Implementing Quality of Service

Document ID: 13747

Contents

Introduction

Which Applications Need QoS?
    Understanding the Characteristics of Applications
Knowing the Network Topology
    Link Layer Header Sizes
Creating Classes Based on the Criteria
Creating a Policy to Mark Each Class
Working from the Edge Towards the Core
Building the Policy to Treat the Traffic
Applying the Policy
Using QoS Policy Manager (QPM) to Monitor the Effects of the Policy
General Purpose QoS Recommendations
Related Information

Introduction

This document provides some high level guidelines for implementing Quality of Service (QoS) in a network that serves as a transport for multiple applications, including delay−sensitive and bandwidth−intensive applications. These applications may enhance business processes, but stretch network resources. QoS can provide secure, predictable, measurable, and guaranteed services to these applications by managing delay, delay variation (jitter), bandwidth, and packet loss in a network.

Which Applications Need QoS?

First, determine which applications are business−critical and require protection. You may have to review all the applications that are competing for network resources. If this is the case, use Netflow Accounting, Network−based Application Recognition (NBAR), or QoS Device Manager (QDM) to analyze the traffic patterns in the network.

NetFlow Accounting provides detail about network traffic and can be used to capture the traffic classification or precedence associated with each flow.

NBAR is a classification tool that can identify traffic up to the application layer. It provides per−interface, per−protocol, and bi−directional statistics for each traffic flow traversing an interface. NBAR also does sub−port classification; looking and identifying beyond application ports.

QDM is a Web−based network management application that provides an easy−to−use graphical user interface for configuring and monitoring advanced IP−based QoS functionality in routers.

Understanding the Characteristics of Applications

It is important to understand the characteristics of the applications that need protection. Some applications tend to be sensitive to latency or packet loss, while others are considered "aggressive" because they are bursty or consume a lot of bandwidth. If the application is bursty, determine if there is a constant burst or a small burst. Is the packet size of the application large or small? Is the application TCP or UDP based?

Characteristic	Guideline
Application that is delay− or loss−sensitive. (Voice and Real Time Video)	Do *not* use weighted random early detection (WRED), traffic shaping, fragmentation (FRF−12), or policing. For this kind of traffic, you should implement Low Latency Queuing (LLQ) and use a priority queue for the delay−sensitive traffic.
Application that is consistently bursty or is a bandwidth hog. (FTP and HTTP)	Use WRED, policing, traffic shaping, or class−based weighted fair queueing (CBWFQ) to guarantee bandwidth.
Application that is TCP−based.	Use WRED to drop packets to cause TCP to back off and then ramp up again using the slow−start algorithm. If the traffic is UDP−based and does not change its behavior when packets are dropped, do not use WRED. Use Policing if you need to rate−limit the application; otherwise just let the packets tail−drop.

Knowing the Network Topology

Some devices may need an IOS upgrade in order to take advantage the QoS features you want to implement. Diagrams of the network topology, router configurations, and software version on each device help you estimate the number of devices requiring an IOS upgrade. Refer to the Cisco Icon Library for icons that can help you create network diagrams.

- Assess the CPU utilization on each router during busy periods to help decide how to distribute QoS features among devices to share the load.
- Classify business−critical traffic types and the interfaces this traffic will traverse. Decide which priority groups or classes to create to realize the QoS goals for your network.
- Determine the maximum delay that the most critical applications can handle and adjust the burst parameters within traffic conditioners (traffic shapers or policers) to accommodate this delay.
- Find out what rates are supported on each interface: PVCs or subinterfaces and configure the bandwidth to match.
- Identify slow links to help determine where the bottlenecks in the network are located and decide how to apply Link Efficiency mechanisms at the appropriate interfaces.
- Calculate the Layer 2 and Layer 3 overhead for each media type that will transport the business critical traffic. This will help calculate the correct amount of bandwidth needed for each class.
- Another key piece of information is whether you want to protect traffic based on application, IP source and destination, or both.

Link Layer Header Sizes

Media Type	Link Layer Header
Ethernet	14 Bytes
PPP	6 Bytes
Frame Relay	4 Bytes

ATM	5 Bytes/Cell

Creating Classes Based on the Criteria

Once you determine which applications need QoS and the classification criteria to use (based on the characteristics of the applications), you are ready to create classes based on this information.

