CoS for MPLS VPNs with IP Quality of Service

Christian Kuster
Data Network Design
TDC Switzerland AG / sunrise
CoS Presentation Swinog 9
29. September 2004

CoS vs. QoS

Class of Service (CoS) is the service provided by a network to the customer traffic.

In IP based networks, this service is achieved by using IP Quality of Service (QoS) mechanisms and features.

More Bandwidth? Is that always the Answer?

PROS

- Increases capacity
- Resolves immedi

CONS

- Short-term solution
- Expensive $$$
- Will not guarantee applications with low latency tolerance such as VoIP and video conferencing
- All applications receive same service, no protection for mission-critical applications
- Emerging applications could jeopardize business critical traffic

How Can QoS Be Applied?

- **Best effort**—no QoS is applied to packets (default behavior)
- **Integrated Services** model—applications signal to the network that they require special QoS
- **Differentiated Services** model—the network recognizes classes that require special QoS

Integrated Services Model (Intserv)

Benefits and Drawbacks of the IntServ Model

+ **RSVP benefits:**
  - **Explicit resource admission control (end-to-end)**
  - **Per-request policy admission control (authorization object, policy object)**
  - **Signaling of dynamic port numbers (for example, H.323)**
- **RSVP drawbacks:**
  - **Continuous signaling due to stateless architecture**
  - **Not scalable**
  - **Needs modified (RSVP supporting) applications**

Differentiated Services Model (DiffServ)

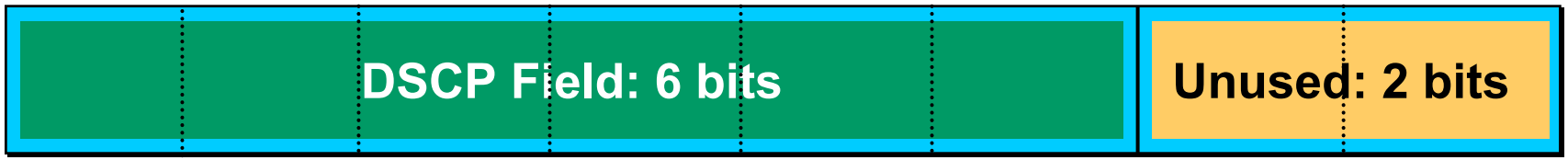
Differentiated Services Model

- TheDifferentiated Services model describes services associated with traffic classes.
- Complex traffic classification and conditioning are performed at network edge, resulting in a per-packet Differentiated Services Code Point (DSCP).
- No per-flow/per-application state exists in the core.
- The core performs only simple “per-hop behaviors” on traffic aggregates.
- The goal is scalability.

Why Is Provisioning Important?

- QoS does not create bandwidth!
- QoS manages bandwidth usage among multiple classes.
- QoS gives better service to a well-provisioned class with respect to another class.

Packet Header Terminology



Former ToS Byte = New DS Field

- **DSCP**: a specific value of the DSCP portion of the DS field. The DSCP is used to select a PHB (Per-Hop Behavior; forwarding and queuing method)
- **DS field**: the IPv4 header ToS octet or the IPv6 traffic class octet when interpreted in conformance with the definition given in RFC 2474. The bits of the DSCP field encode the DSCP, while the remaining bits are currently unused.

DSCP Encoding

- Three pools:
  - “xxxxx0” Standard Action
  - “xxxx11” Experimental/Local Use
  - “xxxx01” EXP/LU (possible std action)
- Default DSCP: “000000”
- Default PHB: FIFO, tail-drop

DSCP Usage

DSCP selects per-hop behavior (PHB) throughout the network:

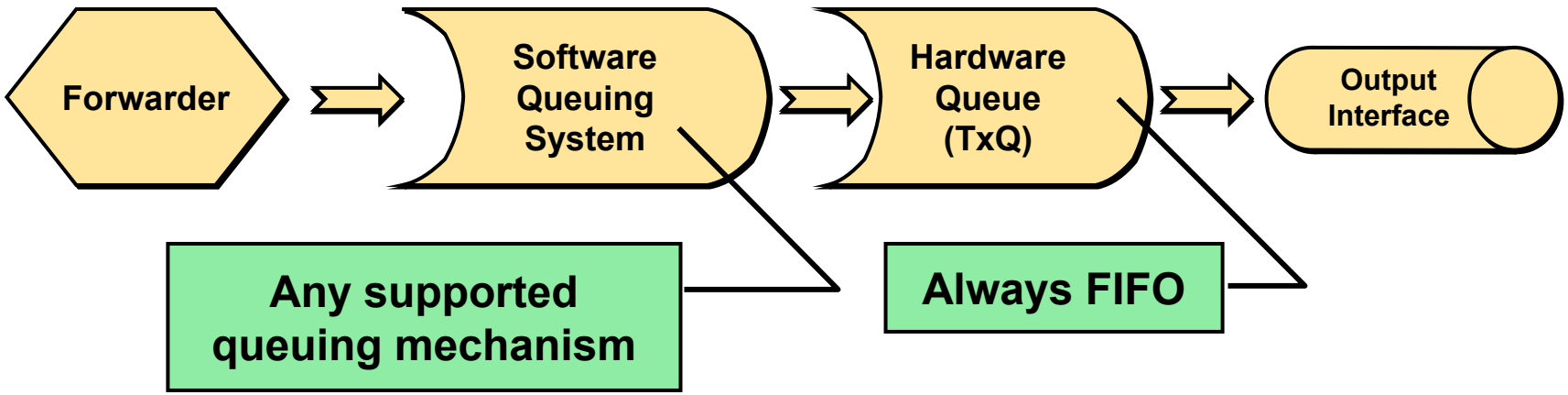
- **Default PHB**
- **Class selector (IP Precedence) PHB**
- **Expedited forwarding PHB**
- **Assured forwarding PHB**

Queuing Overview

Queuing in Cisco IOS

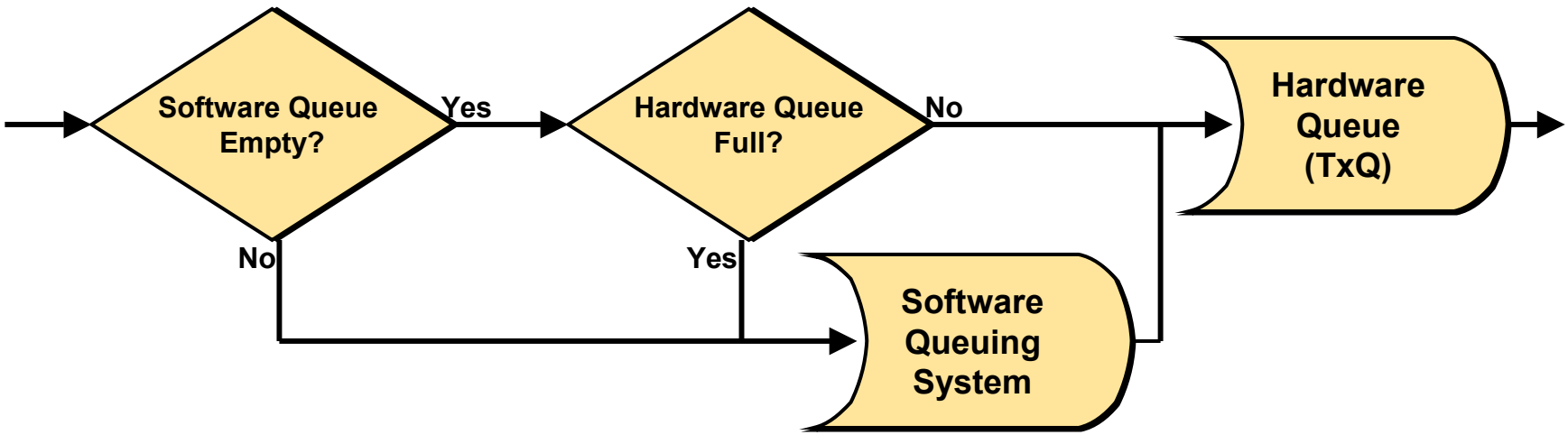
- Cisco routers running Cisco IOS have a number of different legacy queuing mechanisms
- Let's have a look first at the following
  - First In First Out (FIFO)
  - Weighted Fair Queuing (WFQ) with the different distributed versions
  - *Priority Queuing (PQ)*
  - *Custom Queuing (CQ)*
  - *Modified Deficit Round Robin (MDRR)*
  - *IP RTP Prioritization*
- These mechnisms are implemented as software queues

Output Interface Queue Structure



- Each interface has its hardware and software queuing system.
- The hardware queuing system (transmit queue, or TxQ) always uses FIFO queuing.
- The software queuing system can be selected and configured depending on the platform and Cisco IOS version.

