QoS in Axis Video Products

Table of contents

1 Quality of Service

1.1 What is QoS?

Quality of Service (QoS) provides the means to guarantee a certain level of a specified resource to selected traffic on a network. Quality can be defined as e.g. a maintained level of bandwidth, low latency, no packet losses, etc. The main benefits of a QoS aware network can be summarized as:

- The ability to prioritize traffic and thus allow critical flows to be served before flows with lesser priority.
- Greater reliability in the network, thanks to the control of the amount of bandwidth an application may use, and thus control over bandwidth races between applications.

1.2 Requirements for QoS

To use QoS in a network with Axis video products the following requirements need to be met:

- All network switches and routers must include support for QoS. This is important to achieve end-to-end QoS functionality.
- The Axis video products used must be QoS enabled.

1.3 A QoS network scenario

Imagine that the network in *Figure 1* is an ordinary (non-QoS aware) network. In this example, *PC1* is watching two surveillance video streams from cameras *Cam1* and *Cam2*, with each camera streaming at 2.5 Mbps.

Suddenly *PC2* starts a file transfer from *PC3*. In this scenario the file transfer will try to use the full 10 Mbps capacity between the routers *R1* and *R 2,* whilst the video streams will try to maintain their total of 5 Mbps. We can no longer guarantee the amount of bandwidth given to the surveillance system and the video frame rate will probably be reduced. At worst, the FTP traffic will consume all the available bandwidth.
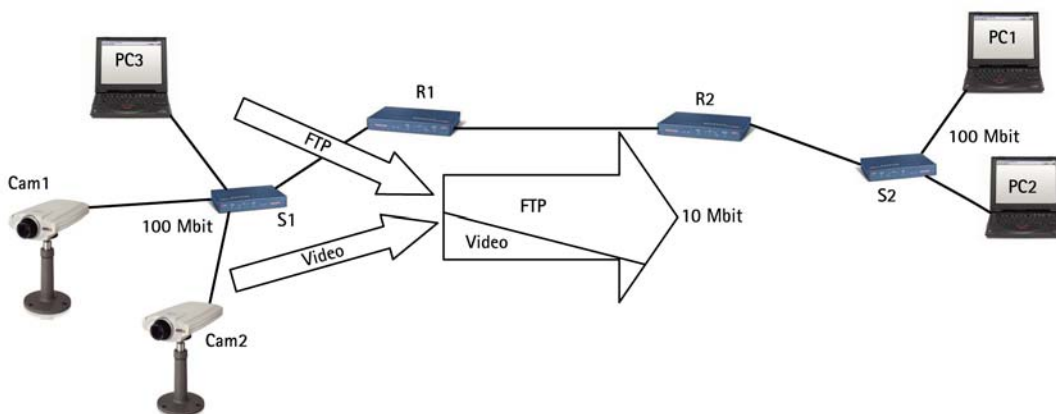

Figure 1

Now, suppose the network in *Figure 2* is QoS-aware and uses the *DiffServ* model (see below). The router *R1* has been configured to devote up to 5 Mbps of the available 10 Mbps for streaming video. FTP traffic is allowed to use 2 Mbps, and HTTP and all other traffic can use a maximum of 3 Mbps.

Using this division, video streams will always have the necessary bandwidth available. File transfers are considered less important and get less bandwidth, but there will still be bandwidth available for web browsing and other traffic. Note that these maximums only apply when there is congestion on the network. If there is unused bandwidth available, this can be used by any type of traffic.

