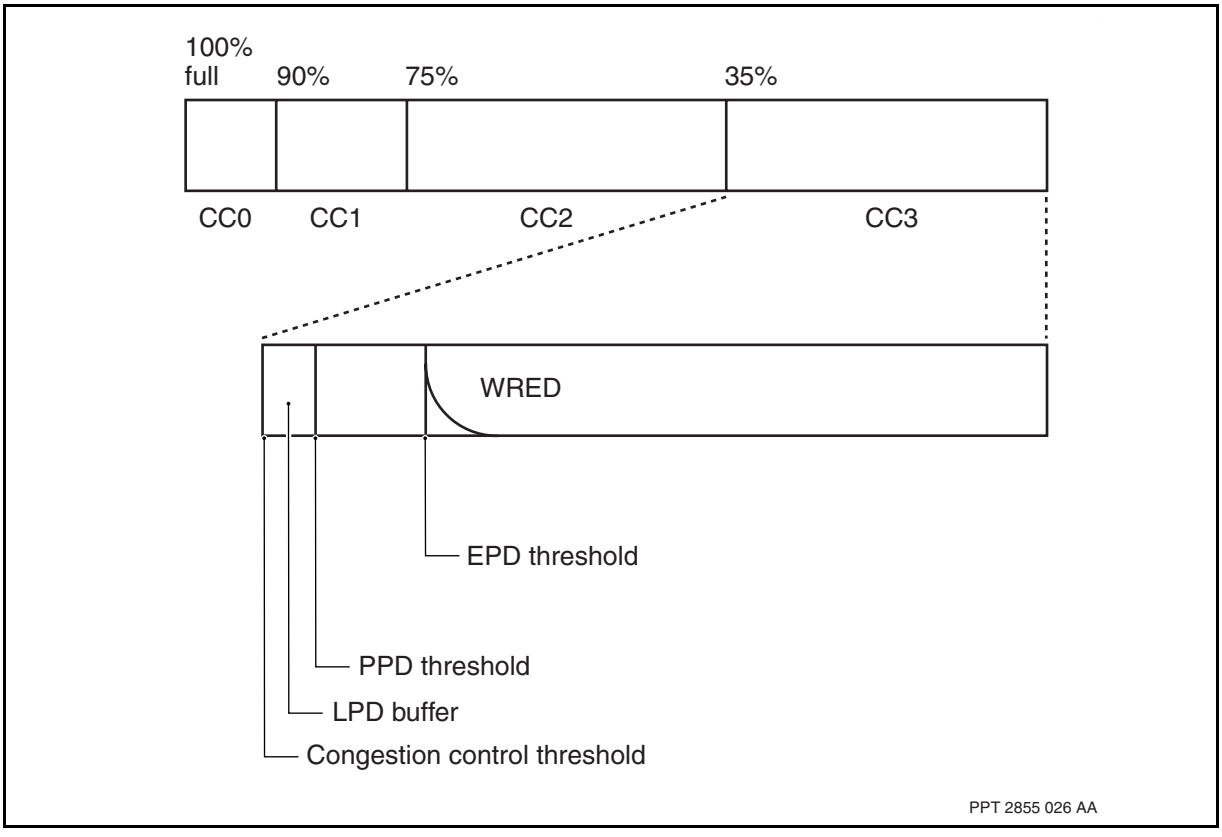
Application of packet-wise discard mechanisms

FP type	Relay point	End point / AAL5 segmentation and reassembly point
ATM IP	LPD, PPD, EPD, WRED	LPD, PPD, EPD, WRED
CQC	PPD	PPD, EPD

Discard thresholds are based on a combination of queue percentage and cell counts based on those percentages.

The figure Overview of packet-wise discard mechanisms (connection queues, no port aggregation) (page 126) shows how mechanisms apply to the congestion control levels for per-VC queues without port aggregation.

Overview of packet-wise discard mechanisms (connection queues, no port aggregation)



Copyright © 2005, Nortel
Nortel Confidential

Overview of late packet discard

LPD permits admission of specific types of cells above the threshold for PPD. These cells include

- an end of message (EOM) cell, where discarding the cell invalidates the frame that follows
- a cell that encapsulates acknowledgment frames (single cell frames), where loss of the acknowledgment frame requires retransmission of frames that the destination received correctly

In most cases, the node does not enqueue cells past the PPD threshold. However, enqueuing EOM and single-frame cells past the PPD threshold is beneficial for the following reasons:

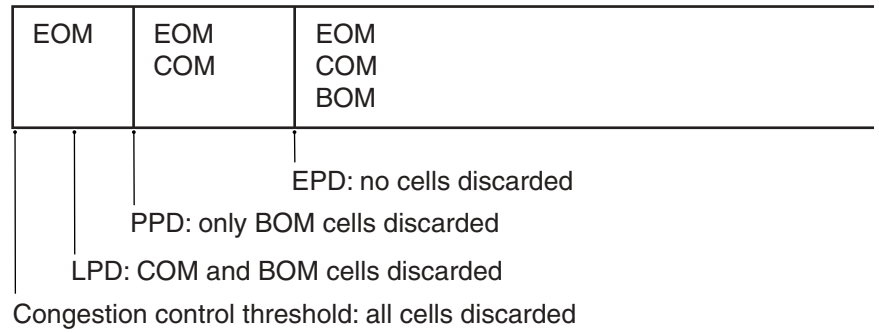
- ensures that EOM cells are available to demarcate frames across the network
- ensures that properly transmitted and received frames are not unnecessarily retransmitted because acknowledgment cells were discarded

LPD defines a small buffer of cells between the PPD threshold and the maximum queue length for the applicable congestion control level. When the queue length exceeds the PPD threshold but remains below the CC level threshold, the node queues only EOM cells and single cell frames. Eventually, if the queue length reaches the congestion control level threshold, the node also discards EOM cells and single cell frames. This condition indicates that network re-balancing is required in the area of this node.

The figure Discard pattern for EOM, COM, and BOM cells (page 128) shows how only EOM cells are permitted below the congestion control threshold and over the PPD threshold within a congestion control level.

Copyright © 2005, Nortel
Nortel Confidential

Discard pattern for EOM, COM, and BOM cells



Overview of partial packet discard

Partial packet discard (PPD) is a control that allows the node to discard cells that belong to frames that have had a cell discarded through congestion control. Both ATM IP and CQC-based FPs support PPD. The PPD threshold is set at

- LPD minus three cells for connection queues
- 99% of the congestion control level for the free list