Bypassing the Software Queue

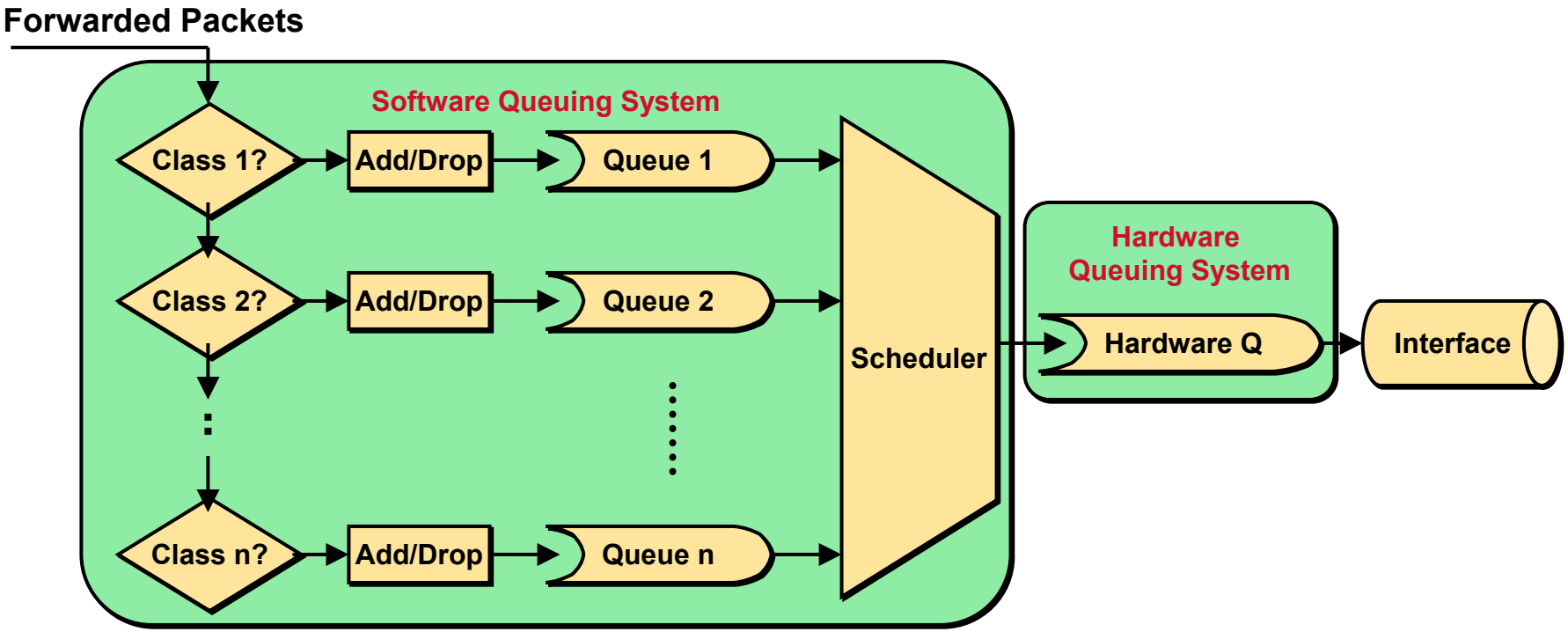


- When a packet is being forwarded, the router will bypass the software queue if:
  - The software queue is empty, and
  - The hardware queue is not full

Hardware Queue (TxQ) Size

- Routers determine the length of the hardware queue based on the configured bandwidth of the interface.
- Long TxQ may result in poor performance of the software queue.
- Short TxQ may result in a large number of interrupts which causes high CPU use and low link use.

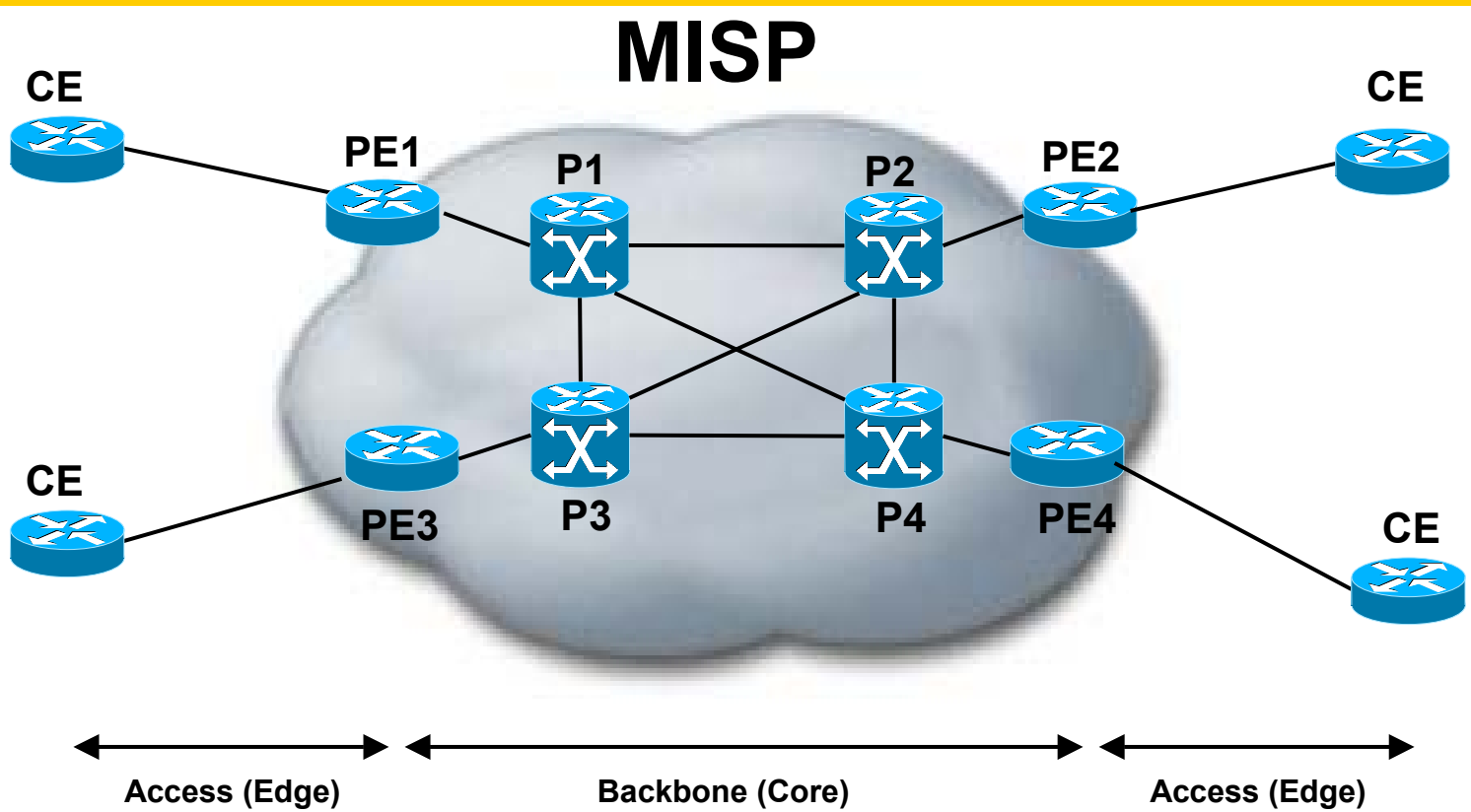
Queuing Components



- Each queuing mechanism has three main components that define it:
  - Classification (selecting the class)
  - Insertion policy (determining whether a packet can be enqueued)
  - Service policy (scheduling packets to be put into the hardware queue)

CoS Implementation

Typical Service Provider Networks



QoS in Service Provider Networks

- Service providers can extend their service offerings by introducing quality.
- Customers can get bandwidth guarantees (like CIR in Frame Relay).
- Customers can get delay guarantees (like CBR in ATM).
- QoS mechanisms have to be deployed where congestion is likely (usually at the network edge).
- The customer traffic is identified based on source or destination IP addresses or application type.