By using QoS we allow the network applications to co-exist on the same network, without consuming each other's bandwidth.
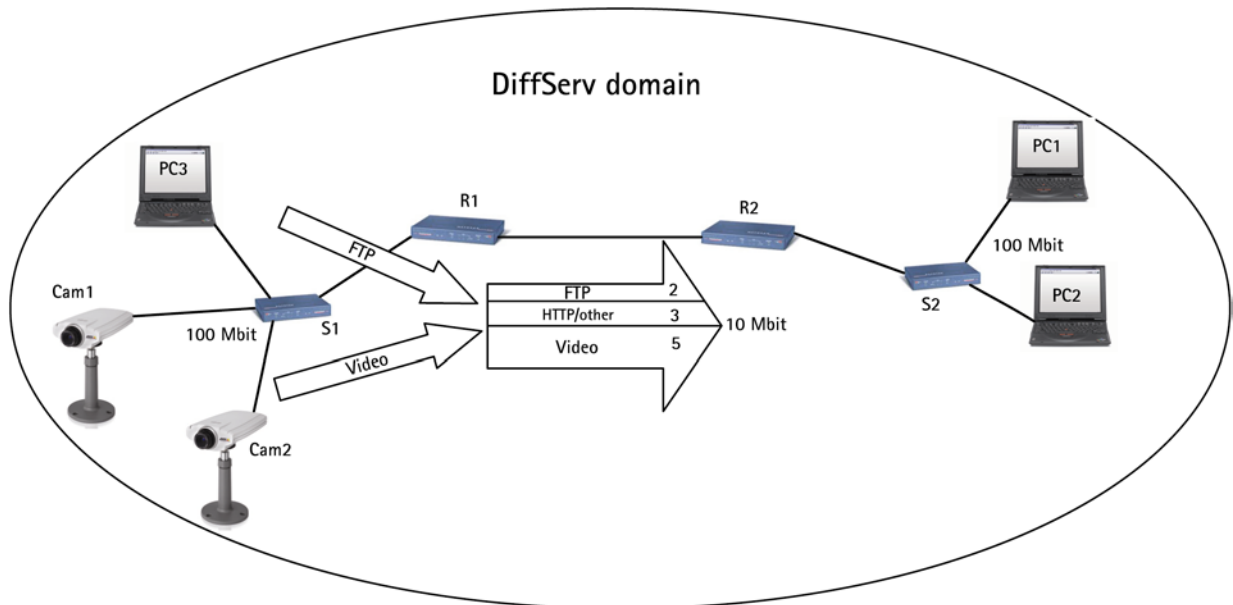

Figure 2

2 QoS models

There are several different ways (models) of implementing QoS. Axis products use the **DiffServ** model, for greater scalability and flexibility. See below for more information.

2.1 The IntServ model

The IntServ model, or *Integrated Services model* uses a protocol called RSVP (Reservation Setup Protocol), which is used to reserve a certain traffic quality in the network.

Prior to use, each application in an IntServ QoS network reserves its own resources and the router either grants or denies the request. When a reservation request is received, the routers need to find a path that can support the IntServ request and also the route that offers the best services.

The major problem with this model is scalability. As the network increases in size, the connections database will grow to enormous proportions and it will be both difficult and ineffective to keep track of all reservations.

Another drawback is the model's inflexibility – the maximum amount of the resource in question that each type of traffic will ever receive is the maximum specified in the model, even if there are other unused resources available.

2.2 The DiffServ model

The *Differentiated Services (DiffServ) model* was introduced in 1998. It is based on two major components:

- Packet marking
- Router queuing disciplines

The applications in a DiffServ network mark their traffic so the router knows which service to apply to the packet. The marking is done in the IP header, by setting a field called the *DSCP* (Differentiated Services Codepoint). This is a 6-bit field that provides 64 different class IDs.

Each DSCP value represents a QoS class, also known as a *behavior aggregate*. Thus, different applications can mark their own traffic with the same DSCP value.

The intelligence of a DiffServ network is setup in the routers, where a particular DSCP value is mapped to a particular routing behavior. This behavior is referred to as a *Per-Hop Behavior* and is implemented in the router using different *queuing disciplines*.

The DiffServ model abandons the concept of states as used in IntServ, and routers operate in a connectionless mode. This adds scalability to the system, since each router works independently, unaware of the network size and complexity. This model is also more flexible, as the classes of service are not strictly defined. Resources over and above the value specified as the maximum when resources are limited can be freely exploited whenever available.

The main benefit of using this model in Axis products is that the products mark their own traffic, instead of letting the boundary nodes of the DiffServ domain do so. The conditioning algorithms in the boundary node cannot, for example, distinguish video over HTTP from audio over HTTP - they will only see HTTP, and probably provide a much lower level of service than actually required.

To allow a node outside the DS domain to classify its own traffic:

- The boundary router must be correctly configured
- The camera must be set up as a *trusted* node

2.2.1 Per-Hop Behaviors

PHBs define the router's forwarding treatment for a certain class of traffic. Classes of traffic are also known as *Behavior Aggregates* (BA). Members of the BA are marked with the same DSCP value. There are four *recommended* PHBs defined, each mapped to one or more DSCP value.