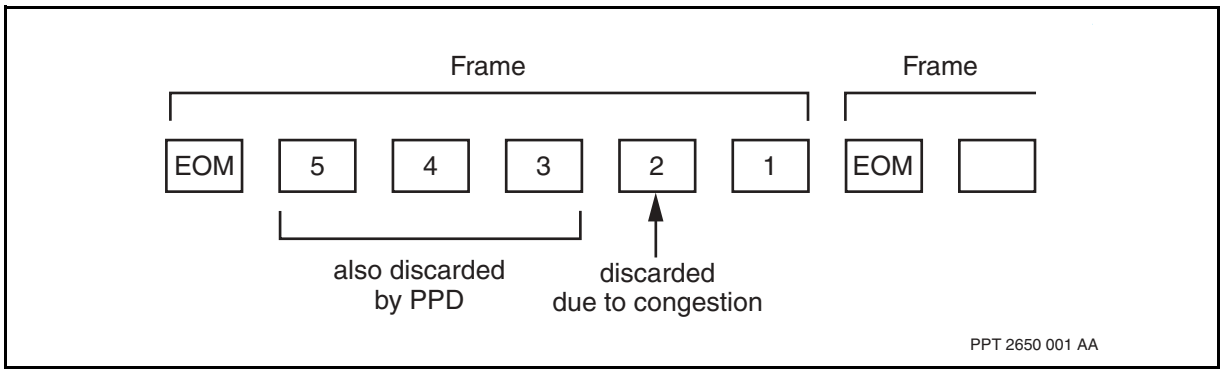
When the queue size reaches the PPD threshold, PPD allows the node to begin dropping cells in the middle of frames, the continuation of message (COM) cells. Once one cell of a frame is discarded, except for the beginning of message (BOM) cell, PPD continues to discard all cells of that frame up to but not including the EOM cell. The frame is already irreparably damaged, and these extra COM cells serve no useful purpose, but add to the load on the network. In situations with long frames and bursty traffic, PPD can substantially improve network goodput by discarding cells that are not required at the destination node.

The node does not discard the EOM cell (it is passed along untouched) to allow the AAL5 reassembly logic to recognize the end of this frame even though the frame contents are damaged. Through the EOM cell, the AAL5 reassembly can also detect the beginning of the next frame and only a single frame is lost. Under PPD logic, if the first cell of a frame, the BOM cell, is discarded, all cells up to and including the EOM cell are also discarded.

Copyright © 2005, Nortel
Nortel Confidential

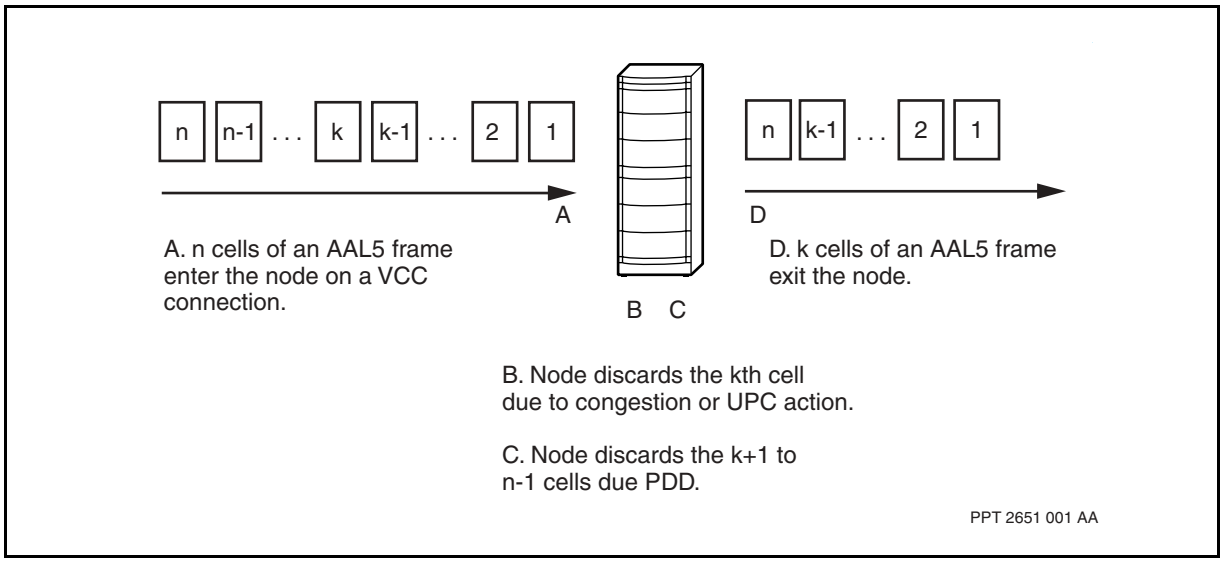
The figure Example of partial packet discard (page 129) gives a simple example of cell discard under PPD. The figure Discard pattern for EOM, COM, and BOM cells (page 128) shows how only EOM and COM cells are permitted below the PPD threshold and above the EPD threshold within a congestion control level.

Example of partial packet discard



The figure Partial packet discard functionality (page 129) illustrates how PPD works. Observe that after PPD applies, Nortel Multiservice Switch nodes still transmit the cells in sequence before the damaged cell: that is, the network still carries some cell traffic that the destination node discards.

Partial packet discard functionality



PPD is configured by enabling or disabling packet-wise discard for a VCC or VPC. You can enable PPD only on either of the following:

- a nailed-up relay point (NRP)

Copyright © 2005, Nortel
Nortel Confidential

- a relay point (RP) on which the connection's end-point performs AAL5 segmentation and reassembly

Overview of early packet discard

Early packet discard (EPD) allows the node to drop entire multi-cell frames when the queue level reaches the EPD threshold for the current congestion control level. Both ATM IP and CQC-based FPs support EPD.

When enabled, the setting for the EPD threshold is relative to the PPD threshold and depends on the service category:

- EPD threshold is 50 cells lower than PPD for connection queues that support CBR and RT-VBR
- EPD threshold is 200 cells lower than PPD for connection queues that support NRT-VBR and UBR
- 85% of the congestion control level for the free list

The benefit of EPD is to preserve buffer space so that the node can complete transmission of the remaining cells for frames already in process. By discarding entire frames, the node maintains buffer capacity for frames that are already partially transmitted. The goal is to ensure that these frames complete transmission before total congestion occurs. This result, in turn, reduces the number of partial frames transmitted across the network (which the destination node discards). In situations with long frames and bursty traffic, EPD can substantially improve network goodput by discarding cells that are not required at the destination node.

The figure Discard pattern for EOM, COM, and BOM cells (page 128) shows how EOM, COM, and BOM cells are permitted below the EPD threshold within a congestion control level.

The exceptions to the EPD process involve single-cell frames. The node continues to queue single-cell frames past the EPD threshold, since these frames are often acknowledgment (ACK) frames. The loss of an ACK frame results in needless retransmission of several frames which the destination has successfully received. See Overview of late packet discard (page 127).

EPD is configured by enabling or disabling packet-wise discard for a VCC or VPC.

High and low priority EPD offset

For per-VC queuing, EPD thresholds for each CC level are specified by means of the EPD offset from the corresponding CC level. The Nortel Multiservice Switch node uses this offset value to derive a connection EPD threshold for the CC level by subtracting the offset from the CC level threshold. The CC level threshold is also referred to as an all packet discard (APD) threshold. As

Copyright © 2005, Nortel Nortel Confidential

described in Overview of partial packet discard (page 128) and Overview of early packet discard (page 130), the CC level threshold is derived from the queue limit.

There are two configurable values: the high priority EPD offset and the low priority EPD offset. These values apply to all connections on the FP. However, each connection may have a different queue limit. For some connections, the queue limit may be too small to support the EPD offset. In these cases, software disables both PPD and EPD for the connection. By reducing the EPD offset through configuration, packet-wise discard is enabled.

