



Introduction to Traffic Management and Quality of Service Technology



Agenda

- **Why Traffic Management Is Important?**
- **What Is QoS?**
- **How to Deploy QoS for Traffic Management?**
- **What Are Some of QoS Enabled Services?**



High Cost of Non-Responsiveness



Brokerage Operations = **\$6.45 Million**

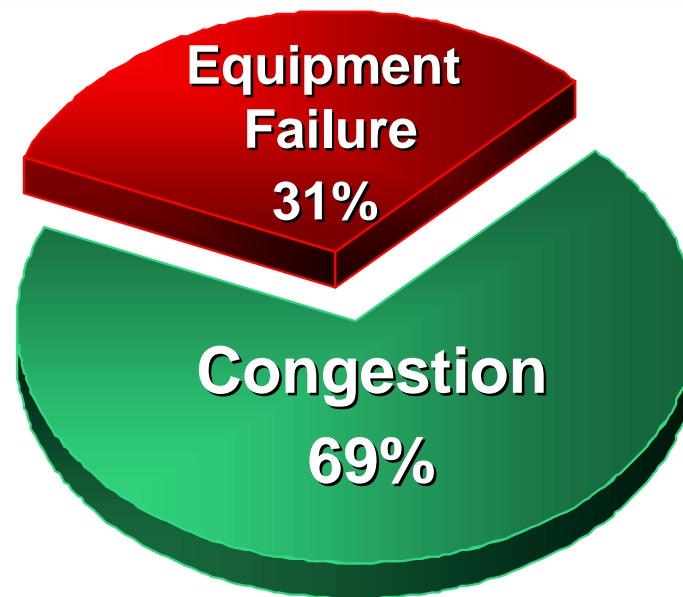
Credit Card Authorization = **\$2.6 Million**