2.2.1.1 The Default PHB
The default PHB provides the traditional *best-effort* service. It is represented by DSCP value 000000 (*NOTE: 6-bits set to 0*) and it must be available at any DiffServ router. This is the forwarding treatment applied to all non-DiffServ aware applications.

2.2.1.2 The Class-Selector PHB
To preserve backward compatibility with the old TOS field definition, the class-selector PHB defines a DSCP value of the form xxx000. The three bits represent the *IP Precedence* defined by a Type of Service aware network. The Class-Selector PHB ensures that DiffServ compliant nodes can coexist with IP Precedence-based nodes.

2.2.1.3 Expedited Forwarding
The Expedited Forwarding (EF) PHB provides a low loss, low latency, low jitter and guaranteed bandwidth service and can be considered the top priority behavior. Applications such as VoIP require this kind of robust service. The recommended DSCP value for EF is 101110.

2.2.1.4 Assured Forwarding
With Assured Forwarding (AF) traffic can be divided into four different classes. Each class is usually

assigned a specific amount of bandwidth.

Within each class it is possible to specify three different drop precedence values. This value denotes the order in which packets will be dropped when there is congestion. A packet with a higher drop precedence will be dropped first.

This gives us an AF matrix, and hence AF is often denoted AF*xy*, where *x* is the class number and *y* is the drop precedence. *Table1.1* shows the DSCP values recommended for the AF PHB.

Drop Precedence	Class 1	Class 2	Class 3	Class 4
Low dP (1)	001010	010010	011010	100010
Medium dP (2)	001100	010100	011100	100100
High dP (3)	001110	010110	011110	100110

Table 1.1

2.2 DiffServ domain

A *DiffServ domain* (DS-domain) is a network that has been set up according to a particular DiffServ specification, and which has a set of DSCP values mapped to certain PHBs.

As each router in a DiffServ domain forwards traffic based on DSCP value mapping, we can only guarantee that we get the correct forwarding treatment within our own DS-domain. As soon as a packet leaves the domain we can no longer guarantee how other routers will map our DSCP values.

The same goes for incoming traffic - we must reclassify incoming traffic to match our rules. This means we must mark the traffic with a DSCP value valid in our domain. This is usually done by looking at the packet header and setting a new DSCP value based on the source address, port numbers, etc.

The marking process is called *traffic conditioning* and is done at the boundary nodes of the DS domain.

2.3 The VLAN 802.1P model

IEEE802.1P defines a QoS model at OSI Layer 2 (L2, Data Link). This is called *CoS*, Class of Service, and adds an extra 3-bit field (called *user-priority*) to the VLAN MAC header. This splits traffic into 8 different classes.

Priority is set up in the switches, which then use different queuing disciplines to forward the packets correctly.

Although the 802.1P CoS works well, it lacks scalability and cannot provide end-to-end guarantees. We cannot assume the L2 protocol to be constant on a larger network or on the Internet. Therefore, most of today's high-end switches implement mapping from L3 (IP DSCP) QoS to L2 CoS.

3 Routers

3.1 The DiffServ implementation in routers

Routers handle the forwarding of network traffic from one network to another. All packets entering from one network (see *Figure 3*) are queued before being processed, to determine which network they should be forwarded to.
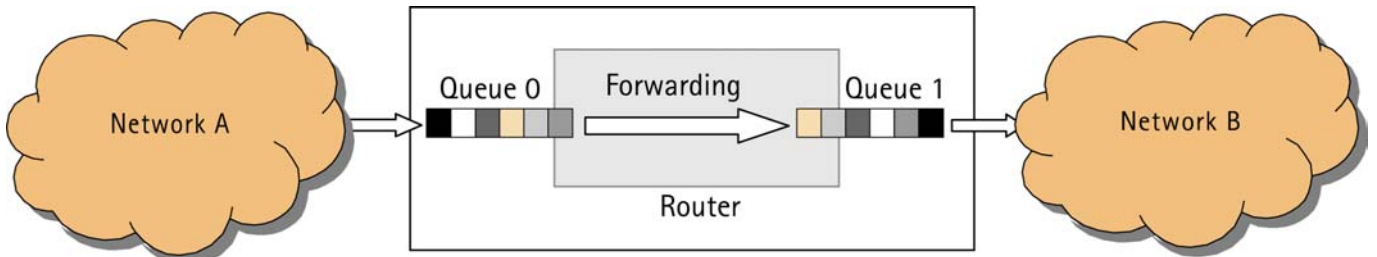

Figure 3

The DiffServ implementation in routers is a way of providing forwarding treatment based on the DSCP of an incoming packet. This is done by using queuing disciplines and prioritizing of packets.

