



## [Quality of Service \(QoS\) on Netgear switches](#)

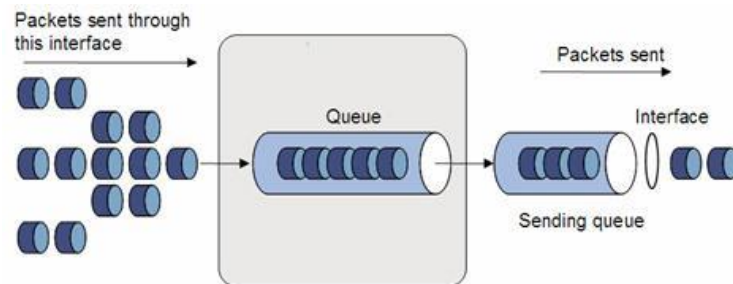
## Section 1 – Principles and Practice of QoS on IP networks

### Introduction to QoS

#### Why?

In a typical modern IT environment, a wide variety of devices are connected together over a common network. Typically, Ethernet is used to connect the devices to packet switches, and those switches forward the packets to their destination as quickly as possible.

By default, all the packets are treated as equal by the switches, and are forwarded in the order that they were received in. If the network is congested by a burst of traffic, the switches will keep packets in their internal memory, in queues called buffers, and forward them when the congestion eases. There will normally be one buffer for each physical port, holding the packets that are to leave the switch (egress) by that port. A buffer is an example of a first-in-first-out (FIFO) queue, so the packets will be forwarded in the order that they were placed in the queue. If the buffer completely fills up before the congestion eases, it will overflow and any further arriving packets will simply be dropped until the congestion eases:



If the network is used for TCP-based applications like email or web-surfing, these applications will continue to work during periods of network congestion, but more slowly. Dropped packets will automatically be re-transmitted, and while a slow response may be annoying for the user it does not make these applications unusable.

However, if the network is also used for making Voice-over-IP (VoIP) phone calls, or Video Conferencing, the impact of network congestion is very different. Video, and particularly Voice, transmissions will quickly become unusable if their packets are delayed or if the delay varies (jitter). And due to the real-time nature of these applications, if their packets are dropped, there is no time to re-transmit them, so the user experiences break-up in their voice or video link.

QoS tries to avoid this problem by prioritizing the forwarding of some packets over others.

QoS is not a network protocol or a single technology, it is a term that encompasses all technologies, protocols or techniques that can be used to prioritize some packets over other packets.

Although any packets can be prioritized, by far the most common use of QoS is to prioritize Voice or Video packets over other packets, as these are the most time-sensitive applications. Voice is even more time-sensitive than video, as video is usually sent heavily compressed, and the decompression process helps to smooth out jitter and even recreate missing packets.

## How?

Getting a switch to prioritize some packets over others involves 3 steps:

1. Classification
2. Marking
3. Queuing

### 1. Classification

Classification is the process of deciding what the priority of a packet should be. This can be as simple as trusting whatever QoS markings are already in the packet header when it is received, or can be based on a complex set of criteria defined by the network administrator. There are no standards for this process.

### 2. Marking

Marking involves setting the value of certain bits in the packet header to indicate what the priority of this packet is. Which bits are set, and what value they are set to, will depend on which layer (2 or 3) markings are being set, and which QoS standard(s) are being used. There are a number of different QoS standards in use, some of which use the same packet header bits as each other to mark priority, but set these bits to different values. This has the potential to cause much confusion when configuring QoS on a network.

### 3. Queuing

Queuing is the process that tries to ensure that packets marked as high priority are forwarded before those marked as lower priority. If network congestion becomes so severe that packet drops are inevitable, the queuing process will also choose which packets are to be dropped. There are many ways to try to achieve both of these objectives, each with advantages and disadvantages.

Note that each switch in a network takes QoS decisions and actions independently, depending on how it has been configured by the network administrator. Each switch can choose to trust the QoS markings that have been applied to packets by another switch, or ignore them, or change them. It can also choose what order to forward marked packets in – although by default packets marked high-priority would be forwarded before those marked low-priority, this can be changed. The QoS standards described below are really only QoS Packet Marking standards – they do not specify the classification and queuing choices that a switch makes.

