Strategic QoS Design

At-A-Glance

The Quality of Service Challenge

Today there is a virtual explosion of rich media applications on the IP network This explosion of content and media types, both managed and un-managed, requires network architects to take a new look at their Quality of Service (QoS) designs.

Step 1: Articulate Business Intent and Application Relevance

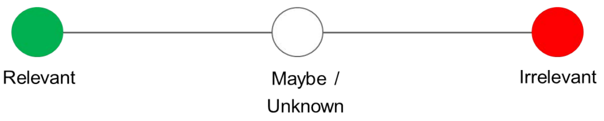
The first step may seem obvious and superfluous, but in actuality it is crucial: clearly define the business objectives that your QoS policies are to enable. These may include any/all of the following:

- Guaranteeing voice quality meets enterprise standards
- Ensuring a high Quality of Experience (QoE) for video
- Increasing user productivity by increasing network response times for interactive applications
- Managing applications that are “bandwidth hogs”
- Identifying and de-prioritizing consumer applications
- Improving network availability
- Hardening the network infrastructure

With these goals in mind, network architects can clearly identify which applications are relevant to their business. Conversely, this exercise will also make it apparent which applications are *not* relevant towards achieving business objectives. Such applications may include consumer-oriented and/or entertainment-oriented applications.

Finally, there may be applications/protocols that can fall into either category of business relevance. For example, HTTP/HTTPS may carry business-relevant traffic or consumer-oriented traffic, and as such cannot be clearly classified in either category. Note: in such cases, deep packet inspection technologies may be able to discretely identify the applications being transported, allowing these to be properly classified in line with business objectives.

Figure 1 Determining Application Business Relevance



Step 2: Define an End-to-End QoS Design Strategy

Once applications have been defined as business-relevant (or otherwise), then the network architect must decide how to mark and treat these applications over the IP infrastructure.

To this end, Cisco advocates following relevant industry standards and guidelines, as this extends the effectiveness of your QoS policies beyond your direct administrative control. That being said, it may be helpful to overview a relevant RFC for QoS marking and provisioning: RFC 4594, “Configuration Guidelines for DiffServ Service Classes.”

These guidelines are to be viewed as industry best-practice recommendations. As such, enterprises and service providers are encouraged to adopt these marking and provisioning recommendations with the aim of improving QoS consistency, compatibility, and interoperability. However, it should be noted that these guidelines are not standards; as such, modifications can be made to these recommendations as specific needs or constraints require.

Thus, to meet specific business requirements, Cisco has made a minor modification to its adoption of RFC 4594: specifically the swapping of Call-Signaling and Broadcast Video markings (to CS3 and CS5, respectively). A summary of Cisco’s implementation of RFC 4594 is presented in Figure 2.

Figure 2 Cisco (RFC 4594-Based) QoS Recommendations

Application Class	Per-Hop Behavior	Queuing and Dropping
Voice	EF	Priority Queue (PQ)
Broadcast Video	CS5	(Optional) PQ
Real-Time Interactive	CS4	(Optional) PQ
Multimedia Conferencing	AF4	BW Queue + DSCP WRED
Multimedia Streaming	AF3	BW Queue + DSCP WRED
Network Control	CS6	BW Queue
Call-Signaling	CS3	BW Queue
Ops/Admin/Mgmt (OAM)	CS2	BW Queue
Transactional Data	AF2	BW Queue + DSCP WRED
Bulk Data	AF1	BW Queue + DSCP WRED
Best Effort	DF	Default Queue + RED
Scavenger	CS1	Min BW Queue

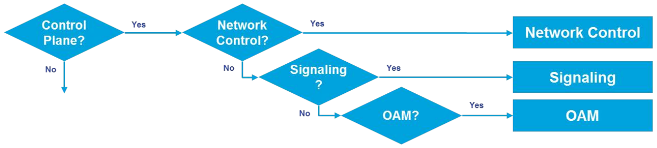
RFC 4594 also provides some application classification rules to help network architects to assign applications to the optimal traffic classes; these are summarized in the following sections:

Business relevant application can be grouped into one of four main categories:

- control plane protocols
- voice applications
- video applications
- data applications

Beginning with the control plane protocols, these may be sub-divided further, as shown in Figure 3.

