

Quality of Service Models

Best Effort · No QoS policies are implemented

Integrated Services (IntServ) · Resource Reservation Protocol (RSVP) is used to reserve bandwidth per flow across all nodes in a path

Differentiated Services (DiffServ) · Packets are individually classified and marked; policy decisions are made independently at each node in a path

Layer 2 QoS Markings

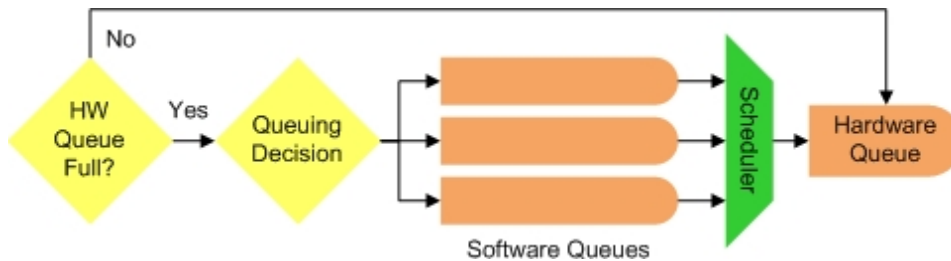
Medium	Name	Type
Ethernet	Class of Service (CoS)	3-bit 802.1p field in 802.1Q header
Frame Relay	Discard Eligibility (DE)	1-bit drop eligibility flag
ATM	Cell Loss Priority (CLP)	1-bit drop eligibility flag
MPLS	Experimental Field (EXP)	3-bit field compatible with 802.1p

IP QoS Markings

Precedence · The first three bits of the IP TOS field are evaluated; compatible with Ethernet CoS and MPLS EXP values

DSCP · The first six bits of the IP TOS are evaluated to provide more granular classification; backward-compatible with IP Precedence

QoS Flowchart



Terminology

Per-Hop Behavior (PHB) · The individual QoS action performed at each DiffServ node according to its configured policy

Trust Boundary · The perimeter beyond which QoS markings are not trusted

Tail Drop · Occurs when a packet is dropped because its queue is full

Policing · Creates an artificial ceiling on the amount of bandwidth that may be consumed; traffic exceeding the cap and be remarked or dropped

Shaping · Similar to policing but buffers excess traffic for delayed transmission; makes more efficient use of bandwidth but introduces a delay

TCP Synchronization · Flows adjust window sizes in synch, wasting bandwidth

Per-Hop Behaviors

Class Selector (CS) · Backward-compatible with IP Precedence values

Assured Forwarding (AF) · Four classes with variable drop preferences

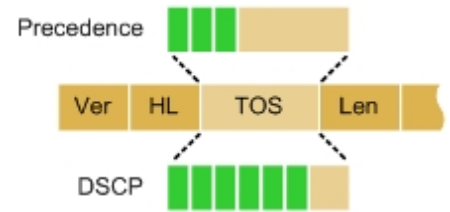
Expedited Forwarding (EF) · Provides priority queuing for delay-sensitive traffic

Congestion Avoidance

Random Early Detection (RED) · Packets are randomly dropped before a queue is full to prevent tail drop; mitigates TCP synchronization

Weighted RED (WRED) · RED with the added capability of recognizing prioritized traffic by its marking

IP Type of Service (TOS)



Precedence Values

Binary	Application
7 111	Reserved
6 110	Routing
5 101	Voice
4 100	Streaming Video
3 011	Call Signaling
2 010	Transactional
1 001	Bulk Data
0 000	Best Effort

DSCP Values

Binary	Prec.	DSCP
56 111000	7	Reserved
48 110000	6	Reserved
46 101110	5	EF
32 100000	4	CS4
34 100010		AF41
36 100100		AF42
38 100110		AF43
24 011000	3	CS3
26 011010		AF31
28 011100		AF32
30 011110		AF33
16 010000	2	CS2
18 010010		AF21
20 010100		AF22
22 010110		AF23
8 001000	1	CS1
10 001010		AF11
12 001100		AF12
14 001110		AF13
0 000000	0	BE