Classes vs. DiffServ

We offer Class 1, Class 2 and Class 3 services in three different combinations:

- **Class 3 is best effort (default PHB)**
  - **DSCP 0 (default)**
- **Class 2 is Assured forwarding (AF PHB)**
  - **DSCP AF41 (34) for in-contract**
  - **DSCP AF43 (38) for out-of-contract**
- **Class 1 is Expedited forwarding (EF PHB)**
  - **DSCP EF (46)**

Profiles

Option (profile) A :

- Class 3 gets 10% of available bandwidth
- Class 2 gets 50% of available bandwidth
- Class 1 gets 40% of available bandwidth with a low-delay guarantee

Profiles

Option (profile) B :

- Class 3 gets 60% of available bandwidth
- Class 1 gets 40% of available bandwidth with a low-delay guarantee

Profiles

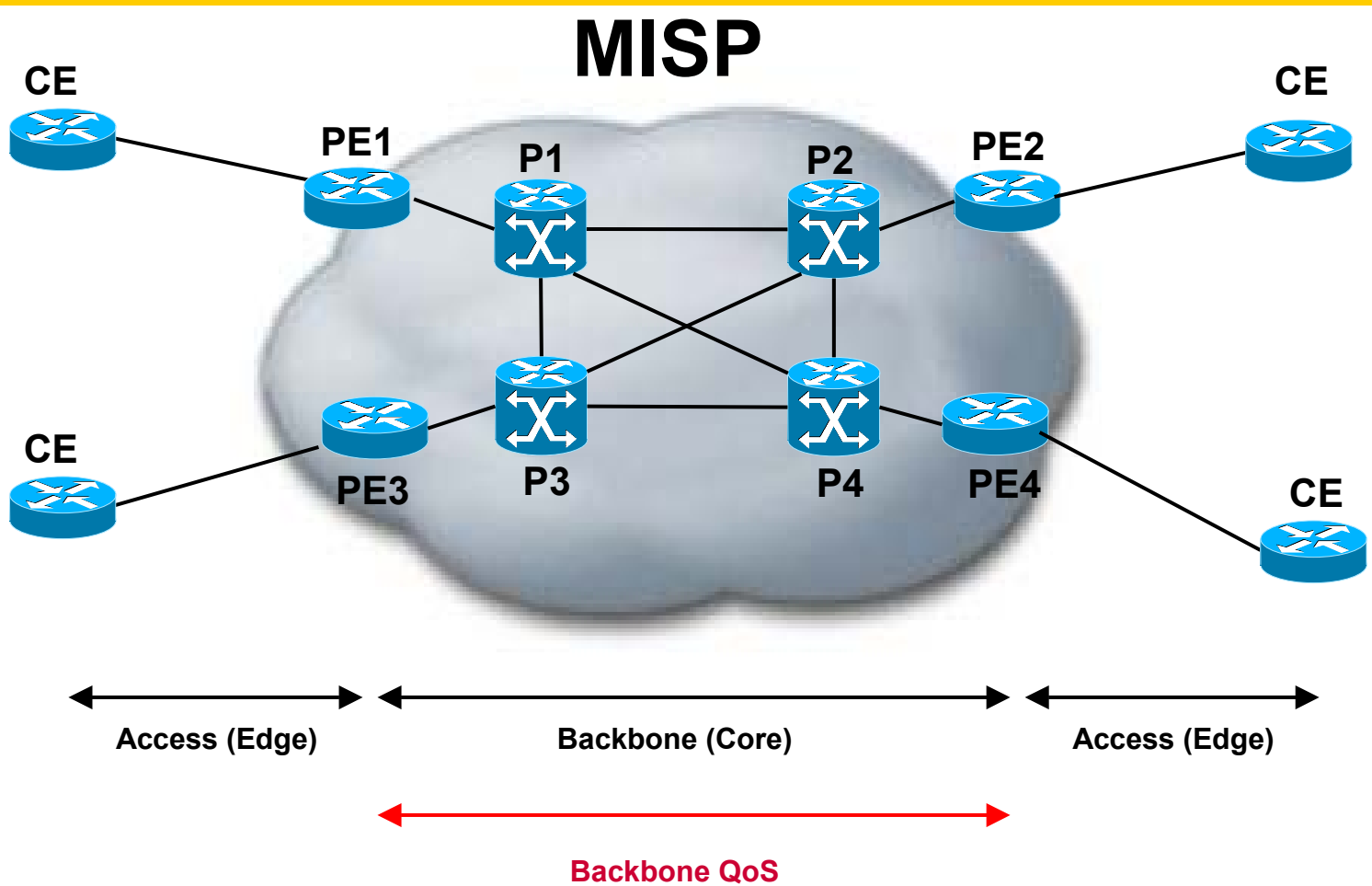
Option (profile) C :

- Class 3 gets 10% of available bandwidth
- Class 2 gets 90% of available bandwidth

Profiles

- IP QoS in the Backbone
  - Between PE – P / P - P routers
- IP QoS in the access part
  - Between PE - CE routers

Backbone QoS



Backbone CoS

- The Key is OverProvisioning
  Offer must be higher than Demand
- The service that traffic receives is dependent upon the ratio of traffic load to available capacity
- More Bandwidth (offer) than traffic (demand) means
  - Low loss
  - Low Latency
  - Low Jitter

Over-Provisioned Backbone

A simple rule of design:

95-Percentile (5-min average Load) <= 50% Link

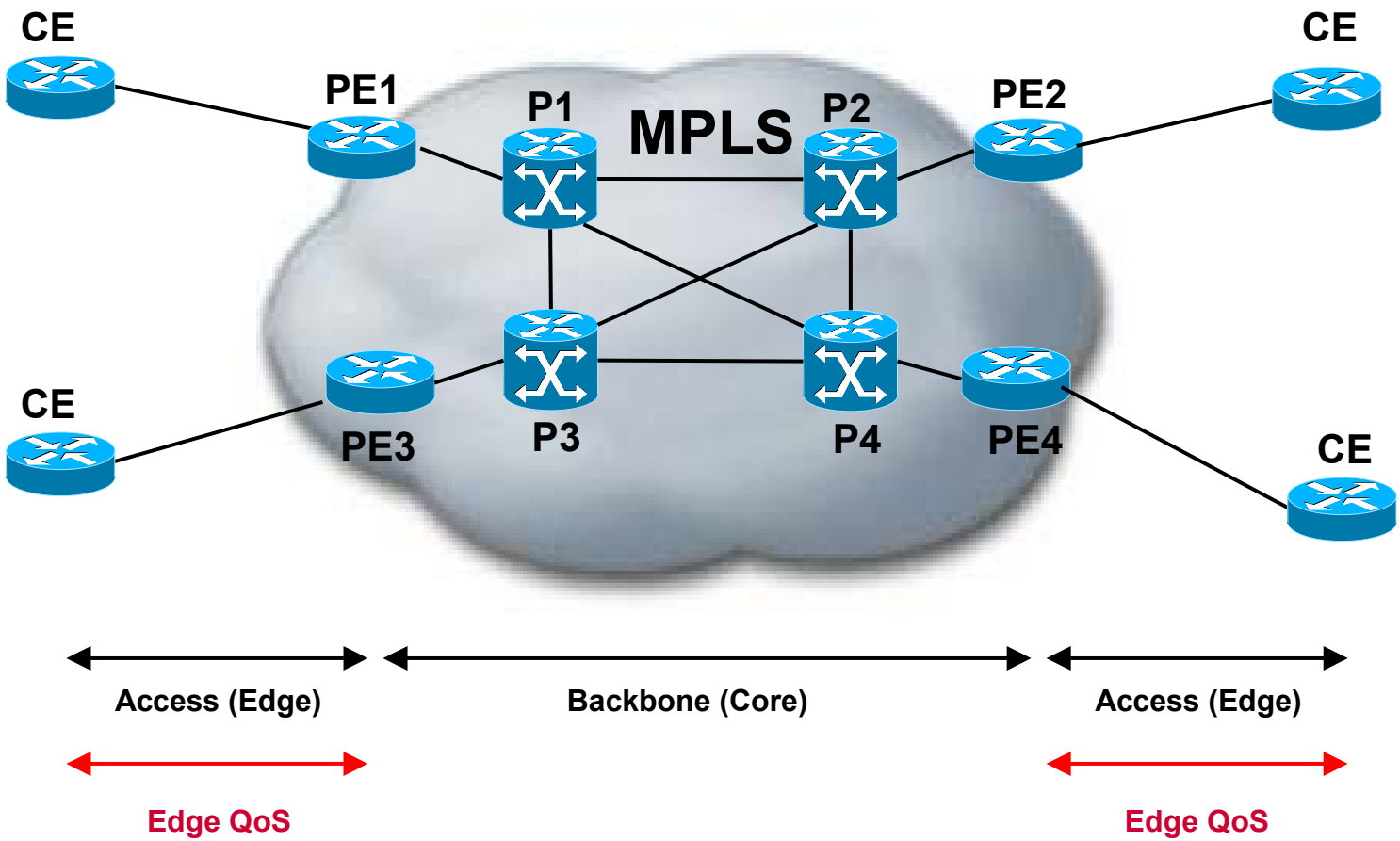
which means

OverProvisioning (OP) > 2

Backbone QoS

- What happens if OP factor <=2 ?
- Prerequisite is a constant monitoring of the backbone links.
- As soon as we approach 50% load on those links, the decision has to be made if the capacity (bandwidth) is increased (e.g. STM-1 to STM-16 / STM-16 to STM-64) or we design and deploy a MPLS aware QoS in the backbone.