This offset for per-VC queuing is configurable for queues as follows:

- one configuration for the exclusive VBR shaper emission priority and the first four (higher priority) unshaped emission priorities; this configuration is the high priority EPD offset
- one configuration for the ABR/VBR shaper and the last two (lower priority) unshaped emission priorities; this configuration is the low priority EPD offset

The connection queue limit is too small to support EPD if the queue limit is less than

(5.71 * EPD_Offset)

This requirement ensures that the derived EPD threshold is never less than 17.5% of the queue limit. The congestion control level 3 (CC3) at 35% of queue limit must be at least twice the EPD offset).

Because of these requirements, you must reconfigure the EPD offsets if you want to decrease the queue limit and still enable transmit packet-wise discard. The EDP offsets are configured through the ATM resource control.

Overview of weighted random early detection

Weighted random early detection (WRED) is a congestion avoidance mechanism that breaks the synchronization effect of multiple TCP sessions. WRED is supported only on Nortel Multiservice Switch 15000 and Multiservice Switch 20000 4-Port OC3 and QRD-based FPs, and Multiservice Switch 7400 2-Port OC3 and MSA32mtp FPs.

For more information, see Packet-wise discard for ATM IP FPs (page 138).

Applications for packet-wise discard

This section provides description of some applications for packet-wise discard. Applications include:

- Partial packet discard at the cell relay point (page 132)

Copyright © 2005, Nortel
Nortel Confidential

- Partial packet discard for AAL5 connections (page 132)
- Intelligent discard at an AAL5 adaptation point (page 133)

Partial packet discard at the cell relay point

PPD at cell relay points discards cells when there are problems with network traffic or cell integrity. This process helps optimize network performance.

Nortel Multiservice Switch nodes discard cells at an interface in response to any of a number of factors, such as congestion, policing through UPC, or problems with data integrity. For example, if one cell in a frame is damaged, the entire frame is damaged and the AAL5 reassembly logic at the destination node discards that frame. By dropping all cells for the frame that follow the lost cell (except the End of Message cell), Multiservice Switch nodes do not transmit (and the network does not carry) cells that AAL5 reassembly logic discards at the destination node anyway. As a result, the bandwidth that the network would have used to carry the cells of a useless frame is available to carry the cells of a useful frame.

Partial packet discard for AAL5 connections

Packet-wise discard is set independently for both the transmit and receive directions. There are differences in application that depend on the FP type.

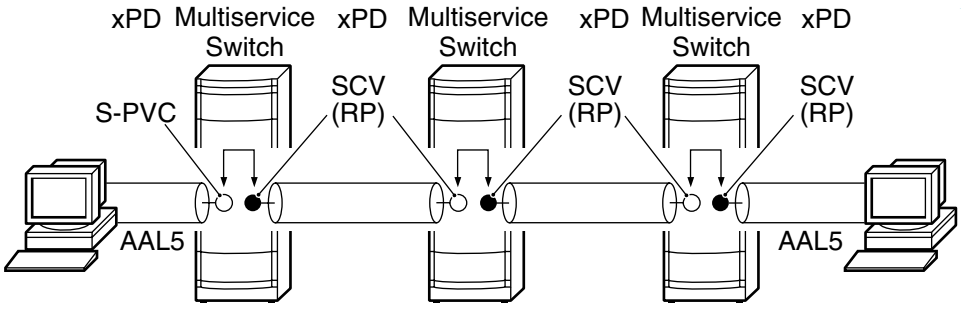
In general, Nortel Multiservice Switch nodes are able to differentiate between AAL5 connections and all other connections. This difference, however, is not visible through operational attributes; if packet-wise discard is enabled through configuration, the operational attribute indicates that it is enabled even if the connection does not carry AAL5 traffic. Nodes enables packet-wise discard only if AAL5 traffic is detected and then only is enabled through configuration.

If packet-wise discard is enabled at an SPVC origin on a PNNI interface, frame discard is requested in the forward direction. When disabled, forward frame discard is not requested in the SPVC call setup.

See the figure Enabling PPD for SPVCs (page 133).

Copyright © 2005, Nortel
Nortel Confidential

Enabling PPD for SPVCs



xPD is one more of LPD, EPD, or PPD. The combination if discard mechanisms depends on the function processor

MSS 2855 028 AA

The following signaling protocols permit PPD for all AAL5 connections in both the transmit and receive directions:

- UNI 3.0
- UNI 3.1
- IISP 1.0v3.x
- IISP 1.0v3.1

Intelligent discard at an AAL5 adaptation point

At an AAL5 adaptation point, Nortel Multiservice Switch nodes support both partial packet discard (PPD) and early packet discard (EPD). Compared to PPD, the EPD process applies a more stringent discard policy to the first cell of an AAL5 frame than to subsequent cells. As a result, EPD reduces the number of incomplete frames in the network and thereby increases goodput. Multiservice Switch software enables PPD and EPD by default at a connection end point that performs AAL5 segmentation and assembly.

The combination of these intelligent packet/frame discard features eliminates frame fragmentation entirely at the adaptation points. The result is high network throughput.

Packet-wise discard and VPTs

Packet-wise discard can be enabled only at the VCC level under a VPT. It cannot be enabled at the VPT level. This characteristic offers greater flexibility, since discard controls can be set independently for each VCC under the VPT. Configuration at the VPT level incurs the restriction of enabling or disabling discard control for all VCCs as a group.

Copyright © 2005, Nortel
Nortel Confidential

Congestion notification

Congestion notification is a traffic management mechanism through which the network signals congestion in the network to subscriber end systems. End systems may respond by reducing the amount of transmitted data, which in turn reduces demands on the network. Congestion notification is a mechanism that helps clear congestion over time.

Overview to congestion notification

Nortel Multiservice Switch nodes have two strategies for congestion notification:

- explicit forward congestion indication (EFCI) header insertion, for ATM node-to-node notification
- EFCI-FCI mapping, for ATM traffic to trunk traffic notification

EFCI is a cell-level congestion indication mechanism. The benefit of EFCI is to notify end systems about impending congestion in the network so that measures can be taken to avoid further severe congestion. EFCI marking involves setting the EFCI bit in the ATM cell header when queue congestion reaches the EFCI threshold.

EFCI header insertion

By using the EFCI field in the cell header, Nortel Multiservice Switch networks pass congestion notifications to the end-systems at the access points. As a cell traverses a node, the node sets the EFCI bit as summarized in the table Policy for setting EFCI (page 134).

Once the EFCI bit is set, it remains set for the lifetime of the cell in the Multiservice Switch network.

Policy for setting EFCI

Transmit/receive queues	Setting EFCI
common queue (link transmit)	EFCI in the transmit direction is automatically set by hardware when the common queue congestion is less than or equal to CC2.
per-VC queues (link transmit)	EFCI in the transmit direction is automatically set by hardware when the per-VC queue congestion is less than or equal to CC2.
processor queues (receiving from link or backplane)	No EFCI setting
backplane transmit queues	EFCI is automatically set by hardware when the backplane queue congestion is less than or equal to CC2.
(1 of 2)	

Copyright © 2005, Nortel
Nortel Confidential

Policy for setting EFCI (continued)

Transmit/receive queues	Setting EFCI
free list exhaustion	EFCI is automatically set by hardware when the free-block congestion is less than or equal to CC2.
The node marks EFCI when there is congestion and does not take into consideration the severity of congestion. Because Multiservice Switch nodes define CC2 as the point at which congestion occurs, EFCI marking occurs at these points. CC3 implies no congestion.	
(2 of 2)	

EFCI-FCI mapping

Nortel Multiservice Switch nodes also forward congestion indication from ATM traffic to trunk traffic using EFCI-FCI mapping. FCI applies to frames that nodes send over the Multiservice Switch trunk protocol.