Figure 3 Control Plane Traffic Classes



- **Network Control**—This traffic class is intended for network control plane traffic, which is required for reliable operation of the enterprise network. Traffic in this class should be marked CS6 and provisioned with a (moderate, but dedicated) guaranteed bandwidth queue. WRED should not be enabled on this class, as network control traffic should not be dropped. Example traffic includes EIGRP, OSPF, BGP, HSRP, IKE, etc.

- **Signaling**—This traffic class is intended for signaling traffic that supports IP voice and video telephony. Traffic in this class should be marked CS3 and provisioned with a (moderate, but dedicated) guaranteed bandwidth queue. WRED should not be enabled on this class, as signaling traffic should not be dropped. Example traffic includes SCCP, SIP, H. 323, etc.

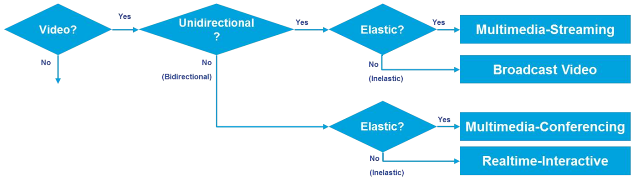
- **Operations/Administration/Management (OAM)**—This traffic class is intended for network operations, administration, and management traffic. This class is critical to the ongoing maintenance and support of the network. Traffic in this class should be marked CS2 and provisioned with a (moderate, but dedicated) guaranteed bandwidth queue. WRED should not be enabled on this class, as OAM traffic should not be dropped. Example traffic includes SSH, SNMP, Syslog, etc.

Provisioning for voice is relatively straightforward:

• **Voice**—This traffic class is intended for voice/audio traffic (VoIP signaling traffic is assigned to the "Call-Signaling" class). Traffic assigned to this class should be marked EF. This class is provisioned with an Expedited Forwarding (EF) Per-Hop Behavior (PHB). The EF PHB-defined in RFC 3246-is a strict-priority queuing service and, as such, admission to this class should be controlled. Example traffic includes G.711 and G.729a, as well as the audio components of multimedia conferencing applications, like Cisco Jabber, WebEx and Spark.

Video—on the other hand—may have unique QoS requirements depending on the type, as illustrated in Figure 4.

Figure 4 Video Traffic Classes



Two key questions need to be answered to determine the optimal traffic classification for a video application :
• is the video unidirectional or bidirectional?
• is the video elastic or inelastic?

"Elastic" flows are able to adapt to network congestion and/or drops (by reducing frame rates, bit rates, compression rates, etc.); "inelastic" flows either do not have such capabilities or—in order to meet specific business configured not to utilize these.

With these two questions answered, video applications may be assigned to their respective traffic classes, including:

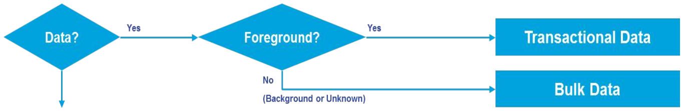
• **Broadcast Video**—This traffic class is intended for broadcast TV, live events, video surveillance flows, and similar "inelastic" streaming video flows Traffic in this class should be marked Class Selector 5 (CS5) and may be provisioned with an EF PHB; as such, admission to this class should be controlled. Example traffic includes live Cisco Enterprise TV (ETV) streams, and Cisco IP Video Surveillance.

• **Real-Time Interactive**—This traffic class is intended for inelastic interactive video applications. Whenever possible, signaling and data sub-components of this class should be separated out and assigned to their respective traffic classes. Traffic in this class should be marked CS4 and may be provisioned with an EF PHB; as such, admission to this class should be controlled. An example application is Cisco TelePresence.