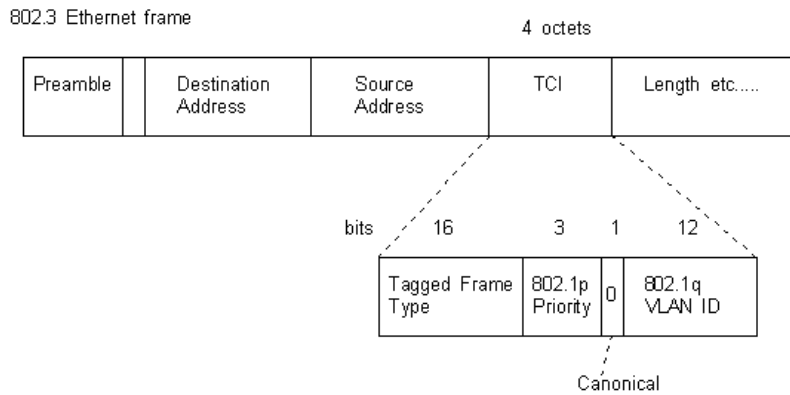
## Layer 2 QoS standards

### 802.1p (dot1p)

There is currently only one QoS standard in widespread use on LANs for priority marking of frames at Layer2.

This is the IEEE 802.1p standard, which specifies 8 priority levels, called Class of Service (CoS) levels. These are numbered 0-7, with 0 being the lowest priority and 7 the highest. CoS level 0 is the default for most frames.

A frames CoS level is marked using 3 bits in the Tag Control Information (TCI) field in the 802.3 Layer 2 frame header:



The 8 CoS priority levels are named as follows. Note that for simplicity these are named the same as the 8 priority levels provided at Layer3 by the IP Precedence standard discussed in the next section:

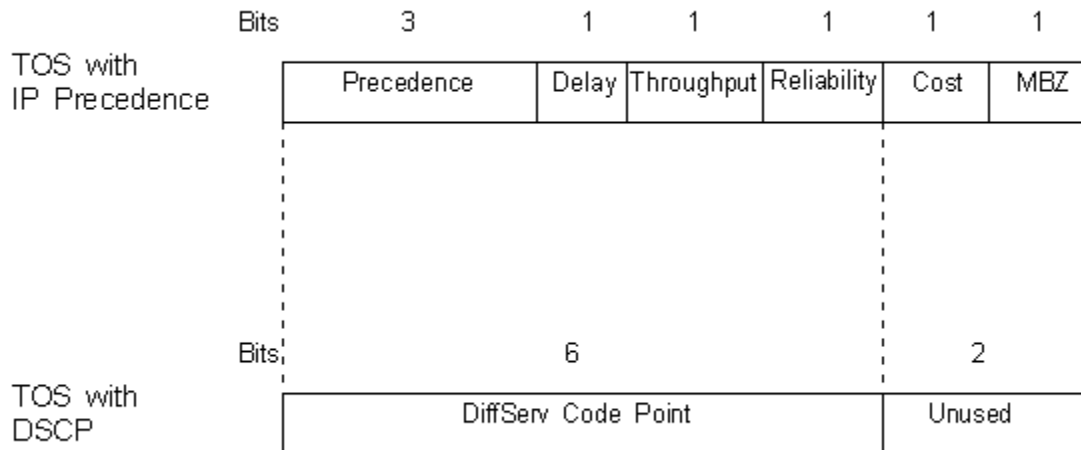
Decimal value	Binary (802.1p bits)	CoS name
0	000	Routine
1	001	Priority
2	010	Immediate
3	011	Flash
4	100	Flash Override
5	101	Critical
6	110	Internetwork Control
7	111	Network Control

## Layer 3 QoS standards

There are currently 3 QoS standards in widespread use on LANs for priority marking of packets at Layer 3. These are:

1. Type of Service (ToS) (now in declining use)
2. IP Precedence
3. Differentiated Services (DiffServ)

All of these standards use the 2<sup>nd</sup> byte in the IP packet header (called the ToS byte) to mark a packet's priority, which can lead to confusion when setting up QoS on a network. When working in an environment where different QoS standards are in use, a good understanding of the use of the ToS byte is required:



### 1. Type of Service (ToS)

The ToS standard, as defined by RFC1349, can be used to mark packets with 5 levels of priority:

Normal-Service

Minimize-Cost

Maximize-Reliability

Maximize-Throughput

Minimize-Delay

Earlier versions of the ToS RFC only defined 4 types of service, they did not include “Minimize-Cost”. The “Minimize-Cost” ToS attempts to minimize the financial cost of passing traffic, not to be confused with normal IP routing protocols which try to route packets based on minimum “Hops” or “Metrics”, often referred to as “Costs”.