Airline Reservations = **\$89,500**

The Cost of Congestion

**Costs of Productivity Loss
Due to Network Downtime**

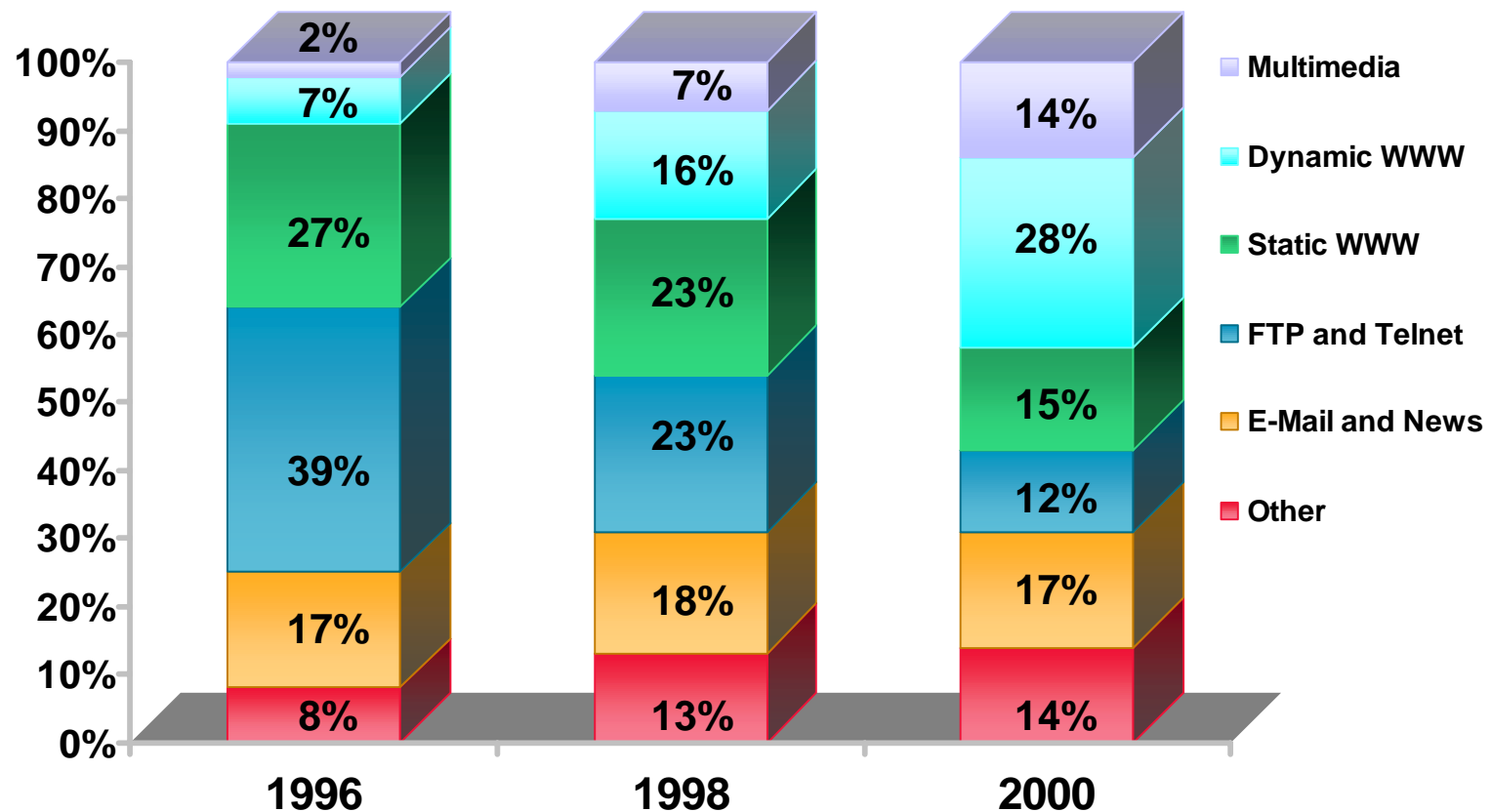


Congestion-related performance degradation has been found to cause the majority of network downtime costs

**Michael Howard
President, Infonetics Research**

©1997 Infonetics Research, Inc.,
Business-Centric Network Management and Downtime Costs 1997

Fundamental Shift Towards Bandwidth-Intensive Applications



Source: The Yankee Group, 1996

Not All Traffic Is Equal

| | Voice | FTP | ERP and Mission-Critical |
|-----------------------|-----------------|------------------|--------------------------|
| Bandwidth | Low to Moderate | Moderate to High | Low |
| Random Drop Sensitive | Low | High | Moderate To High |
| Delay Sensitive | High | Low | Low to Moderate |
| Jitter Sensitive | High | Low | Moderate |

Traffic Is Grouped into SLAs

Agenda

- **Why Traffic Management Is Important?**
- **What Is QoS?**
- **How to Deploy QoS for Traffic Management?**
- **What Are Some of QoS Enabled Services?**



What Is Quality of Service?

“

Collection of technologies which allows applications/users to request and receive **predictable service levels** in terms of data throughput capacity **(bandwidth)**, latency variations **(jitter)** and **delay**

”

Agenda

- **Why Traffic Management Is Important?**
- **What Is QoS?**
- **How to Deploy QoS for Traffic Management?**
- **What Are Some of QoS Enabled Services?**



QoS Models

- **Provisioned QoS**
Differentiated services
- **Signaled/dynamic QoS**
Integrated services(RSVP)

Step 1: Identify Traffic and it's Requirements

- **Network audit**
What is running and when?
- **Business audit**
How important is it for business?
- **Application audit**
What are it's requirements from network?
- **Service levels required**

Network Audit

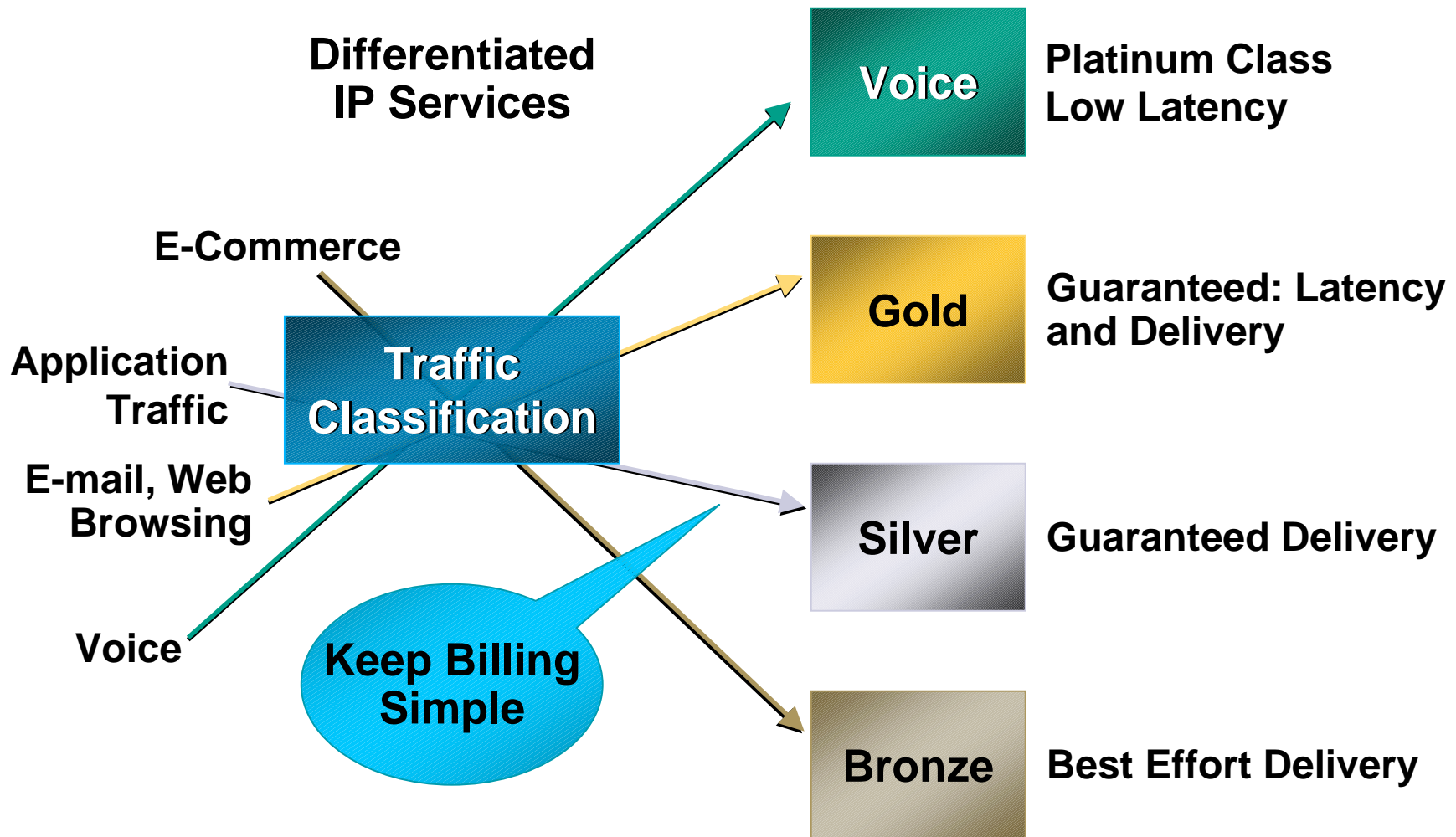
- **NetFlow**
Provides information on various traffic flows in the network
- **Protocol discovery**
Discovers what bandwidth intensive applications are running on the network
- **Sniffer**