Edge (access) QoS



Edge (access) QoS

-Customer traffic classification
-IP packet marking (DSCP)
-Class based queuing (CBWFQ, CBLLQ)
-Class-based policing for class 1
-out-of-contract marking for class 2
-Differentiated dropping (Class-based WRED)
-Link fragmentation and Interleave (LFI – FRF.12)

-Class based queuing based on DSCP marking (CBWFQ, CBLLQ)
-Class-based policing for class 1
-Differentiated dropping (Class-based WRED)
-Link fragmentation and Interleave (LFI – FRF.12)

Implementation

- CoS (IP QoS) is standards (Diffserv) compliant
- CoS uses the Cisco IOS Modular QoS CLI (MQC)
- Class-based low latency queuing (CBLLQ) for Class 1
- Class-based weighted fair queuing (CBWFQ) for Class 2 and 3
- Weighted random early detection (WRED) for Class 2 and 3
- FRF.12 Link fragmentation and interleave (LFI) on links up to 768 kbps if Class 1 is present

The invisible classes

What about management traffic (smtp, telnet, tftp etc…) and router internal traffic (keepalives, BGP etc.)?

The two invisible classes :

- **MGT class (for all management traffic)**
- **RP class (for all internal traffic)**

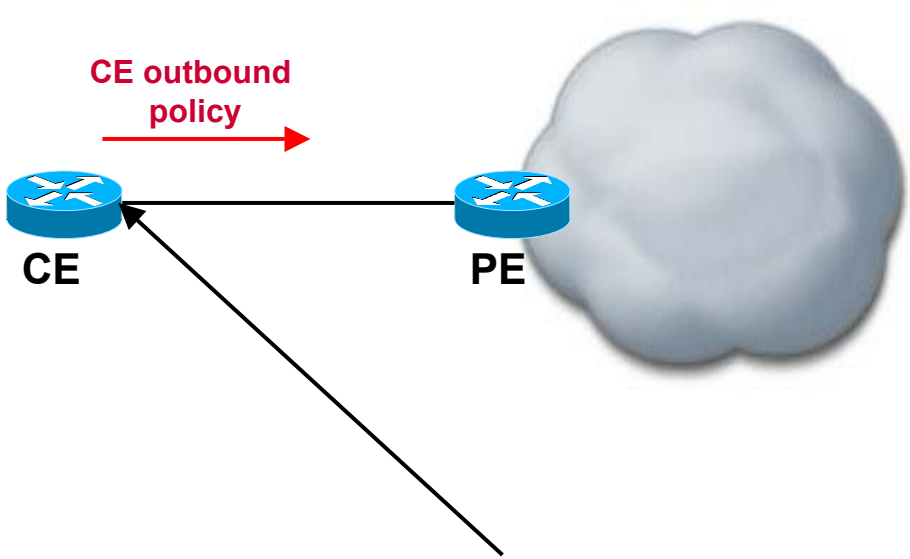
These classes are always configured.

All classes together

So in reality, we have a 5 class model.

- 3 Customer classes
  - class 1, class 2, class 3
- 2 internal classes
  - class MGT class RP

QoS Configuration on CE side



- Decide which CoS profile (CoS classes) used on this site
- Classify customer traffic to the appropriate CoS Class

•Classification is done via the MQC
=> class maps

Classification options

Our CoS option uses the following classification options in the class maps :

- **NBAR – for application**
  - `match protocol` *`protocol`*
- **access lists – for IP addresses / application**
  - `match access-group` *`named-access-list`*
- **IP precedence / DSCP – adapt customer marking**
  - `match ip precedence` *`prec1 [prec2 [prec3 [prec4]]]`*
  - `match ip dscp dscp1` *`[dscp2 [dscp3 [dscp4]]]`*

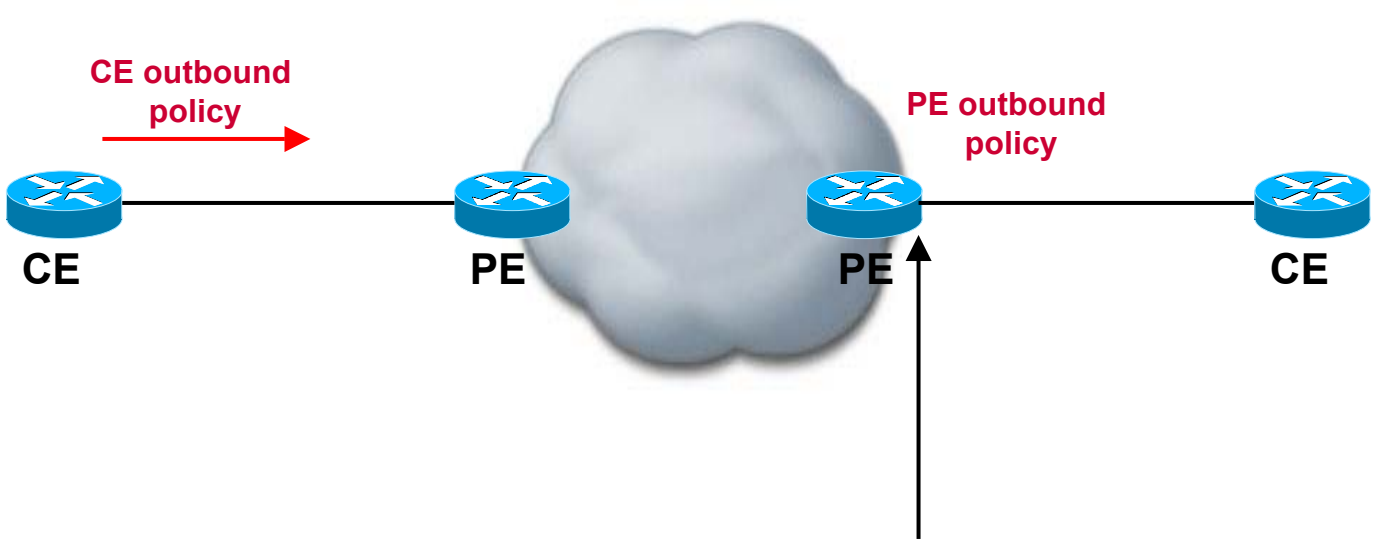
CE service policy

- What is defined in the service policy ?
  - Queuing strategy
    - CBLLQ for Class 1
    - CBWFQ for the other classes
  - Intelligent dropping
    - CBWRED for Class 2 and Class 3
  - Rate limiting
    - CB policing with dropping exceeding traffic for Class 1
    - CB policing with remarking exceeding traffic for Class 2

LFI (FRF.12) and FRTS

- In addition to the service policy, we have to use LFI (FRF.12) on links up to 768 kbps.
- LFI with FRF.12 requires Frame Relay traffic shaping (FRTS).

QoS Configuration on PE side



- Decide which CoS profile (CoS classes) used for this attached site
- The configuration is always the same for a certain profile and a certain link speed
    - Packets are already correctly DSCP marked, no need to classify based on the customer specific requirements

QoS Configuration on PE side

Basically, the same mechanisms are used on the PE side as on the CE.

Due to different platforms (10K / 7600 hardware based, 7500 distributed software platform), the configuration look slightly different (nested policy-maps, no FRTS but class-based shaping, etc)

Classification is done solely based on the DSCP value of a packet

Case Study
Example 1

- The customer experiences problems (slow response times, time-outs, session drops) with his SQL (Oracle) and Citrix applications from time to time, especially when File transfers of his inventory systems are running.
- The traditional approach :
  - Sell more bandwidth, although 90% of the time, bandwidth utilization is only at 50%