The required ToS is marked in a packet header by setting or un-setting bits 4-7 of the ToS byte (note that bit 8, the MBZ (must be zero) bit, is always zero):

Bits, ToS byte:	1	2	3	4	5	6	7	8
	IP Precedence	IP Precedence	IP Precedence	Minimize Delay	Maximize Throughput	Maximize Reliability	Minimize Cost	Unused MBZ
Normal Service	0	0	0	0	0	0	0	0
Minimize Delay	0	0	0	1	0	0	0	0
Maximize Throughput	0	0	0	0	1	0	0	0
Maximize Reliability	0	0	0	0	0	1	0	0
Minimize Cost	0	0	0	0	0	0	1	0

## 2. IP Precedence

The IP Precedence standard uses the first 3 bits of the ToS byte to mark packets with 8 levels of priority, numbered 0-7, with 0 being the lowest priority and 7 the highest.

These levels are given the same names as the 8 priority levels provided at Layer2 by the 802.1p standard :

Decimal value	Binary (bits 1-3 ToS byte)	IP Precedence name
0	000	Routine
1	001	Priority
2	010	Immediate
3	011	Flash
4	100	Flash Override
5	101	Critical
6	110	Internetwork Control
7	111	Network Control

Bits, ToS byte:	1	2	3	4	5	6	7	8
	IP Precedence	IP Precedence	IP Precedence	Minimize Delay	Maximize Throughput	Maximize Reliability	Minimize Cost	Unused MBZ
Routine	0	0	0	0	0	0	0	0
Priority	0	0	1	0	0	0	0	0
Immediate	0	1	0	0	0	0	0	0
Flash	0	1	1	0	0	0	0	0
Flash Override	1	0	0	0	0	0	0	0
Critical	1	0	1	0	0	0	0	0
Internetwork Control	1	1	0	0	0	0	0	0
Network Control	1	1	1	0	0	0	0	0

Because IP Precedence and ToS use different bits in the ToS byte to mark the priority of a packet, they can co-exist in the same packet header without interfering with each other.

## 3. Differentiated Services (DiffServ)

The DiffServ standard uses the first 3 bits of the ToS byte to mark packets with 6 levels of priority, called Class Selectors (CS). These are called “default”, CS1, CS2, CS3, CS4 and “Expedited Forwarding(EF)”, and to maintain backward compatibility they correspond exactly to the first 6 priority levels specified by the IP Precedence standard:

bit 1	bit 2	bit 3	Decimal value	DiffServ Class Selector	IP Precedence name
0	0	0	0	default	Routine
0	0	1	1	CS1	Priority
0	1	0	2	CS2	Immediate
0	1	1	3	CS3	Flash
1	0	0	4	CS4	Flash Override
1	0	1	5	EF	Critical
1	1	0	6		Internetwork Control
1	1	1	7		Network Control

The DiffServ standard also uses the next 3 bits (bits 4,5 and 6) of the ToS byte to mark packets with a Drop Precedence.

The standard specifies that bit 6 always has a value of zero and is ignored.

This leaves 2 bits (bits 4 and 5), giving 4 possible values of Drop Precedence:

bit 4	bit 5	Decimal value	DiffServ Drop Precedence
0	0	0	default
0	1	1	Low Drop Probability
1	0	2	Medium Drop Probability
1	1	3	High Drop Probability

The first 6 bits of the ToS byte, used to mark packets with a DiffServ Class Selector and Drop Precedence, are called the Differentiated Services Code Point (DSCP).

The final 2 bits of the ToS byte may be unused, or they may be used for Explicit Congestion Notification (ECN):

Bits:1	2	3	4	5	6	7	8
Priority	Priority	Priority	Drop Precedence	Drop Precedence	Drop Precedence	Unused	Unused
Class Selector			Drop Precedence		Ignore	Unused or ECN	
Differentiated Services			Code Point (DSCP)				

There is a naming standard for various combinations of Class Selector and Drop Precedence. This is:

“Assured Forwarding (AF)” <decimal value of CS bits><decimal value of Drop Precedence bits(ignoring bit6)>

For example, “AF31” means CS3 and Low Drop Probability, ie. DSCP = 011010 = 26

Per Hop Behaviour	Class Selector	DiffServ Code Point (DSCP)			IP Precedence
Default	default	000000 (0)			0
Assured Forwarding		<b>Low Drop Probability</b>	<b>Medium Drop Probability</b>	<b>High Drop Probability</b>	
	CS1	AF11	AF12	AF13	1
		001010 (10)	001100 (12)	001110 (14)	
	CS2	AF21	AF22	AF23	2
		010010 (18)	010100 (20)	010110 (24)	
	CS3	AF31	AF32	AF33	3