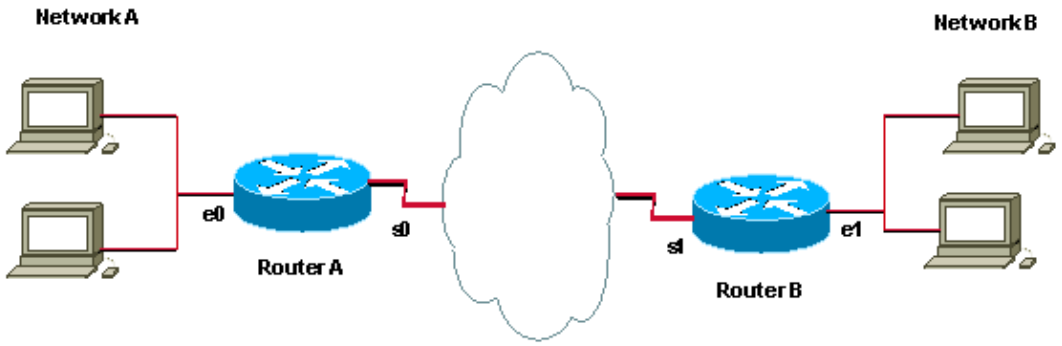
Creating a Policy to Mark Each Class

Create a policy to mark each class of traffic with the appropriate priority values (use differentiated services control point (DSCP) or IP Precedence). The traffic will be marked as it comes into the router on the ingress interface. The markings will be used to treat the traffic as it leaves the router on the egress interface.

Working from the Edge Towards the Core

Work from the router closest to the traffic towards the core. Apply your marking on the ingress interface of the router. In the topology below, Router A is the obvious place to mark traffic and apply policy for data sourced from Network A and destined for Router B. The traffic will be marked as it comes into Router A's Ethernet0 interface, and the QoS policy will be applied on Router A's Serial0 interface as it leaves the router. If the same policy should be applied in both directions (so that traffic sourced from Network B and destined for Network A receives the same treatment), the traffic coming from Network B should be marked as it comes into the Router B's Ethernet1 interface and treated as it leaves the router on the Serial1 interface.

Once traffic is marked on the ingress interface on one router, it maintains the same markings as it traverses multiple hops (unless it is re−marked). Usually, traffic only needs to be marked once. QoS policies can be applied on additional hops based upon these markings. You should only have to re−mark in the event that traffic is arriving from an untrusted domain.



Building the Policy to Treat the Traffic

Now that you have marked the traffic, you can use the markings to build a policy and do traffic classification on the rest of the network segments. We recommend keeping the policy simple by using no more than four classes.

If possible, implement and test a QoS implementation in a lab environment. Deploy it in the live network after you are satisfied with the results.

Applying the Policy

Apply the policy in the appropriate direction. Decide if the policy needs to be applied in one direction or in both directions. Always mark and treat traffic as close to the source as possible, as described in the Creating a Policy to Mark Each Class section of this document.

We recommend that you apply the same policy in both directions in order to filter traffic arriving from and destined to both sides of the site. This means you should apply the same policy outbound on the serial interface of RouterA and the serial interface of RouterB.

Using QoS Policy Manager (QPM) to Monitor the Effects of the Policy

Use QPM as a complete system for centralized policy control and automated, reliable policy deployment.

General Purpose QoS Recommendations

Below is a list of QOS categories and some of the more widely− used QOS features associated with each category.

Category	Associated QoS Features
QoS Service Model	Provisioned (Diffserv) QoS when possible or signaled (RSVP) when necessary.
Classification/Marking	Diffserv Code Points or qos−group ID.
Congestion Management	LLQ or CBWFQ.
Congestion Avoidance	Diffserv−compliant WRED.
Link Efficiency	MLPPP, LFI, FRF.11, FRF.12, CRTP
Signaling	RSVP, QPPB
Traffic Conditioners/Policing	Class Based Policer and Generic Traffic Shaping (GTS) or Frame−Relay Traffic Shaping (FRTS).
Configuration/Monitoring	QPM, Modular QoS Command Line Interface (CLI), QDM

Related Information

- **QoS Support Page**
- **IP Routed Protocols Support Page**
- **IP Routing Support Page**
- **IS−IS Support Page**
- **Technical Support − Cisco Systems**

Contacts & Feedback | Help | Site Map
© 2014 − 2015 Cisco Systems, Inc. All rights reserved. Terms & Conditions | Privacy Statement | Cookie Policy | Trademarks of Cisco Systems, Inc.

Updated: Feb 15, 2008 Document ID: 13747