Queuing Comparison Chart

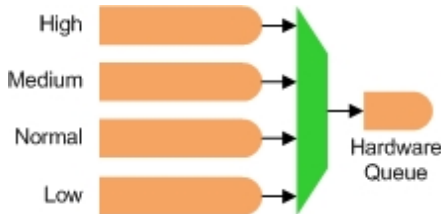
	FIFO	PQ	CQ	WFQ	CBWFQ	LLQ
Default on interfaces	>2 Mbps	No	No	<=2 Mbps	No	No
Number of queues	1	4	Configured	Dynamic	Configured	Configured
Configurable classes	No	Yes	Yes	No	Yes	Yes
Bandwidth allocation	Automatic	Automatic	Configured	Automatic	Configured	Configured
Provides for minimal delay	No	Yes	No	No	No	Yes
Modern implementation	Yes	No	No	No	Yes	Yes

First In First Out (FIFO)



- » Packets are transmitted in the order they are processed
- » No prioritization is provided
- » Default queuing method on high-speed (>2 Mbps) interfaces
- » Configurable with the **tx-ring-limit** interface configuration command

Priority Queuing (PQ)



- » Provides four static queues which cannot be reconfigured
- » Higher-priority queues are always emptied before lower-priority queues
- » Lower-priority queues are at risk of bandwidth starvation

LLQ Configuration Example

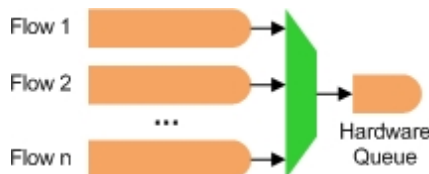
```
! *** Class definitions ***
class-map match-all Voice
! Matches packets by DSCP value
match dscp ef
!
class-map match-all Call-Signaling
match dscp cs3
!
class-map match-any Critical-Apps
match dscp af21 af22
! Matches packets by access list
match access-group name Mgmt_LAN
!
class-map match-all Scavenger
match dscp cs1
!
! *** Policy creation ***
policy-map Foo
class Voice
! Priority queue policed to 33%
priority percent 33
class Call-Signaling
! Allocate 5% of bandwidth
bandwidth percent 5
class Critical-Apps
bandwidth percent 20
! Extend queue size to 96 packets
queue-limit 96
class Scavenger
! Police to 64 kbps
police cir 64000
conform-action transmit
exceed-action drop
class class-default
! Enable WFQ
fair-queue
! Enable WRED
random-detect
!
! *** Policy Application ***
interface Serial0
service-policy Foo
```

Custom Queuing (CQ)



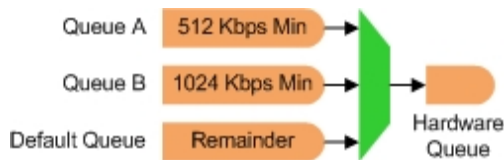
- » Rotates through queues using Weighted Round Robin (WRR)
- » A configurable number of bytes is processed from each queue per turn
- » Prevents queue starvation but does not support delay-sensitive traffic

Weighted Fair Queuing (WFQ)



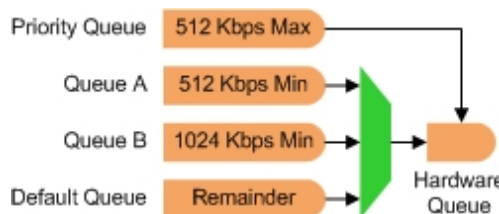
- » Queues are dynamically created per flow to ensure fair processing
- » Statistically drops packets from aggressive flows more often
- » No support for delay-sensitive traffic

Class-Based WFQ (CBWFQ)



- » Provides the benefits of WFQ with administratively configured queues
- » Each queue is allocated an amount or percentage of bandwidth
- » No support for delay-sensitive traffic

Low Latency Queuing (LLQ)



- » CBWFQ with the addition of a policed strict priority queue
- » Highly configurable while still supporting delay-sensitive traffic

Troubleshooting

```
show policy-map
show interface
show queue <interface>
show mls qos
```