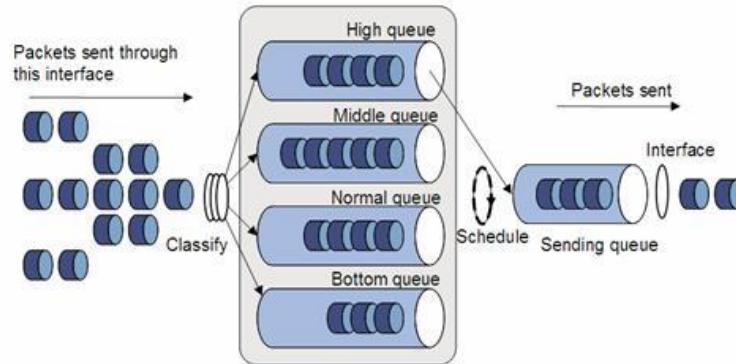
		011010 (26)	011100 (28)	011110 (30)	
	CS4	AF41	AF42	AF43	4
		100010 (34)	100100 (36)	100110 (38)	
<b>Expedited Forwarding</b>	EF	101110 (46)			5

Note that although the DSCP is 6 bits long and could therefore hold 64 values, this table shows that only 14 of those possible values are valid values.

## Queuing

Instead of a single buffer queuing outgoing packets at each physical egress port on a switch, imagine 2 queues at each egress port. The switch places packets that have been marked as high priority into one queue, and other packets into the other queue. The egress port always transmits packets from the high priority queue first, and only when that queue is empty does it start transmitting packets from the other queue. This will clearly achieve the objective of some packets being passed through the network faster and more reliably than others, particularly during network congestion.

This is an example of a very simple queuing and queue servicing schedule, and some switches do implement QoS exactly like this. However, more expensive switches will have more than 2 queues per port, to allow for more than 2 levels of priority:



Ideally, there would be a separate physical queue for each possible QoS marking that a packet might have. However, this would make the switch very expensive, so instead there is a table in the switch settings that maps QoS markings to physical queues. For example, cheap switches with only 2 queues per port usually map 802.1p or IP Precedence markings 0-4 to the low priority queue, and markings 5-7 to the high priority queue. Some switches allow this mapping to be changed by the administrator, and some do not.

### Queuing Servicing Schedules:

The queue servicing schedule described in the example above can lead to high priority traffic monopolizing the entire network bandwidth leaving nothing for other traffic. There are many other more sophisticated queue servicing schemes that attempt to avoid this and other problems in a variety of ways.



Bear in mind that queue servicing is completely local to an individual switch, so different manufacturers may use different schemes, implemented in different ways, and named in different ways. This can lead to some confusion for administrators, but does not impede communication between different manufacturer's products.

Only two queue servicing schedules are available on Netgear switches:

1. **“Strict”** : This is the simple schedule described in the example above. The highest-priority queue will be emptied first, then the next-highest priority etc. If there are a lot of high priority packets, the lowest-priority queue may never get serviced.
2. **“Weighted”** : This is the default setting on Netgear switches. All the queues get serviced on a round-robin basis, but more packets will be sent from the high priority queues during each servicing turn than from the lower-priority queues. High priority packets are still more likely to be sent quickly, and less likely to get dropped, than lower priority packets, but even the lowest-priority queue will always get serviced.

### **Congestion Management:**

If a network becomes congested by a lot of traffic, packets may be placed in the sending queues faster than they can be sent out the physical interface. If the congestion only lasts for a short time, the queues will act as buffers, storing the packets until the congestion eases and they can be sent.

However, if the congestion lasts too long, the queues will fill up. Any more packets that arrive will have to be dropped (deleted), since there is nowhere to store them. This is called “Taildrop”, and is the only congestion management method available on Netgear switches.

Other, more sophisticated methods, such as Weighted Random Early Detection (“WRED”) start dropping some packets before the queues are full, to avoid a sudden transition from forwarding-all-packets to dropping-all-packets.

### **Rate Limiting / Traffic Shaping:**

In some circumstances it is necessary to limit the rate at which packets are sent out a physical interface to less than the actual speed of the interface. This usually arises because of contractual arrangements between network service providers and customers. It is not really part of QoS, but because it is achieved by using the same queuing / queue servicing system as QoS, it is often configured at the same time.

### **Policing:**

In some circumstances, it is necessary to limit the rate at which certain packets are received by an interface. This usually arises because of contractual arrangements between network service providers and customers. Typically, the service provider will use Policing to limit how quickly the customer can send data, and use Shaping to limit how quickly the customer can receive data. The customer may also use Shaping to ensure that they comply with the contract and do not incur a financial penalty or suffer a lot of dropped packets by sending data too quickly to the service provider.