Case Study
Example 1

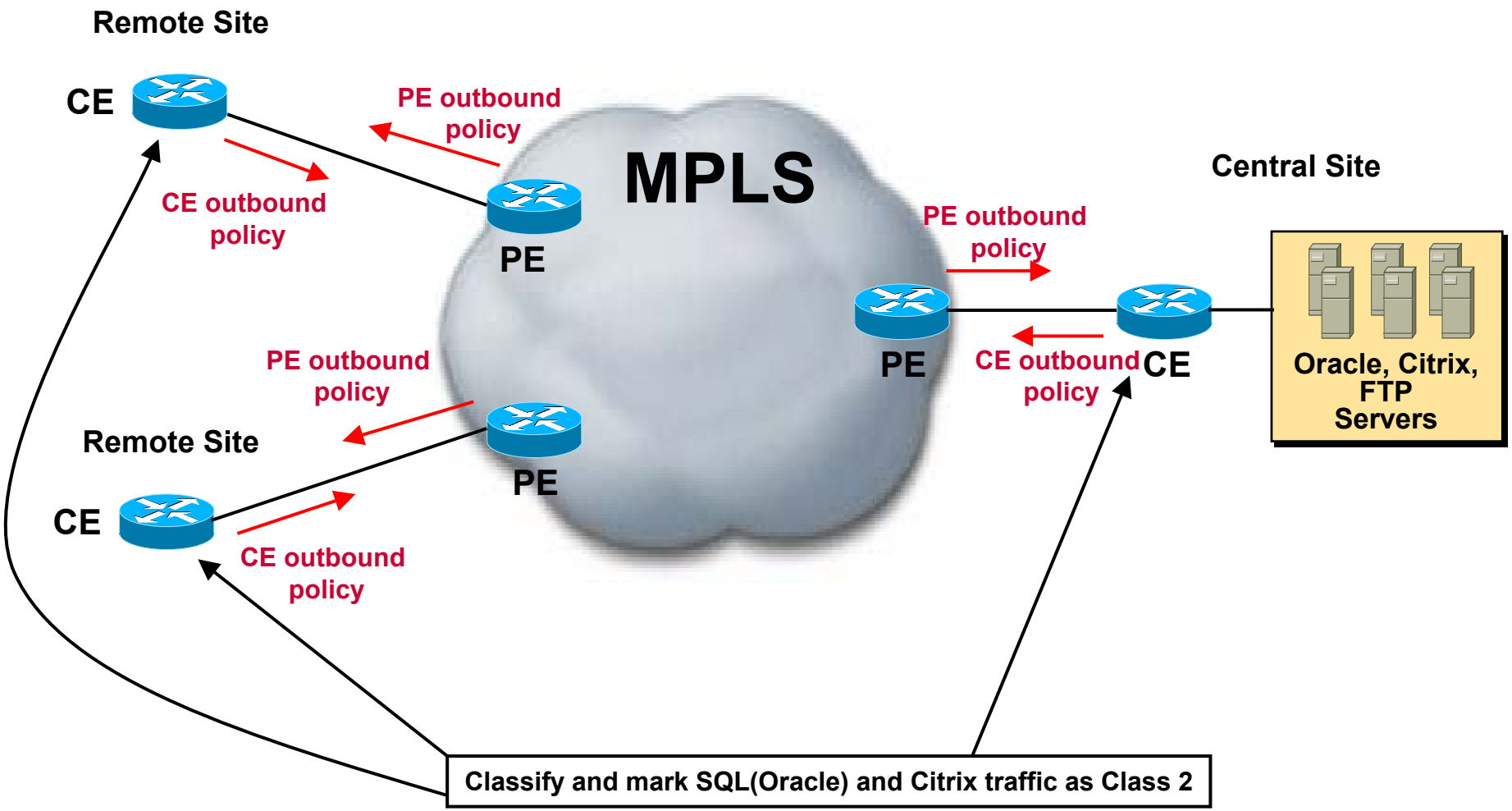
- **Even after the bandwidth upgrade, the customer still is experiencing the same problems**
  - **FTP aggressively eats up all available bandwidth**

➨ **The customer is very unhappy ☹**

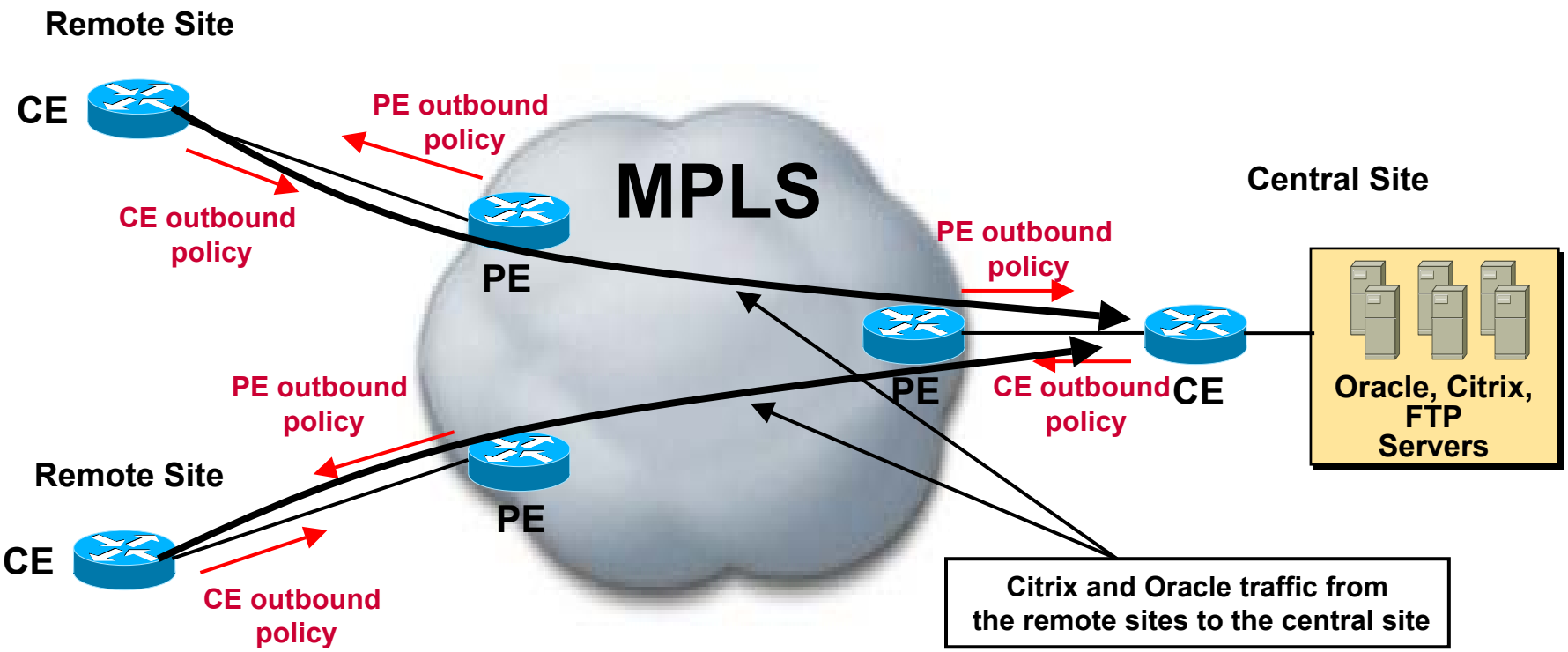
Case Study
Example 1

- The sunrise CoS approach :
  - Sell customer CoS profile C with Class 2 and Class 3
  - Citrix and SQL-oracle traffic is put into Class 2, the rest into Class 3

Case Study Example 1



Case Study
Example 1



- The key is appropriate PE-CE bandwidth provisioning, e.g.
  - 2 remote site à 256 kbps
  - ➨ central site 512 kbps

Case Study
Example 1

- **Result :**
  - **If there is no Oracle and Citrix traffic (e.g. during the night), the File transfer runs at full speed.**
  - **If there is Oracle and Citrix traffic, the File transfer traffic is throttled back to a maximum of 10% of the bandwidth, depending on the amount of Oracle and Citrix traffic.**