Multiservice Switch network elements set a forward congestion indication (FCI) in a frame header to indicate pending or current network congestion. Mapping occurs when an incoming AAL5 frame with EFCI inserted into at least one of its cells is converted into Multiservice Switch node frames and destined for another FP on the shelf. For example, EFCI-FCI mapping occurs when sending frames over a trunk mapped to an ATM connection. Mapping occurs automatically (no configuration required).

Multiservice Switch node PORS and DPRS use mapping to affect close loop control at the service or connection level. Frame relay does not use FCI. Nodes use EFCI to map to the forward explicit congestion notification (FECN) and backward explicit congestion notification (BECN).

Packet-wise discard for CQC-based FPs

CQC-based FPs provide three packet-wise discard mechanisms that the node triggers within each of the congestion control levels for each queue.

This chapter presents information in the following sections:

- Characteristics of CQC packet-wise discard (page 135)
- EFCI on CQC-based FPs (page 137)

Characteristics of CQC packet-wise discard

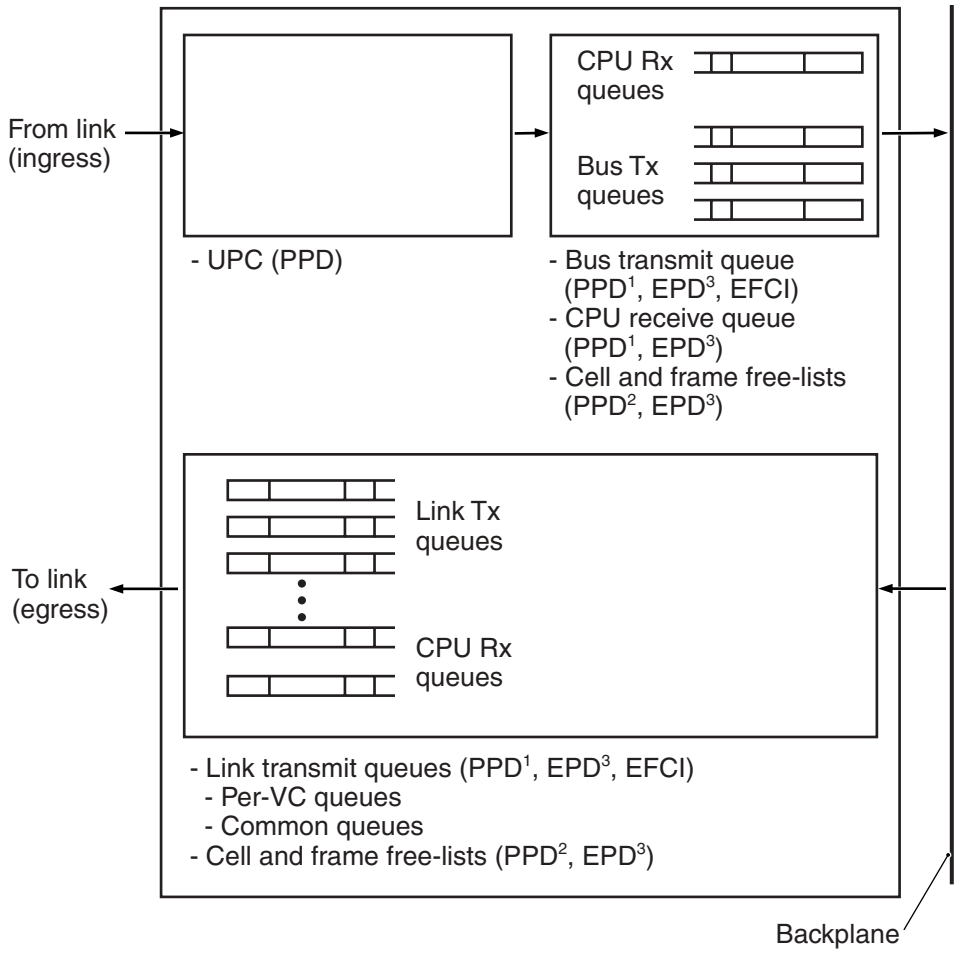
CQC-based FPs use the following packet-wise discard mechanisms:

- late packet discard (LPD)
- partial packet discard (PPD)
- early packet discard (EPD)

Copyright © 2005, Nortel
Nortel Confidential

On CQC-based FPs, end-point EPD is automatically enabled whenever frame-to-cell segmentation is performed. LPD is also present, such that the LPD threshold equals the PPD threshold. The figure Application points for packet-wise discard and EFCI: CQC-based FP (page 136) shows where packet-wise discard mechanisms apply.

Application points for packet-wise discard and EFCI: CQC-based FP



Note 1: For the PQC queues, the PPD feature is optional, through configuration, for ATM cell relay points.
Note 2: For the PQC free lists, the PPD feature always applies for AAL5 adaptation points, and is optional, through configuration, for cell relay points.
Note 3: In PQC, the EPD feature applies only to AAL5 adaptation points and is always turned on.

PPT 2855 031 AA

Copyright © 2005, Nortel
Nortel Confidential

LPD on CQC-based FPs

LPD is present on CQC-based FPs so that the node can buffer end-of-message (EOM) cells for improved goodput. The LPD buffer defines the PPD threshold as follows:

CC level (in cells) - LPD buffer (in cells) = PPD threshold

PPD at cell relay points: CQC variant

CQC-based FPs support only PPD on a cell relay connection (that is, on a connection between interfaces that do not provide AAL5 segmentation and reassembly). PPD is available in one of the following ways:

- configured through the forward and backward frame discard parameters
- signaled through the AAL-type equals AAL5

You can configure PPD for both the transmit and receive directions. See NN10600-710 *Nortel Multiservice Switch 7400/15000/20000 ATM Configuration Management*.

Enabling PPD on a non-AAL5 connection could cause discard of all traffic whenever congestion is encountered.

If PPD is set up on a PVC, then PPD is applicable only to the associated connecting point. PPD has no effect along a VPC.

PPD for AAL5 connections over CQC-based FPs

On CQC-based FPs, PPD is available at VCC cell transfer points. PPD is not available at frame-cell conversion points, and does not apply to VPCs. You can enable packet-wise discard only for connections that carry AAL5 segmentation traffic. If you enable packet-wise discard for connections that carry other traffic types, traffic discard may occur. On CQC-based FPs, EPD is automatically enabled at frame-cell conversion points and is independent of configured settings.

EPD on CQC-based FPs

On CQC-based ATM FPs, end-point EPD is automatically enabled whenever frame-to-cell segmentation is performed.

EFCI on CQC-based FPs

In CQC-based ATM FPs, EFCI marking applies to cells in the following queues:

- common queue
- per-VC queue
- Nortel Multiservice Switch node's bus queue

Copyright © 2005, Nortel
Nortel Confidential

EFCI marking is always performed. The threshold value is set to 35% of the total queue length. There are no statistical attributes to view.

Packet-wise discard for ATM IP FPs

ATM IP FPs provide four packet-wise discard mechanisms that the node triggers within each of the congestion control levels for each queue.