Policing is also not really part of QoS, but compliance (or otherwise) with a policing policy is often used as part of the criteria for what QoS markings are applied to a packet, ie. rather than simply dropping packets that arrive too quickly, the service provider might mark them with a low priority and a high drop probability before forwarding them.

## Section 2 – Configuring QoS on Netgear switches

The QoS features available on Netgear switches vary dramatically from model to model. The following examples will refer to a GSM7328S, a high-end layer 3 switch, running firmware 7.3.3.1, as this has rich QoS features.

On this switch, the QoS menu is divided into 2 submenus, “CoS” and “DiffServ”. Unfortunately, these terms do not directly correspond their use in the Section 1:

The CoS menu items are about:

1. Trusting (or not trusting) the existing QoS markings of inbound packets.
2. Queuing outbound packets.

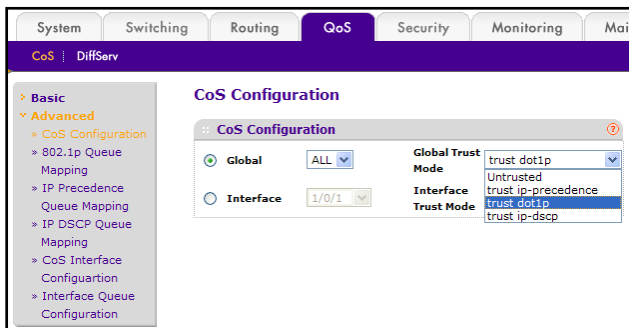
The DiffServ menu items are about:

1. Classification of inbound packets, based on a very wide range of criteria.
2. Marking inbound packets with the desired QoS markings.

### The CoS menu

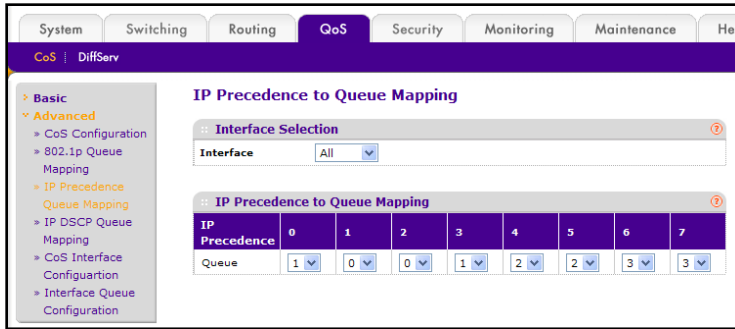
#### 1. Trusting

The “CoS Configuration” menu controls the trusting (or not trusting) or the existing QoS markings on inbound packets. The default is to trust “dot1p” (802.1p layer 2 priority) markings on inbound packets on all physical interfaces. The other options are to trust IP Precedence markings, trust DSCP (DiffServ) markings, or not to trust any markings. This trust setting can be set globally (all ports) or separately on each interface:



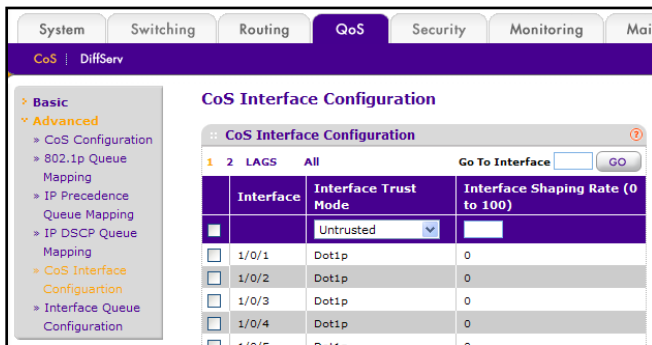
#### 2. Mapping

The 3 “Queue Mapping” menus control the mapping of packet priority markings to physical outbound queues. There is usually no need to change these settings from the default:



### 3. Shaping

The “CoS Interface Configuration” menu provides another menu where the trust settings of each interface can be viewed and changed. It also provides the ability to limit, or “shape”, the maximum data rate at which data will be sent out each interface (value is % of physical interface speed):

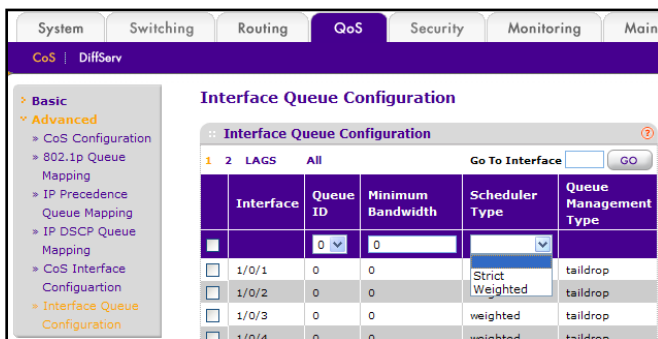


### 4. Queue Servicing

The “Interface Queue Configuration” menu allows the queue servicing schedule to be adjusted individually for each queue on each interface. The scheduling types available are “strict” and “weighted”, as described in Section 1. Default is “Weighted”.