Case Study
Example 1

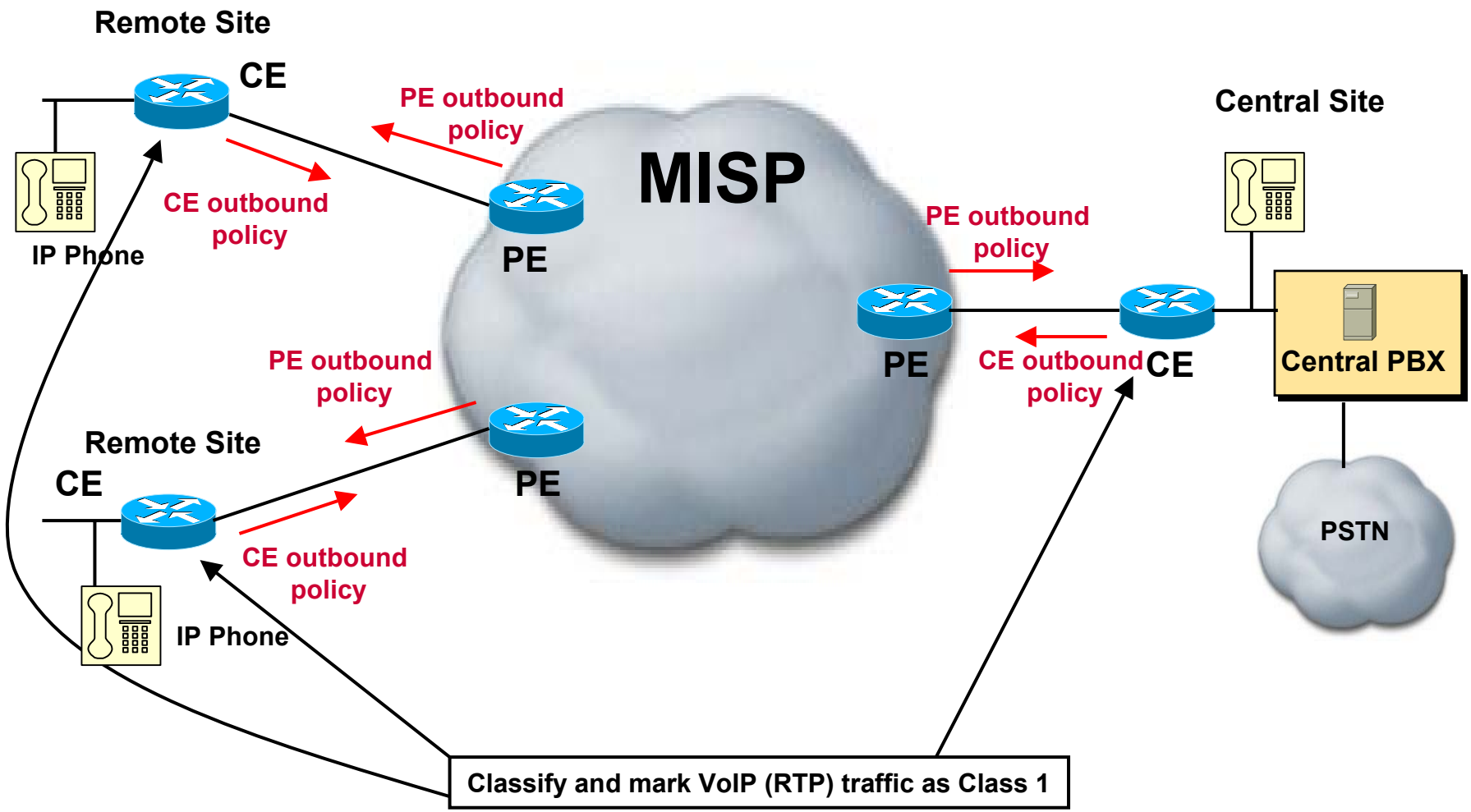
- **Result :**
  - **The customer is paying less than for a bandwidth upgrade AND is getting the desired result (quality performance of Citrix and Oracle).**

➨ **The customer is very happy ☺**

Case Study
Example 2

- A customer has the plan to deploy IP Telephony (VoIP) within his VPN to get rid of his expensive and old PSTN infrastrucure (e.g. a PBX in each branch, dedicated lines etc.)
  - He is happy with the current quality and performance, just wants to add VoIP.
  - ➨ CoS is a MUST for VoIP, so he goes for profile B: Class 1 for VoIP, the rest is Class 3

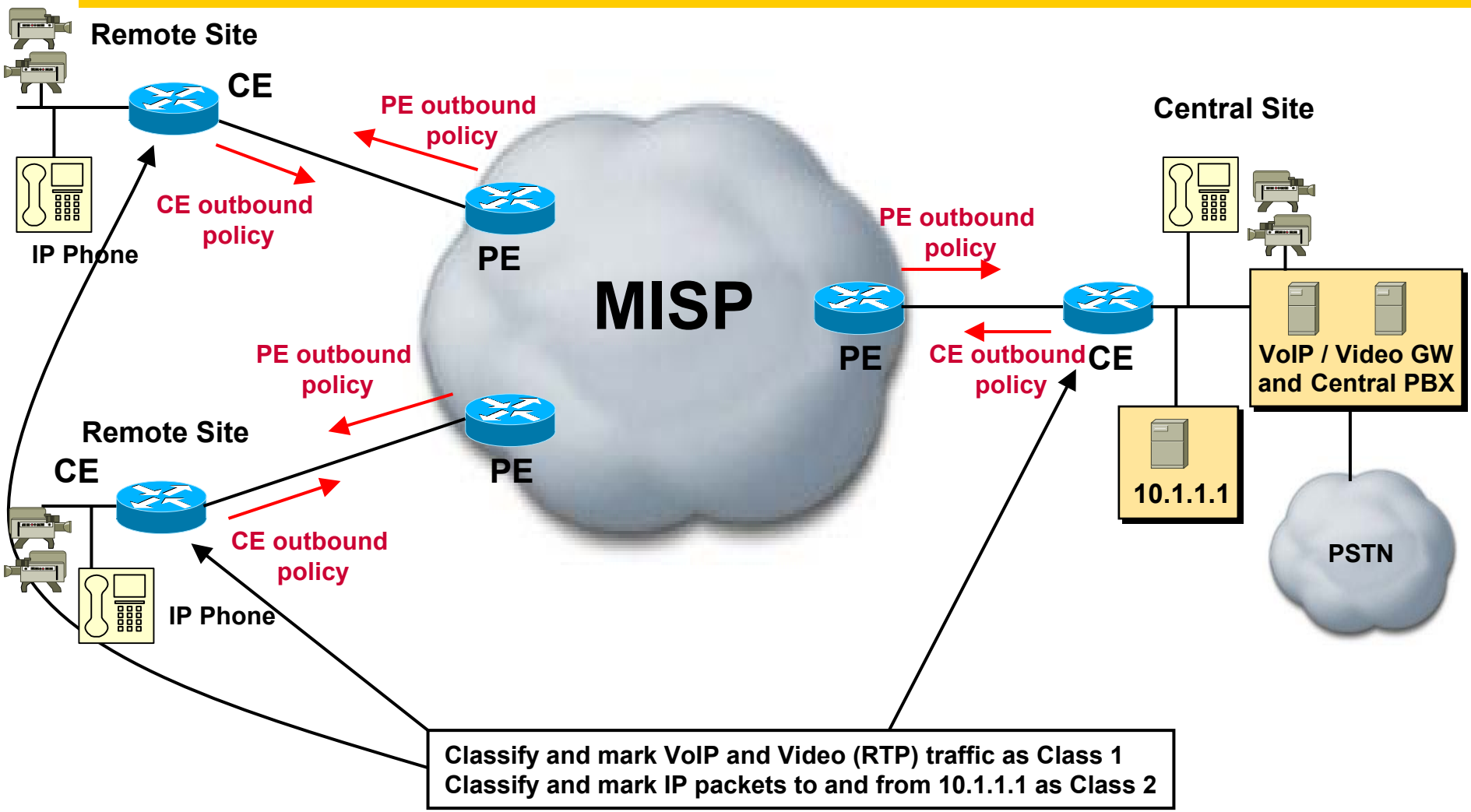
Case Study Example 2



Case Study
Example 3

- A customer has a central server with different important applications. He wants to make sure that all of the traffic to and from this server has has priority over other data traffic.
- In addition, he plans to roll out video conferencing and/or VoIP
- The perfect customer for profile A
  - Class 1 for Voice/Video
  - Class 2 for priority data
  - Class 3 for the rest