The following topics are discussed in this section:

- Characteristics of ATM IP packet-wise discard (page 138)
- LPD on ATM IP FPs (page 142)
- EPD on ATM IP FPs (page 142)
- PPD on ATM IP FPs (page 142)
- WRED on ATM IP FPs (page 143)
- AAL5 auto-detection (page 144)
- Configuring packet-wise discard for ATM IP (page 145)
- EFCI on ATM IP FPs (page 145)

Characteristics of ATM IP packet-wise discard

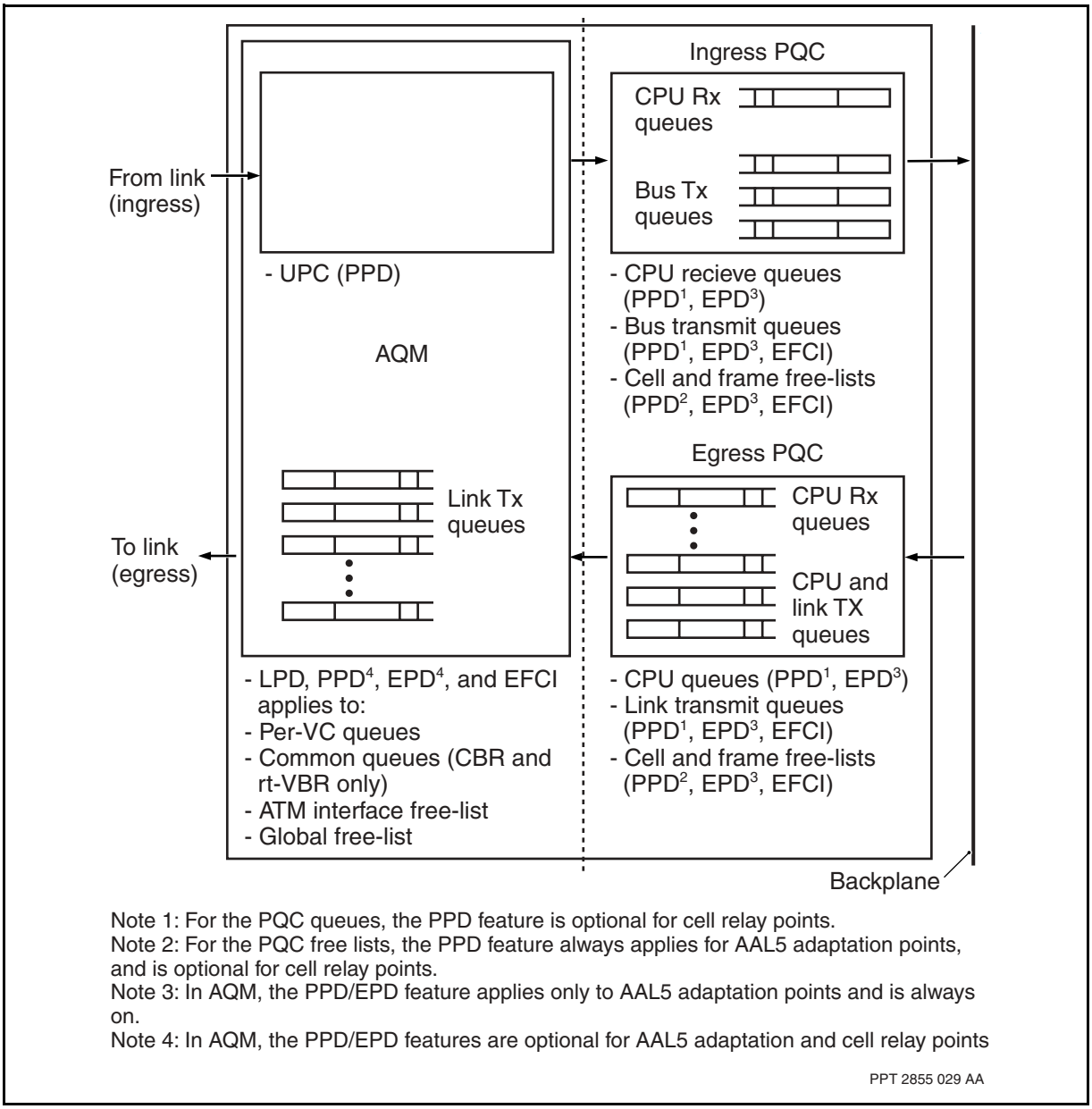
ATM IP FPs use the following packet-wise discard mechanisms:

- late packet discard (LPD)
- partial packet discard (PPD)
- early packet discard (EPD)
- weighted random early detection (WRED)

The figure Application points for packet-wise discard and EFCI: ATM IP FP (page 139) shows where packet-wise discard mechanisms apply.

Copyright © 2005, Nortel
Nortel Confidential

Application points for packet-wise discard and EFCI: ATM IP FP



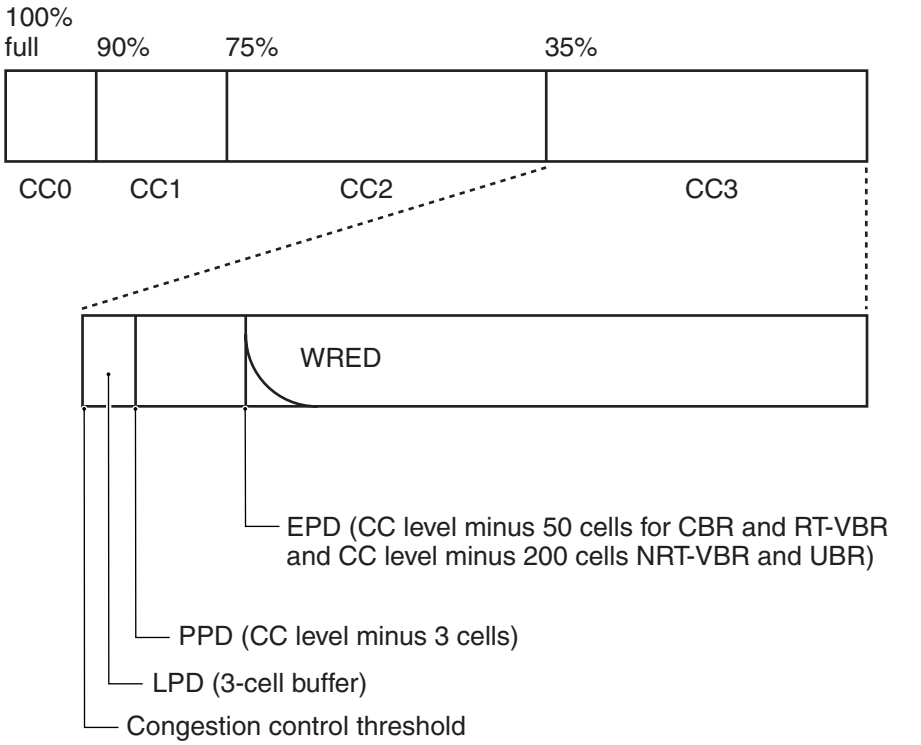
Note 1: For the PQC queues, the PPD feature is optional for cell relay points.
Note 2: For the PQC free lists, the PPD feature always applies for AAL5 adaptation points, and is optional for cell relay points.
Note 3: In AQM, the PPD/EPD feature applies only to AAL5 adaptation points and is always on.
Note 4: In AQM, the PPD/EPD features are optional for AAL5 adaptation and cell relay points

PPT 2855 029 AA

The figure ATM queue manager packet-wise discard mechanisms: connection queues (page 140) shows how packet-wise discard mechanisms apply within a congestion control level for a connection queue without port aggregation. The figure ATM queue manager packet-wise discard mechanisms: free list (page 141) shows how packet-wise discard mechanisms apply within a congestion control level for the free list.