The scheduling can be further refined by assigning a minimum % of the available bandwidth to the queues on an interface. This could be used to ensure that there is always enough bandwidth for a video link, for example, or to ensure that low-priority traffic always has a minimum amount of bandwidth available.

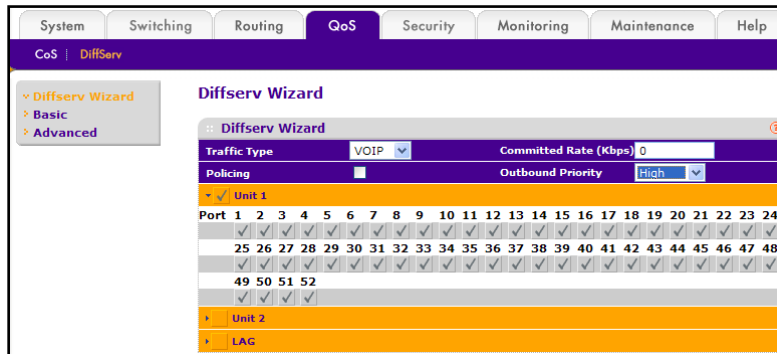
There is usually no need to change these settings from the default.



## The DiffServ Menu

### 1. The DiffServ Wizard

The settings in the diffserv menus can appear complex. The “DiffServ Wizard” automatically creates a simple setup, which can be edited later if required:



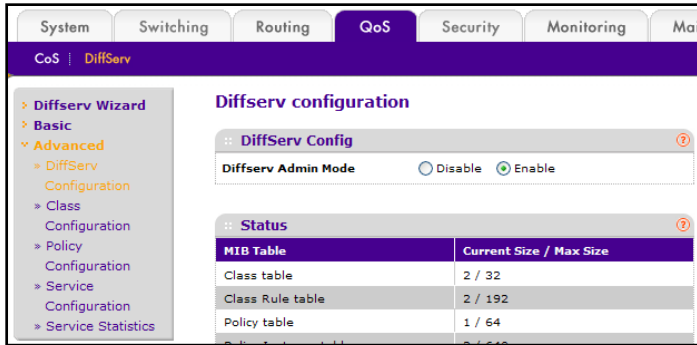
The following is an extract from the help screen on the switch for the DiffServ Wizard, which illustrates the steps required to set up DiffServ manually:

The DiffServ Wizard enables DiffServ on the switch by creating a traffic class, adding the traffic class to a policy, and then adding the policy to the ports selected on DiffServ Wizard page. The DiffServ Wizard will:

- Create a DiffServ Class and define match criteria used as a filter to determine if incoming traffic meets the requirements to be a member of the class.
  - Set the DiffServ Class match criteria based on Traffic Type selection as below:
    - VOIP - sets match criteria to UDP protocol.
    - HTTP - sets match criteria to HTTP destination port.
    - FTP - sets match criteria to FTP destination port.
    - Telnet - sets match criteria to Telnet destination port.
    - Every - sets match criteria all traffic.
  - Create a DiffServ Policy and adds the DiffServ Policy to the DiffServ Class created.
  - If Policing is set to YES, then DiffServ Policy style is set to Simple. Traffic which conforms to the Class Match criteria will be processed according to the Outbound Priority selection. Outbound Priority configures the handling of conforming traffic as below:
    - High - sets policing action to markdscp ef.
    - Med - sets policing action to markdscp af31.
    - Low - sets policing action to send.
- If Policing is set to NO, then all traffic will be marked as specified below:
- High - sets policy mark ipdscp ef.
  - Med - sets policy mark ipdscp af31.
  - Low - sets policy mark ipdscp be.
- Each port selected will be added to the policy created.