Case Study
Example 3



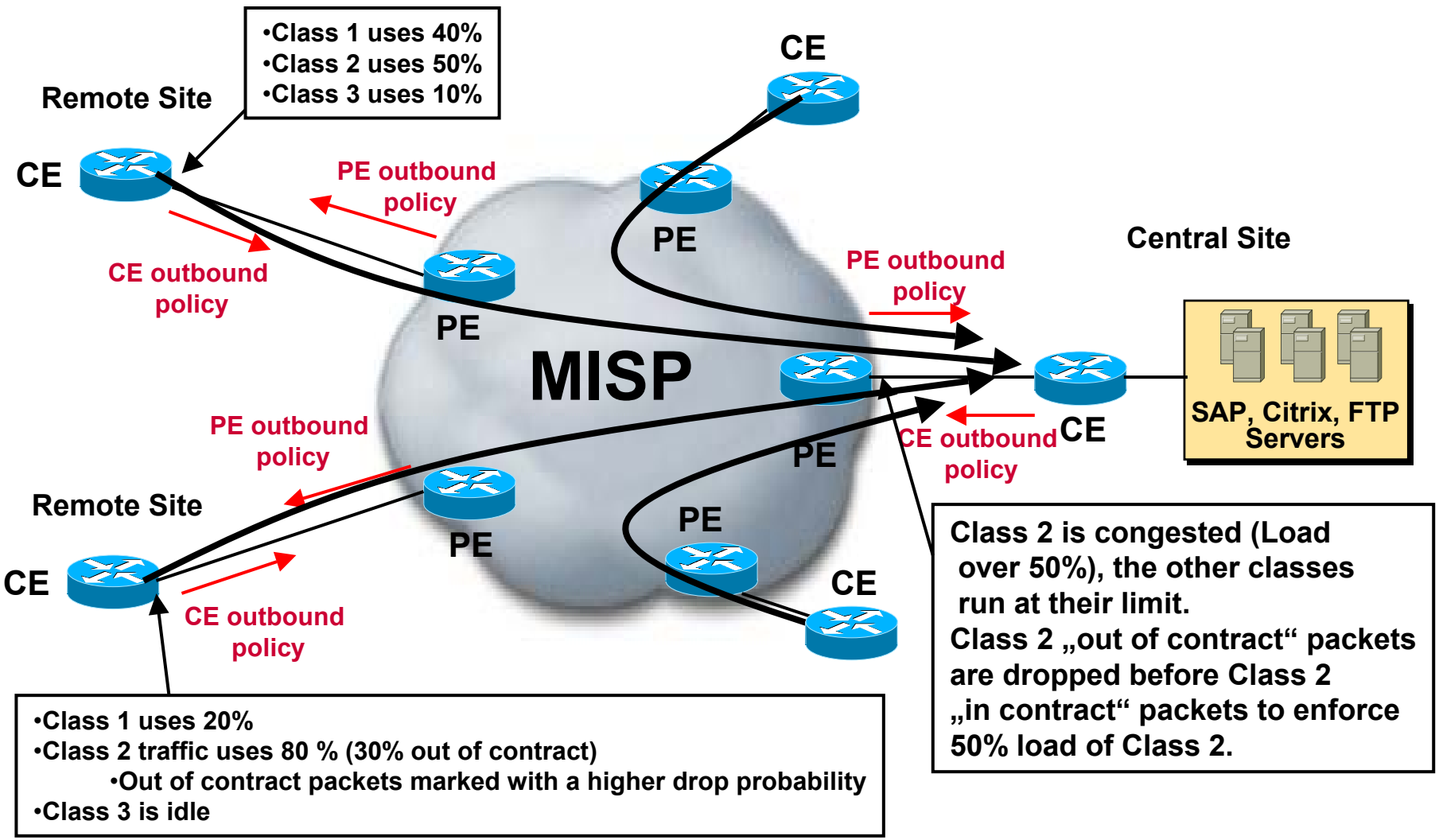
Class 2 “out of contract” handling

- Class 1 is fixed at 40% of the link bandwidth, excess traffic is dropped ➡ no burst capability
  - To ensure constant low delay and jitter
- Class 2 can burst above 50% (profile A) or 90 % (profile C) if the other classes do not fully use their allocated bandwidth
- Class 3 also can burst above it’s allocated bandwidth (10 % profiles A and C, 60% profile B) if room left on the link.

Class 2 “out of contract” handling

- Class 3 is always treated (marked) the same way (no matter if bursting above the allocated bandwidth or “in contract”)
  - Class 3 packets are always dropped first in case of congestion on a link
- Class 2 packets are treated differently if they are “in contract” or “out of contract”
  - “Out of contract” traffic is dropped before “in contract” traffic in case of congestion

Class 2
“out of contract” handling



Common errors

- Classification of customer traffic not working correctly
  - Wrong Class provisioning (classification of customer traffic)
- Overload of a Class (packets drops within a class)
  - Access Link bandwidth too small

Common errors

- **Check reporting reporting tool for Class throuput and packets drops**

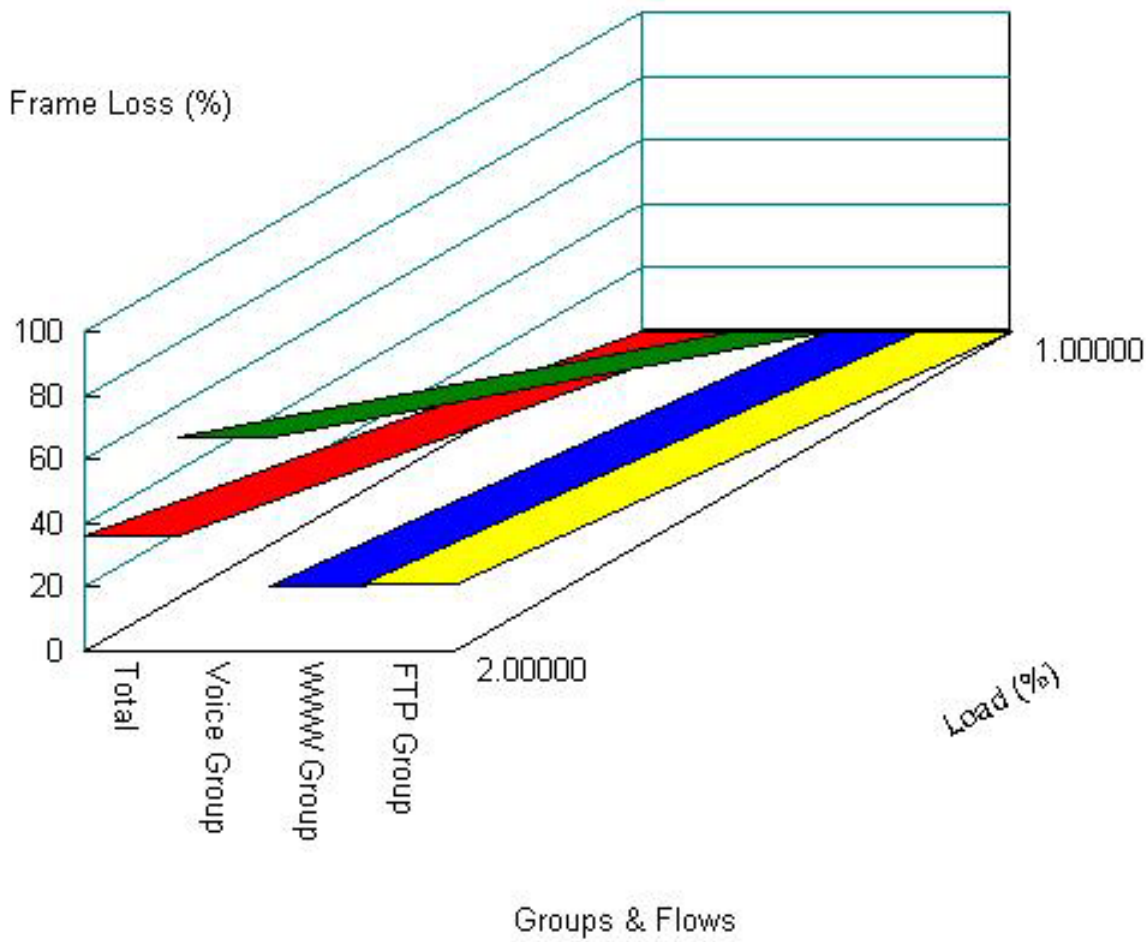
alternatively

- **The one and only router command which tells you all :**

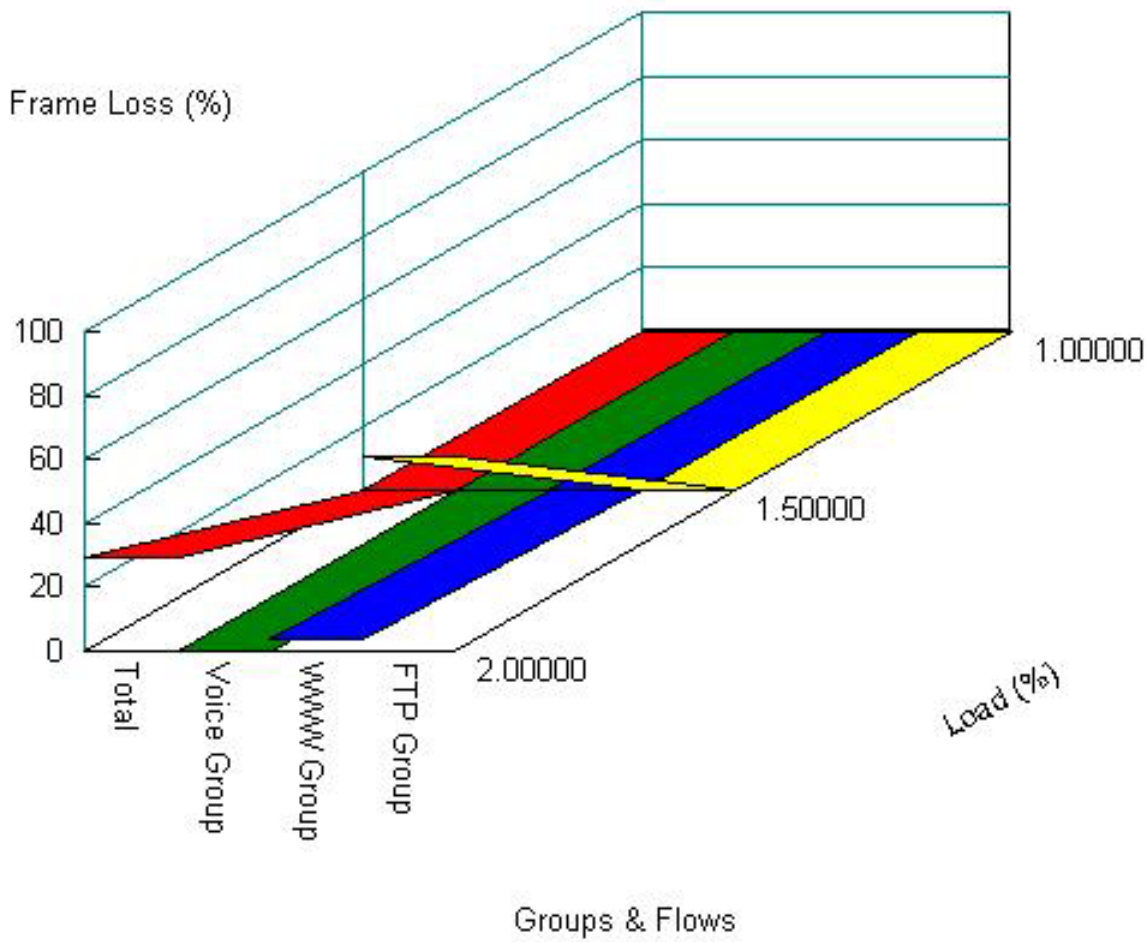
`show policy-map interface` *`interface`*