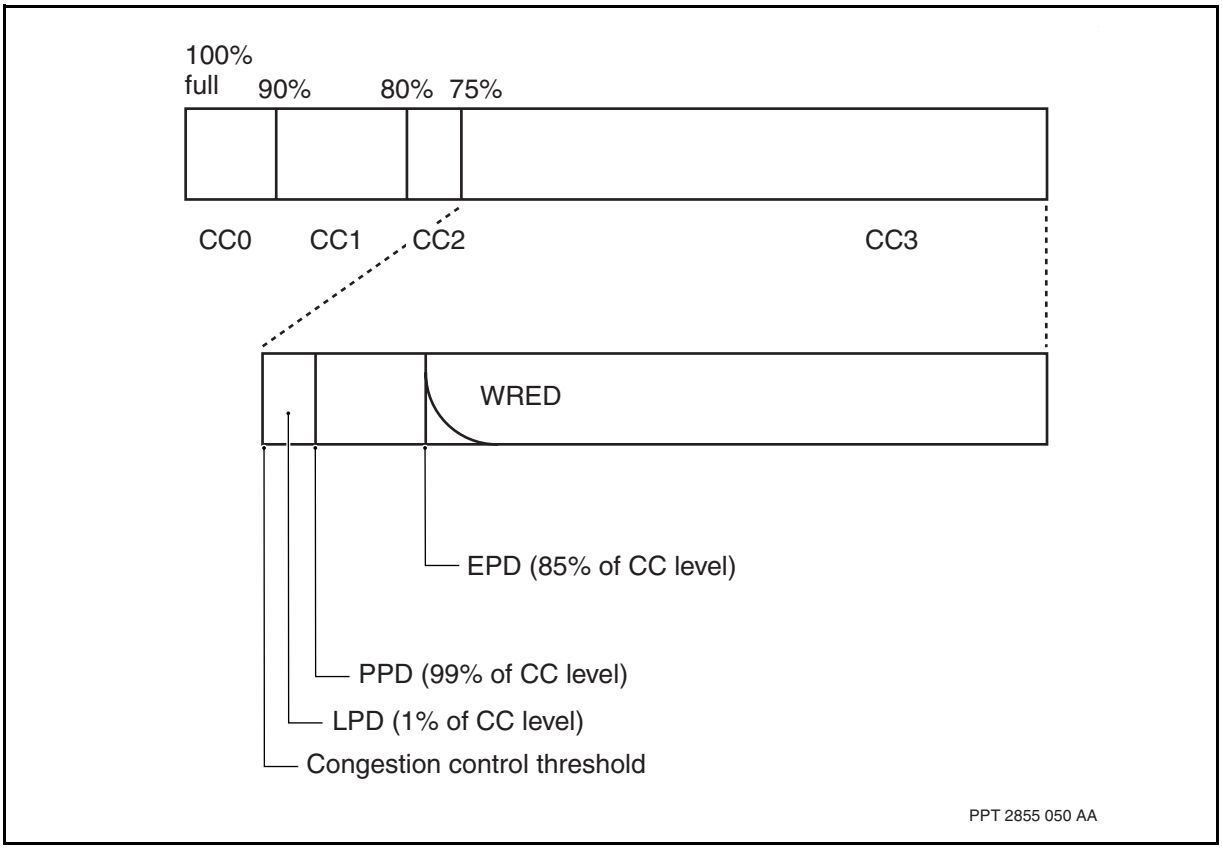
Copyright © 2005, Nortel
Nortel Confidential

ATM queue manager packet-wise discard mechanisms: connection queues



Copyright © 2005, Nortel
Nortel Confidential

ATM queue manager packet-wise discard mechanisms: free list



Nortel Multiservice Switch nodes replicate this arrangement of packet-wise discard mechanisms for all congestion control levels for all queues in the ATM queue manager (AQM).

User traffic maps to discard priority 1, 2, and 3, depending on the node configuration and the connection cell loss priority (CLP) setting. Explicit forward congestion indication (EFCI) marking occurs at 35% of the maximum transmit queue length (independently of the congestion control thresholds). EPD and PPD functions occur at specific points relative to each congestion control level.

Example of packet-wise discard on ATM IP

In this example, assume that traffic for service category RT-VBR CLP0 maps to discard priority 1, and EPD for these cells occurs at the congestion control level 1 (CC1) minus 50 cells. Traffic for service category NRT-VBR CLP0 maps to discard priority 2, and EPD for these cells occurs at the congestion control level 2 (CC2) minus 200 cells.

Copyright © 2005, Nortel
Nortel Confidential

For the connection queues, EPD occurs at a specific threshold that is relative to the PPD threshold (see Overview of early packet discard (page 130)). For the ATM interface free list, the EPD threshold is 85% of the threshold for each congestion control level.

The figure ATM queue manager packet-wise discard mechanisms: connection queues (page 140) shows the application of packet-wise discard in relation to a per-VC queue filling. On ATM IP FPs, LPD, PPD, EPD, and WRED apply to the following buffer allocations:

- per-VCC within a VPT
- VCC connection (as the figure ATM queue manager packet-wise discard mechanisms: connection queues (page 140) shows)
- tandem VPC connection point (as the figure ATM queue manager packet-wise discard mechanisms: connection queues (page 140) shows)
- ATM interface free lists

LPD on ATM IP FPs

On ATM IP FPs, LPD is available at frame segmentation and reassembly points and at tandem switches. You configure LPD by enabling or disabling packet-wise discards for a VCC or VPC connection. If enabled, the LPD limit is three cells higher than the PPD threshold for connections (three cells less than the CC level). For the free list, LPD is set at one percent of the congestion control level. See the figures ATM queue manager packet-wise discard mechanisms: connection queues (page 140) and ATM queue manager packet-wise discard mechanisms: free list (page 141).

EPD on ATM IP FPs

In ATM IP FPs, EPD is configurable at frame SAR points and at tandem switches. Once the queue reaches the EPD threshold for the current congestion control level, the node discards the next BOM cell for frames with the corresponding discard priority. The node also discards all subsequent cells belonging to that frame, including the EOM cell. This discard process continues until the queue level falls below the EPD threshold for the current congestion control level.

PPD on ATM IP FPs

In ATM IP FPs, PPD is available at frame SAR points as well as at tandem switches. ATM IP FPs also have a PPD auto-detect feature.

Single-cell frames are still queued past the PPD level, since single cell frames are often TCP ACK frames. The loss of an ACK frame would mean that the TCP level would needlessly retransmit several frames which have been successfully received.

Copyright © 2005, Nortel
Nortel Confidential

PPD is configured by enabling or disabling packet-wise discards for a VCC or VPC connection. PPD may be configured independently in the transmit and receive direction. If enabled, the PPD threshold is equal to the CC level threshold minus the LPD offset.