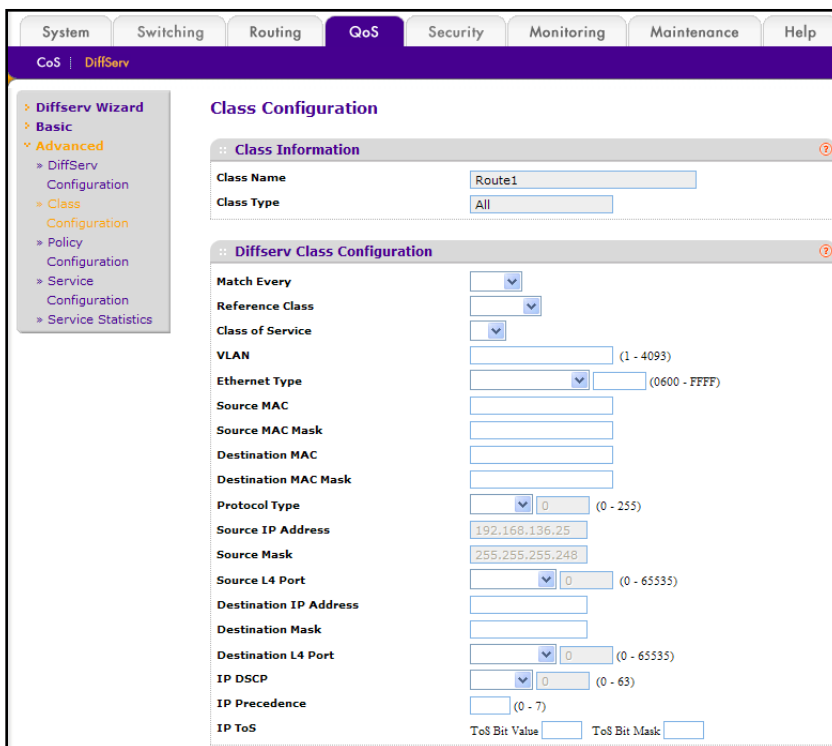
### 2. Enable/Disable

The “DiffServ Configuration” menu simply enables or disables DiffServ on the switch. DiffServ is enabled by default. If DiffServ is disabled, the current settings are not lost:



### 3. Classification

The first step of setting up DiffServ manually. The “Class Configuration” menu allows a class of packets to be defined, by specifying detailed criteria which the switch uses to determine which packets are members of the class. Create a new class name, edit that class, and the following screen is presented:



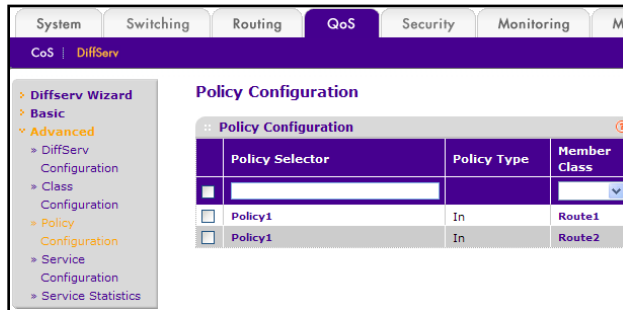
This looks complex, but it is basically just a list of information that can be found in the layer 2 and layer 3 headers of a packet. If the information in a packet's headers match the information entered on this screen, the packet is a member of this class.

The “Reference Class” option can be used to specify another class configured on this switch. Any packets that are members of the reference class will also be a member of this class.

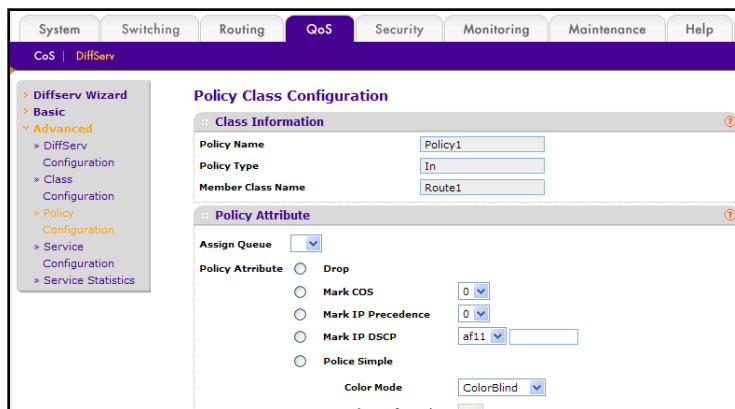
#### 4. Actions: Dropping, Marking and Policing

The second step of setting up DiffServ manually. This step involves specifying what action to take for each packet, based on the packet's membership of classes specified in the previous step.