- **Information about**
  - **How many packets/bytes classified per class**
  - **How many packet/byte drops per class**
  - **and much much more…**
- **Remember to check CE and PE side !!!**

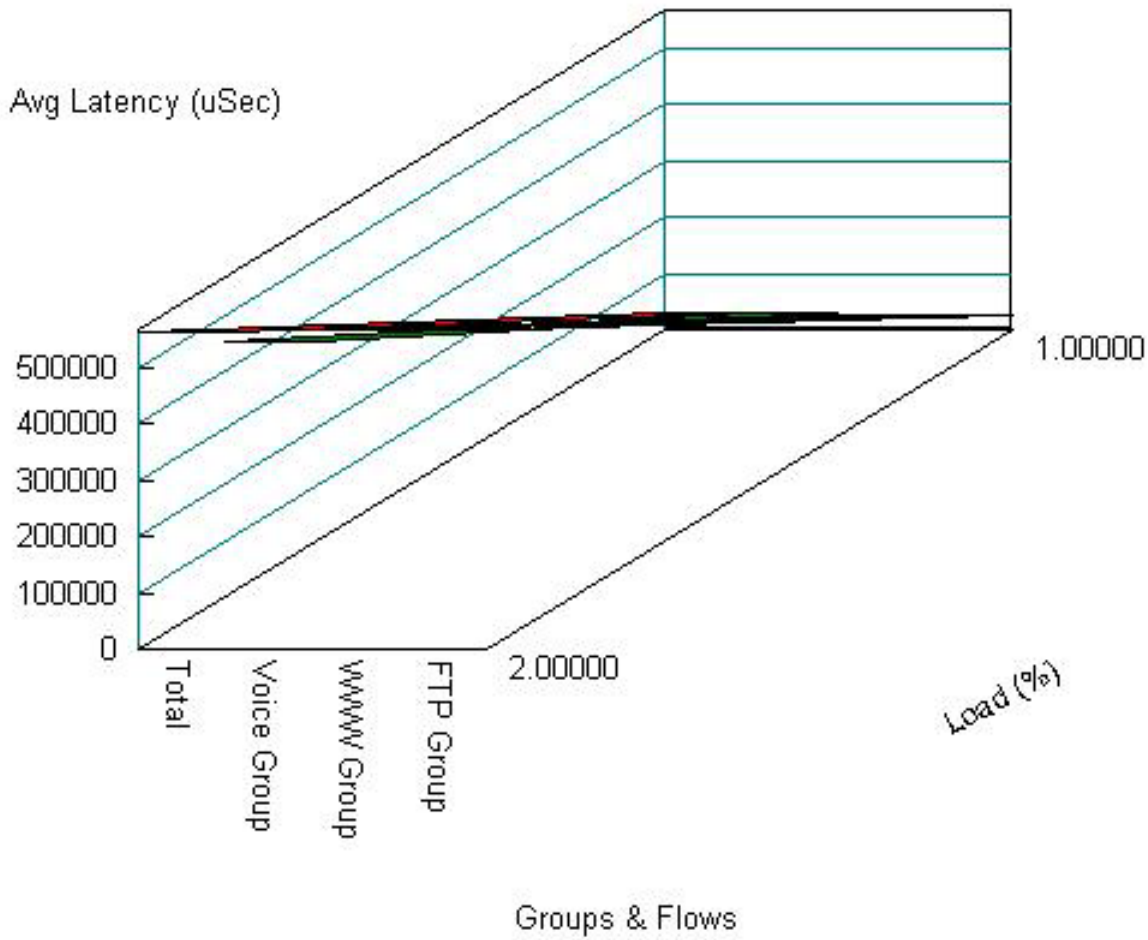
Packet Loss without CoS



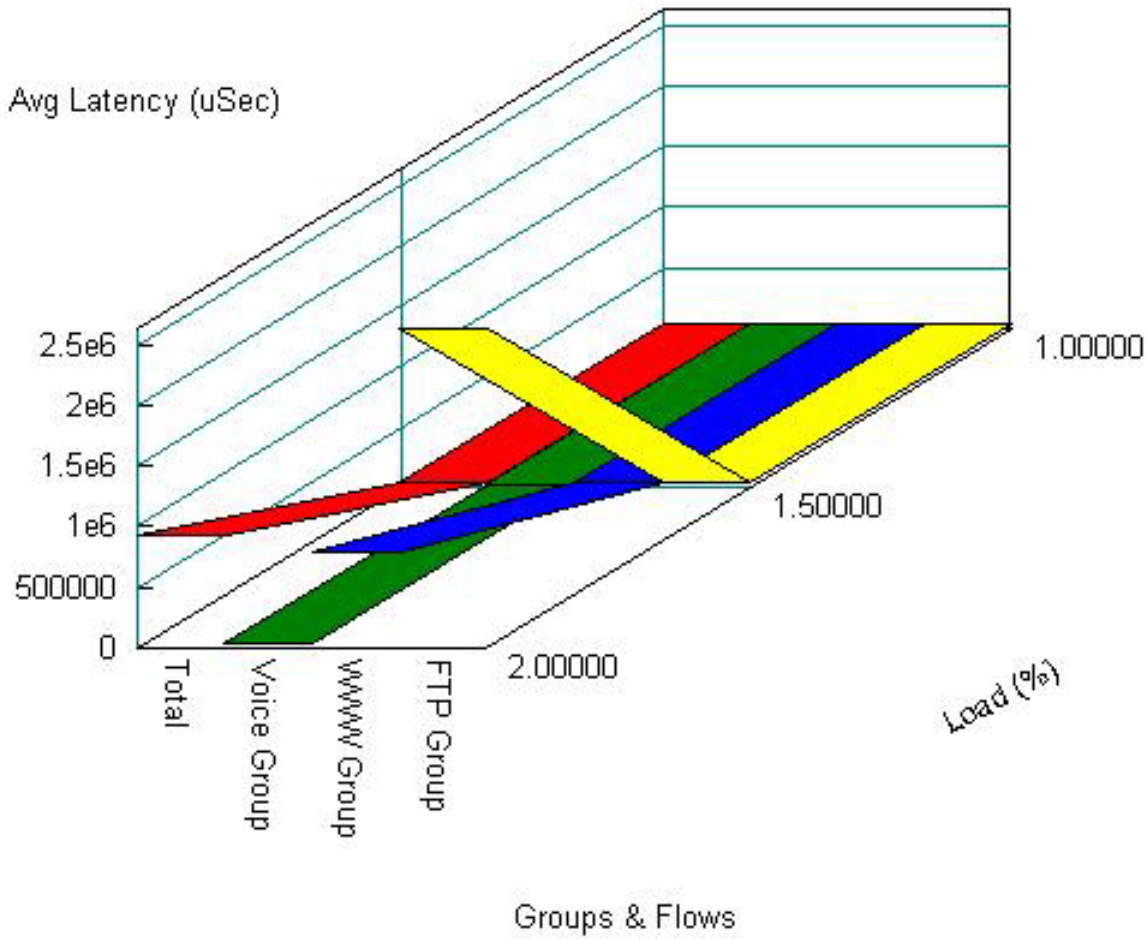
Packet Loss with CoS



Delay without CoS



Delay with CoS




Rapid
A fatal exception 0E has occurred at 0028:C0011E36 in VXD VMM(01) +
00010E36. The current application will be terminated.
* Press any key to terminate the current application.
* Press CTRL+ALT+DEL again to restart your computer. You will
lose any unsaved information in all applications.
Press any key to continue _

Thank you