Note: Router configuration is brand dependent and is not covered by this document.

3.2 Queue disciplines

To be able to provide the mechanism desired by DiffServ the routers implement different ways of handling its queues. These different techniques are known as *queuing disciplines* and are algorithms designed to provide different service guarantees.
To exemplify the use of different queuing disciplines we will briefly explain FIFO queues and priority queues.

3.2.1 FIFO

This is the simplest queuing discipline and is usually the default setting in network routers. FIFO describes the simple First-In-First-Out behavior of the queue.
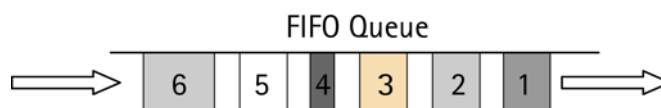

Figure 4

Figure 4 shows a FIFO queue where packets enter from the left and exit to the right. The first packet entering will be the first packet sent out on the port. Other packets will be placed at the end of the queue and must wait their turn. This discipline is too basic to provide any QoS service.

3.2.2 Priority queuing

Priority queuing is a relatively simple way of implementing differentiated service classes in a router. Instead of using one FIFO queue only, several are used, and each is assigned a different priority. A *classifier* determines which queue to put the incoming packet in, and the *scheduler* puts the packets out onto the network.
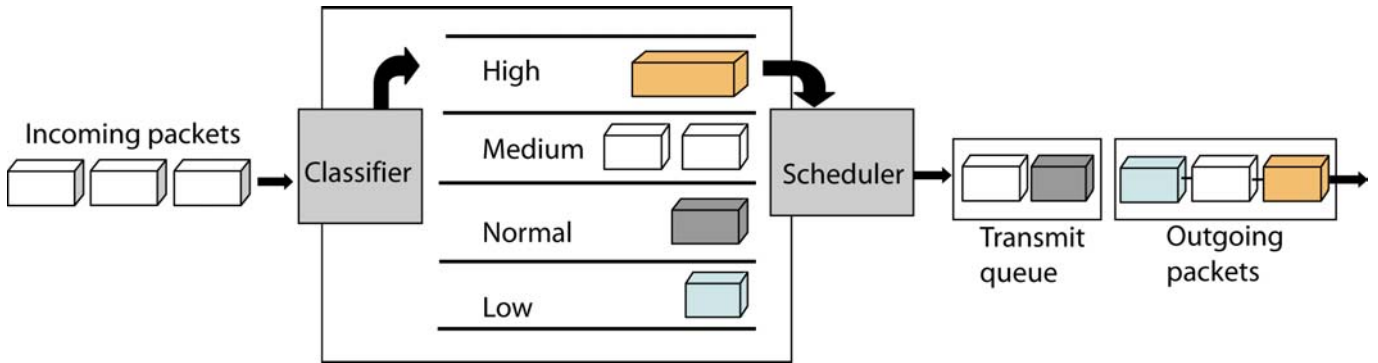

Figure 5

The classifier parses the header of incoming packets and decides which priority queue to put the packet in. In a DiffServ network this decision is based on the DSCP value of the incoming packet.

The scheduler ensures that the high priority queue gets served first, the medium priority queue next, and so on. This discipline allows the traffic to be prioritized.

4 QoS in Axis products

4.1 Overview

Axis has opted to use the DiffServ model, to provide greater scalability and flexibility. This is implemented by:

- Marking network traffic
- Classifying network applications in QoS classes
- Providing a user interface

4.2 The Axis QoS classes

Axis uses four QoS classes. The members of a class all mark their traffic with the same DSCP value and thus receive the same forwarding treatment from routers.

4.2.1 Live Video

This class consists of applications that stream Motion JPEG video streams over HTTP, and MPEG-4 video streams over RTP, RTP/RTSP and RTP/RTSP/HTTP.

4.2.2 Live Audio

This class consists of applications that generate audio flows, and is only present in products that support audio.

4.2.3 Event/Alarm

This class consists of applications that generate event and alarm traffic. The class handles FTP, HTTP, SMTP and TCP events.