• **Multimedia Conferencing**—This traffic class is intended for elastic interactive multimedia collaboration applications. Whenever possible, signaling and data sub-components of this class should be separated out and assigned to their respective traffic classes. Traffic in this class should be marked Assured Forwarding (AF) Class 4 (AF41) and should be provisioned with a guaranteed bandwidth queue with DSCP-based Weighted-Random Early Detect (DSCP-WRED) enabled. Traffic in this class may be subject to policing and re-marking. Example applications include Cisco Jabber, WebEx and Spark.

• **Multimedia Streaming**—This traffic class is intended for elastic streaming video applications, such as Video-on-Demand (VoD). Traffic in this class should be marked AF Class 3 (AF31) and should be provisioned with a guaranteed bandwidth queue with DSCP-based WRED enabled. Example applications include Cisco Digital Media System Video-on-Demand (VoD) streams, E-Learning videos, etc.

Figure 5 Data Traffic Classes



When it comes to data applications, there is really only one key question to answer (as illustrated in Figure 5):
• Is the data application "foreground" or "background"?

"Foreground" refers to applications from which users expect a response—via the network—in order to continue with their tasks; excessive latency to such applications will directly impact user productivity.

Conversely, "background" applications—while business relevant—do not directly impact user productivity and typically consist of machine-to-machine flows.

• **Transactional Data**—This traffic class is intended for interactive, "foreground" data applications Traffic in this class should be marked AF Class 2 (AF21) and should be provisioned with a dedicated bandwidth queue with DSCP-WRED enabled. This traffic class may be subject to policing and re-marking. Example applications include data components of multimedia collaboration applications, Enterprise Resource Planning (ERP) applications, Customer Relationship Management (CRM) applications, database applications, etc.

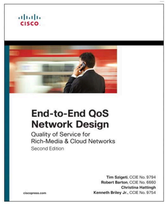
• **Bulk Data**—This traffic class is intended for non-interactive "background" data applications Traffic in this class should be marked AF Class 1 (AF11) and should be provisioned with a dedicated bandwidth queue with DSCP-WRED enabled. This traffic class may be subject to policing and re-marking. Example applications include: E-mail, backup operations, FTP/SFTP transfers, video and content distribution, etc.

With all business-relevant applications assigned to their respective traffic classes, then only two types of traffic classes are left to be provisioned:

• **Best Effort** (the Default Class)—This traffic class is the default class. The vast majority of applications will continue to default to this Best-Effort service class; as such, this default class should be adequately provisioned. Traffic in this class is marked Default Forwarding (DF or DSCP 0) and should be provisioned with a dedicated queue. WRED is recommended to be enabled on this class.:

• **Scavenger**—This traffic class is intended for all applications that have been previously identified as business-irrelevant. These may include video applications that are consumer and/or entertainment-oriented. The approach of a "less-than Best-Effort" service class for non-business applications (as opposed to shutting these down entirely) has proven to be a popular, political compromise. These applications are permitted on business networks when bandwidth is available; however, as soon as the network experiences congestion, this class is the most aggressively dropped. Traffic in this class should be marked CS1 and should be provisioned with a minimal bandwidth queue that is the first to starve should network congestion occur. Example traffic includes Netflix, YouTube, Xbox Live/360 Movies, iTunes, BitTorrent, etc

For more details, see:
http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/QoS_SRND_40/QoSIntro_40.html
And the Cisco Press Book: **End-to-End QoS Network Design** (Second Edition)-Chapter 10


End-to-End QoS
Network Design
Quality of Service for
Rich-Media & Cloud Networks
Second Edition
Tim Szigeti, CCIE No. 9794
Robert Barton, CCIE No. 6660
Christina Hattingh
Kenneth Briley Jr., CCIE No. 9754
ciscopress.com