The "Policy Configuration" menu item allows new action policies to be created, and one or more classes to be associated with each policy:



Edit a policy, and the following screen is presented:

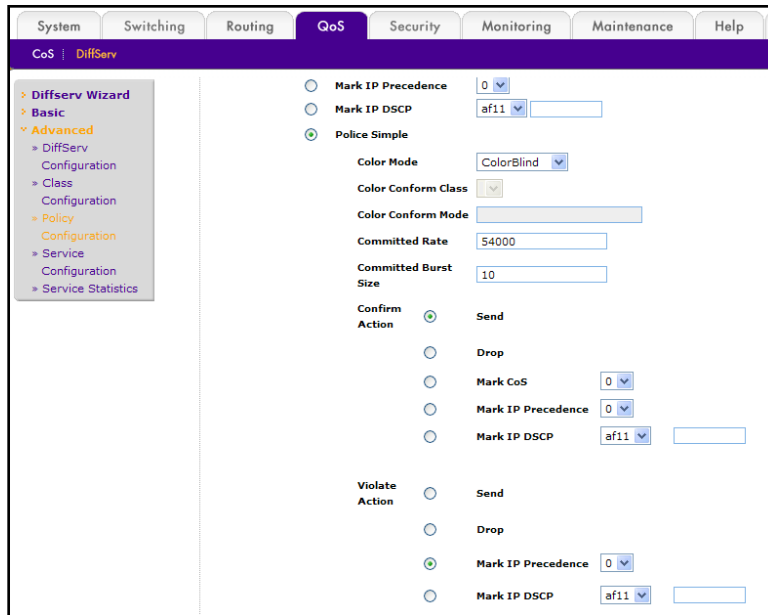


Any one of the following 6 actions can be chosen:

- (i) The "Assign Queue" setting directly specifies what output queue the packet will be placed in. This bypasses the "marking" stage of QoS.
- (ii) The "Drop" setting specifies that the packet will be dropped. It will simply disappear.
- (iii) The "Mark COS" setting specifies what QoS priority marking will be written into the Layer 2 header of the packet. Any existing marking will be overwritten.
- (iv) The "Mark IP Precedence" setting specifies that an IP Precedence QoS priority marking will be written into the Layer 3 of the packet. Any existing marking will be overwritten.
- (v) The "Mark IP DSCP" setting specifies that a DSCP QoS priority marking will be written into the Layer 3 of the packet. Any existing marking will be overwritten.
- (vi) The "Police Simple" options are explained below.

## Policing

The purpose of the “Police Simple” settings is to compare the rate at which the classified packets are received by an interface to a specified “committed rate” and “committed burst size”, and to decide if packets are arriving slower than the specified rate (Conforming Packets), or are arriving faster than the specified rate (Non-conforming or Violating packets).



Note: in the screenshot, “Confirm Action” should read “Conform Action”.

In this example, if packets that are members of the “Route1” class arrive at less than 54000 Kbps (kilo bits per second) they will be deemed “Conforming” and will be sent normally.

If the packets arrive faster than 54000 Kbps, but only do so in bursts of 10 KB (kilo bytes) or less, they will also be deemed “Conforming” and sent.

If the packets keep arriving faster than 54000 Kbps, they will be deemed “Violating” and, in this example, they will be marked with an IP Precedence of Zero and sent. In many real-world scenarios, they would simply be dropped.

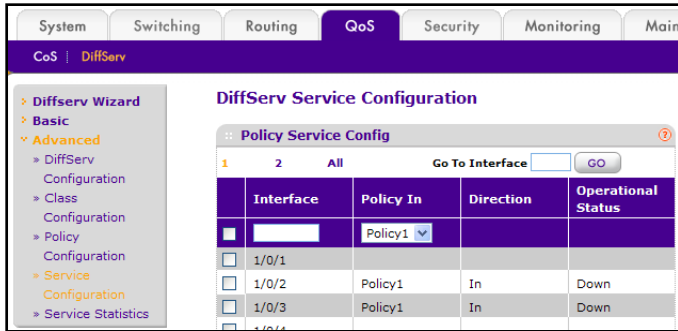
There is an alternative method of Policing using colours which is not discussed here, see RFC 2697.

It is important to understand how the rate of incoming packets is measured for policing purposes. The specified Committed Rate is used to calculate a minimum inter-packet time. If the time between the arrival of 2 packets is smaller than this time (and the Committed Burst Size has been exceeded), the second packet is deemed to be Violating. This leads to the surprising result that even if only 2 small packets were sent, if they were too close together the second one could be deemed Violating.

## **5. Assigning a Policy to a physical interface**

The final step in setting up DiffServ manually. The “Service Configuration” menu item allows policies to be assigned to one or more physical interfaces. Only the packets that enter the switch

though the specified ports will be checked by the policy. Policies can also be assigned to Link Aggregation Groups (LAGS), but not to virtual interfaces used for VLAN routing:



One could ask, why not just apply all policies to all interfaces? The primary answer to this is that checking packets consumes switch resources, so to avoid slow switch throughput only packets that need to be checked should be checked.