4.2.4 Management

This class consists of applications that generate management traffic. The class can handle FTP, HTTP, HTTPS and SNMP management traffic.
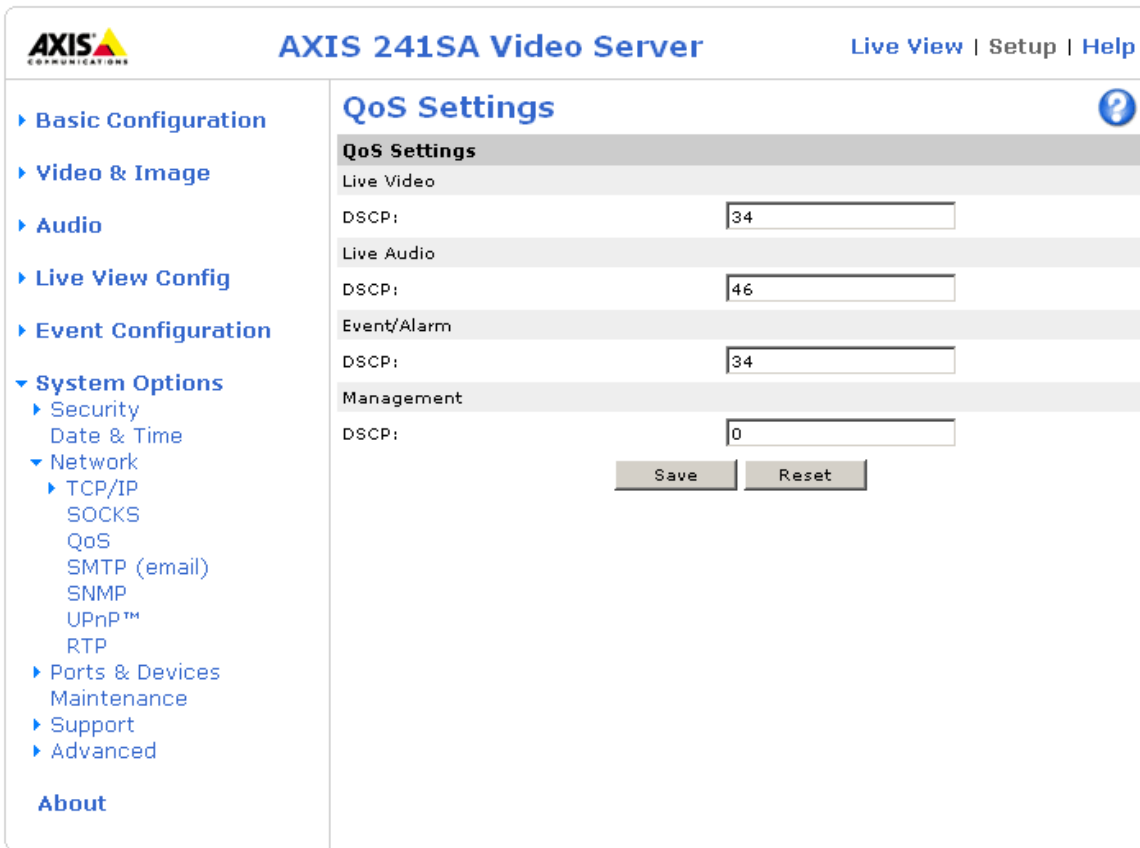
4.3 The HTTP user interface

Figure 6

Figure 6 shows an example of a screenshot of the HTTP user interface for QoS. It shows the different QoS classes supported by the product and lets the user specify the DSCP (DiffServ Codepoint) value for each class.

The default DSCP value is 0, which is the same as *the default PHB* and which can be interpreted as QoS disabled.

5 Performance improvements

5.1 Maintaining video frame rates with QoS

The graph below illustrates a possible scenario in which performance is maintained/improved on a QoS enabled network.

a) Video is first streamed over a quiet, non-QoS network, where the frame rate can be assumed to be approximately 30 frames per second. This is illustrated by line A.

b) Network congestion is then simulated by e.g. adding a non-responsive UDP stream. The frame rate drops sharply, to a practically unusable level, as shown by line B.