PPD and LRC errors

On transmit, the Nortel Multiservice Switch node sets the AAL5 frame length to zero any time an LRC error is detected for that given frame. LRC errors may occur when PPD in previous nodes in the connection discards cells belonging to the same node frames covered by LRC. LRC errors can also occur between node FPs over the backplane, an these errors can also cause the node to set AAL5 frame length to zero.

The PPD has no direct influence on the generation of AAL5 frames with a length set to zero. On receive, the node counts the number of aborted frames. If the node is connected to the switch of a third-party vendor, the abort may results from causes other than LRC errors.

PPD for AAL5 connections over ATM IP FPs

On ATM IP FPs, PPD and EPD apply to individual VCCs and to VCCs within VPCs at all connection points, including tandem VPC connections. Both PPD and EPD in the transmit and receive directions are configurable. You can enable packet-wise discard for any connection (VCC or VPC) on ATM IP FPs since these FPs can automatically detect AAL5-segmented traffic on a connection. An ATM IP FP enables packet-wise discards only if it detects AAL5-segmented traffic.

WRED on ATM IP FPs

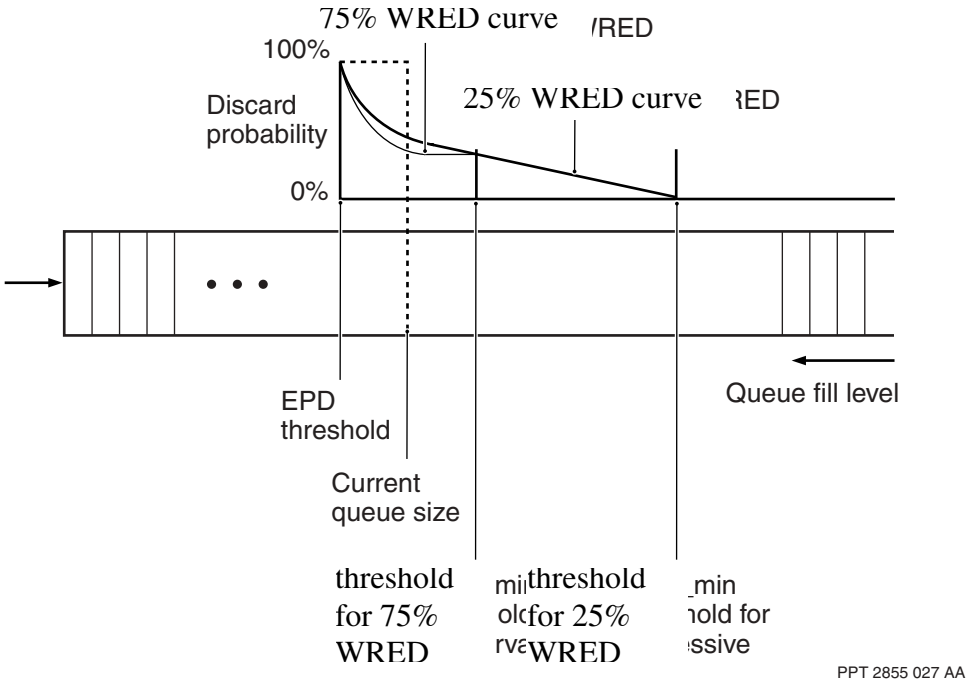
WRED is a congestion avoidance mechanism on a per VC basis within the CC3 level that breaks the synchronization effect of multiple TCP sessions. WRED is supported only on Nortel Multiservice Switch 15000 and Multiservice Switch 20000 4-port OC3 nodes, and Multiservice Switch 7400 2-port OC3 nodes, and MSA32mtp FPs. WRED improves overall network goodput. While WRED is more important to free list first-in first-out (FIFO) exhaustion than for per-VC queuing, it is still useful in per-VC queuing where multiple TCP sessions are multiplexed onto a single virtual circuit.

In WRED, as the queue size grows beyond a minimum threshold, the ATM IP FP can optionally drop arriving packets with a probability that increases as the queue size increases. Typically, WRED begins to drop packets with very low probability at a low queue level, and increases to probability 1 as queue size reaches the EPD threshold. WRED on ATM IP FPs, as the figure ATM IP FP WRED mechanism (page 144) shows, behaves like a randomizing EPD. As with EPD, the node discards all cells for the frame.

Copyright © 2005, Nortel
Nortel Confidential

Furthermore, WRED offers an additional capability such that different connections may have different weighted dropping probabilities. WRED applies to per-VC queuing. Operators can disable WRED, or configure WRED aggressiveness as one of three values from most conservative to most aggressive. The minimum WRED threshold is set to either 25%, 50% or 75% of the EPD level. This setting defines WRED aggressiveness (most aggressive, medium aggressive or most conservative). The most aggressive setting activates WRED when the queue level reaches 25% of the EPD threshold. The most conservative setting waits for a higher queue level (75% of the EPD threshold) before WRED is active.

ATM IP FP WRED mechanism



AAL5 auto-detection

ATM IP FPs auto-detect at any cell relay point that a given VCC is carrying AAL5-segmented cells. The operator can request packet-wise discard functions for a given connection. However, these discard functions are enabled only if the AAL5 auto-detect indicates that this VCC is carrying AAL5 traffic.

In addition, ATM IP FPs provide the ability to detect AAL5 segmentation occurring on VCCs within a VPC, even when the VCCs are not separately configured at this link, for example at a VPC tandem switching point. Again, as with VCCs, packet-wise discards are enabled only if AAL5 auto-detect discovers an AAL5 VCC.

Copyright © 2005, Nortel
Nortel Confidential

Configuring packet-wise discard for ATM IP

EPD, PPD and LPD are turned on or off at the same time through a AAL5 packet-wise discard option configured on a per connection basis or signalled in the call setup. Packet-wise discard may be enabled or disabled in the transmit or receive direction, independently. As well, packet-wise discard functions are only activated when AAL5 traffic is detected on that VCC or on a VCC within that VPC.

End-point EPD (EPD at a frame-to-cell segmentation and reassembly point) is available only if configured.

WRED can be turned on only when packet-wise discard is enabled. Further, WRED can be disabled even when packet-wise discard is enabled. The aggressiveness of the WRED function is provisioned on a per-VC basis.

In Nortel Multiservice Switch nodes, WRED aggressiveness is provisioned by two attributes: *txWredMode* and *txWredThreshold* under AtmIf/n Vcc/x.y [Vpt/x Vcc/y] Vcd Tm. The words conservative, medium aggressive, and aggressive are not visible to the user. Therefore, to configure WRED

- to disabled, you need to provision *txWredMode* as disabled
- to conservative, you need to provision *txWredMode* as enabled and txWredThreshold as 75
- to medium aggressive, you need to provision *txWredMode* as enabled and txWredThreshold as 50
- to aggressive, you need to provision *txWredMode* as enabled and txWredThreshold as 25

EFCI on ATM IP FPs

ATM IP FPs perform EFCI marking for the following queues:

- link common queue
- per-VC queue
- Nortel Multiservice Switch node's bus queue
- cell and frame free lists

EFCI marking is always performed. The threshold value is set to 35% of the total queue length. Statistics are available at the ATM interface level for ATM IP FPs. EFCI statistics are not available for CQC-based FPs. ATM IP FPs exchange congestion information with FR-UNI FPs. For example, ATM IP FPs can immediately report congestion in the transmit queue congestion in the opposite direction toward the FR-UNI FP. This feature leads to faster feedback when compared with older methods of waiting for the return of the BECN signal from the far-end frame relay source.

Copyright © 2005, Nortel
Nortel Confidential

Packet-wise discard for APC-based FPs

Network congestion occurs when incoming traffic exceeds outgoing link capacity. To control congestion on APC-based FPs, AAL5 cells are discarded to free up the needed buffer space through the following packet-wise discard mechanisms:

- Partial packet discard for APC-based FPs (page 146)
- Early packet discard for APC-based FPs (page 146)
- Weighted random early detection for APC-based FPs (page 146)

In addition to the three packet-wise discard mechanisms, Nortel Multiservice Switch nodes use thresholding for CLP0 and CLP1 cell types based on ATM service categories. For more information, see Per-VC queuing on APC/PQC-based FPs (page 87).

PPD, EPD, and WRED are enabled or disabled on a per-connection basis. The default setting is disabled. If the connections are not AAL5 connections, packet-wise discard must be disabled. If enabled on non-AAL5 connections, cell discard will occur after its queue length reaches its CLP0+1 threshold and will continue indefinitely.

Partial packet discard for APC-based FPs

PPD is a congestion control mechanism that allows the node to discard all incoming cells belonging to incomplete frames, except for the EOM cell. This avoids the waste of buffer space by incoming AAL5 cells belonging to frames that are already corrupted.

Early packet discard for APC-based FPs

For the APC-based FPs, the EPD threshold is set to be 85% of the CLP0+1 threshold.

Weighted random early detection for APC-based FPs

WRED is a congestion control mechanism that allows a node to discard AAL5 cells when a minimum buffer threshold is exceeded. The minimum WRED threshold is set to 25% of the EPD level.

In WRED, as the queue size grows beyond a minimum threshold, the APC or PQC-based FP can optionally drop arriving packets with a probability that increases as the queue size increases. Typically, WRED begins to drop packets with very low probability at a low queue level, and increases to probability 1 as queue size reaches the EPD threshold. Given a certain queue fill beyond the minimum threshold, AAL5 connections belonging to different ATM service categories can have different discard probabilities (different weights) of being discarded. As with EPD, the node discards all cells for the frame.

Copyright © 2005, Nortel Nortel Confidential

Packet-wise discard for GQM-based FPs

GQM-based FPs use packet-wise discard mechanisms the same as other Nortel Multiservice Switch node ATM FPs and include:

- early packet discard (EPD)
- late packet discard (LPD)
- partial packet discard (PPD)
- weighted-random early detection (W-RED)

The W-RED operates similarly to the way the 4-port OC-3/STM-1 ATM FP uses it. A single W-RED curve (or discard probability function) is available but each connection may specify the W-RED thresholds at 25%, 50%, or 75% of the EPD threshold. A W-RED threshold is where non-zero probability begins and increases as the buffer occupancy grows. A W-RED threshold of 25% is the most aggressive since discards due to W-RED begin sooner.

W-RED thresholds can only be specified for connections which use per-VC queuing. Common queued connections may still enable W-RED, but the W-RED threshold is fixed at 25%.

User control of packet-wise discard capabilities apply only to the transmit direction.

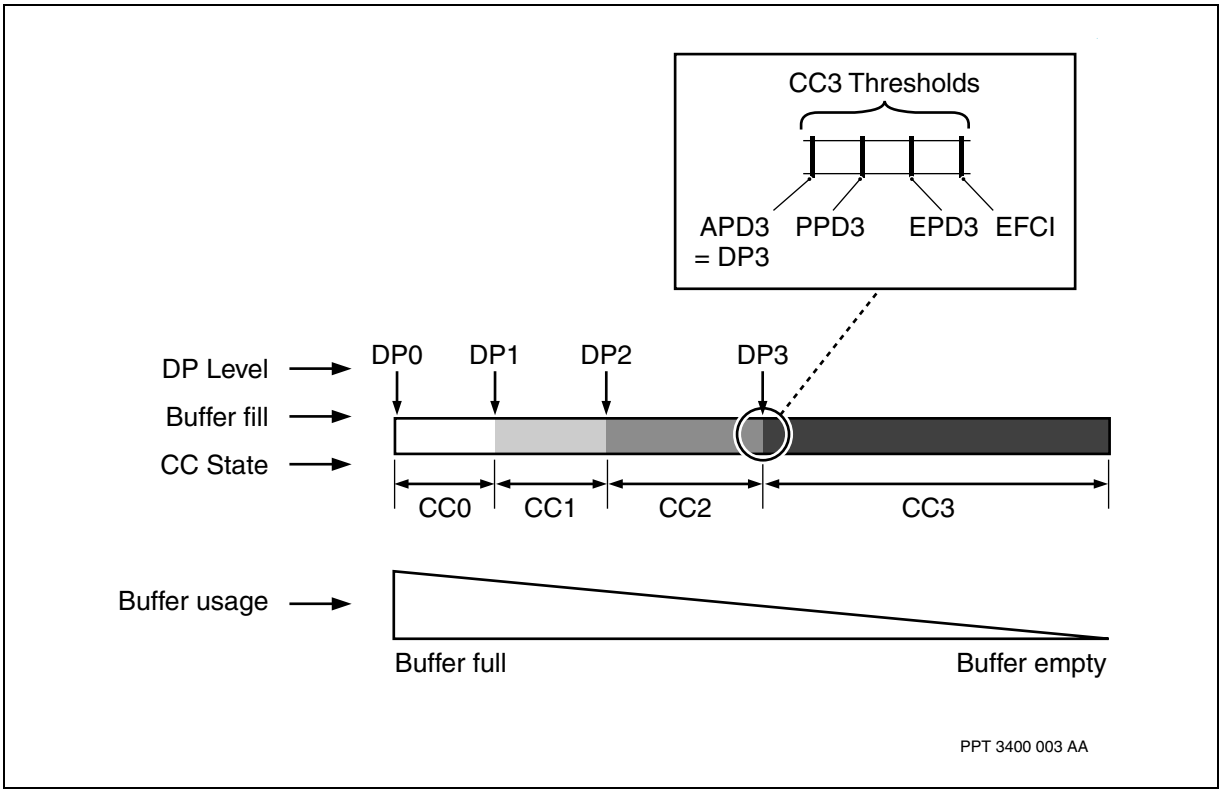
Packet-wise discard on VPs is not supported. GQM-based FPs cannot distinguish different VCs inside a VP to perform packet-wise discards on individual VCs inside the VP.

Packet-wise discard on VCs is only supported at the egress end. Ingress packet-wise discard is not supported on GQM-based FPs.

Refer to the figure Discard priority thresholds and congestion control states for GQM-based FPs (page 148) for the queue congestion control levels.

Copyright © 2005, Nortel
Nortel Confidential

Discard priority thresholds and congestion control states for GQM-based FPs



Copyright © 2005, Nortel
Nortel Confidential

Nortel Multiservice Switch 7400/15000/20000

ATM Queuing and Scheduling Fundamentals

Copyright © 2005 Nortel.
All Rights Reserved.

Publication: NN10600-707
Document status: Standard
Document issue: 7.1S1
Document date: October 2005
Product release: PCR7.1 and up
Job function: Product Fundamentals
Type: NTP
Language type: U.S. English

NORTEL, the globemark design, and the NORTEL corporate logo are trademarks of Nortel.