c) Video is then streamed over a DiffServ-enabled network, with the same congestion also added here. The video traffic is unaffected by the congestion and the frame rate is maintained, as shown by line C.
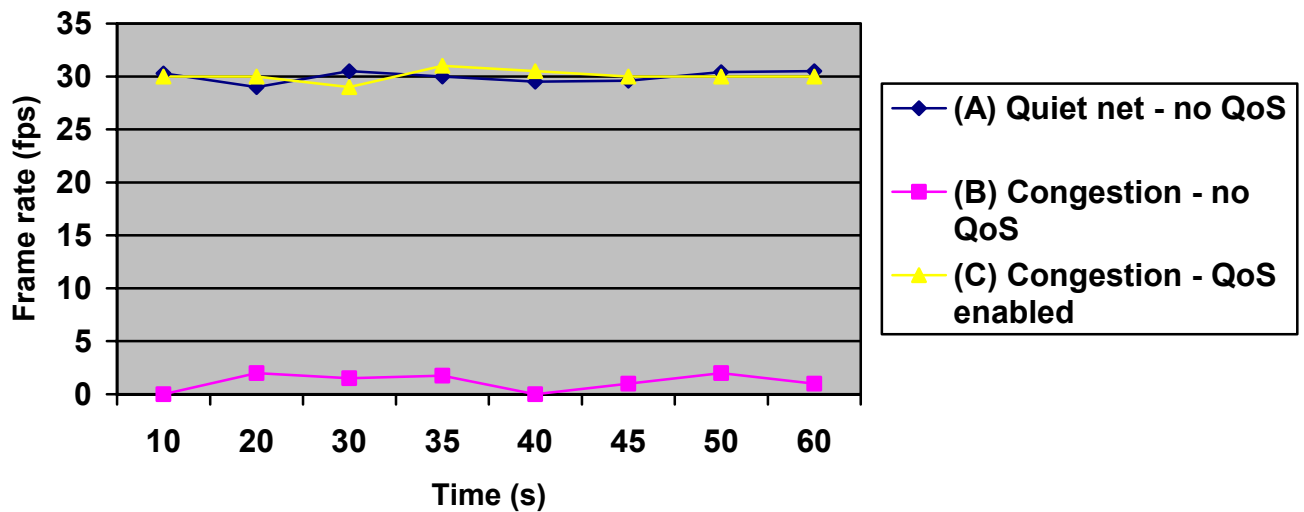
6 QoS over Internet/WAN

To get true DiffServ QoS over a WAN, we must consider the route taken by the network traffic.
To provide end-to-end QoS the whole path must be DiffServ-enabled. The route often goes via one or more ISP backbones. The level of QoS an ISP can provide is usually stated in the SLA, *Service Level Agreement*, between the end user and the ISP.

7 TCP ECN

7.1 What is TCP ECN?

ECN is an acronym for **Explicit Congestion Notification** and is an improvement of the congestion control in TCP.

The basic idea is for TCP to report that congestion is about to occur, before the queue actually overflows. The router does this by marking a field in the IP header called **CE** (**Congestion Experienced**). When the sender receives this signal the flow is slowed down. The traditional method of indicating congestion is for routers to drop packets that lead to the re-transmission of packets.

The ECN model will only be used when both communication end-points support it. This is achieved by negotiating at TCP connection initialization time.

By introducing ECN support re-transmission over the network is minimized, which leads to higher throughput and less delay.

7.2 TCP ECN in AXIS video products

TCP ECN is enabled by default in AXIS video products and can be disabled from the plain config web page (root.Network.tcpECN).

8 Abbreviations

DS   Differentiated Service
FPS   Frames Per Second
PTZ   Pan/Tilt/Zoom
ECN   Explicit Congestion Notification
LAN   Local Area Network
WAN   Wide Area Network
RTP   Real Time Protocol
RTSP   Real Time Streaming Protocol
BA   Behavior Aggregate
PHB   Per Hob Behavior
QoS   Quality of Service
SLA   Service Level Agreement
ISP   Internet Service Provider
CoS   Class of Service
L2   OSI Layer 2
L3   OSI Layer 3
CE   Congestion Experienced
TOS   Type of Service

9 Useful links

- http://www.ietf.org/rfc/rfc2475.txt - An Architecture for Differentiated Services
- http://www.ietf.org/rfc/rfc2597 - Assured Forwarding PHB Group
- http://www.ietf.org/rfc/rfc2598.txt - An Expedited Forwarding PHB
- http://www.ietf.org/rfc/rfc3168.txt - The Addition of Explicit Congestion Notification (ECN) to IP
- http://www.ietf.org/rfc/rfc1633.txt - Integrated Services in the Internet Architecture: an Overview