How much bandwidth should I guarantee to my mission-critical applications?

Are there any non-mission-critical applications I should police?



Step 2: Divide the Traffic into Classes and Color It



What Is a Class?

- **Single user**
Mac address, IP address...
- **Department, customer**
Sub net, interface...
- **Application**
Port numbers, URL...

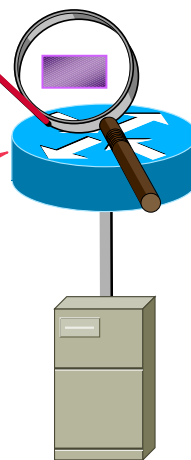
Network-Based Application Recognition



Link Utilization

| | |
|------------|------------|
| Citrix | 25% |
| Netshow | 15% |
| Oracle | 10% |
| FTP | 30% |
| HTTP | 20% |

- Protocol discovery analyzes application traffic patterns in real time
- NBAR classifies network traffic using application information
- Enables downstream actions based on QoS policies via random early detection class-based queuing, and policing
- New applications easily supported by loading Packet Description Language Modules

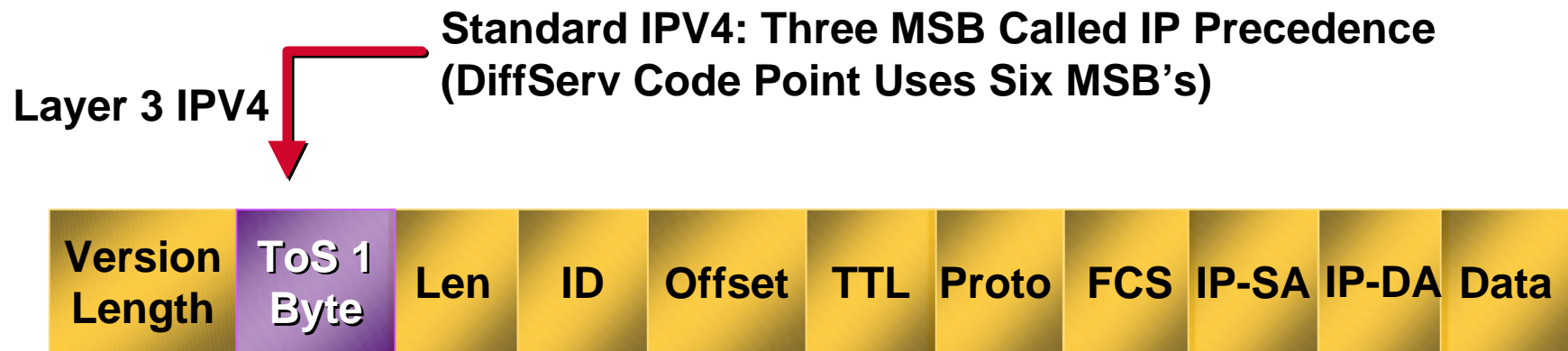


**Mark Citrix Real-Time as
GOLD Service and Police FTP**

Guarantee Bandwidth for Citrix

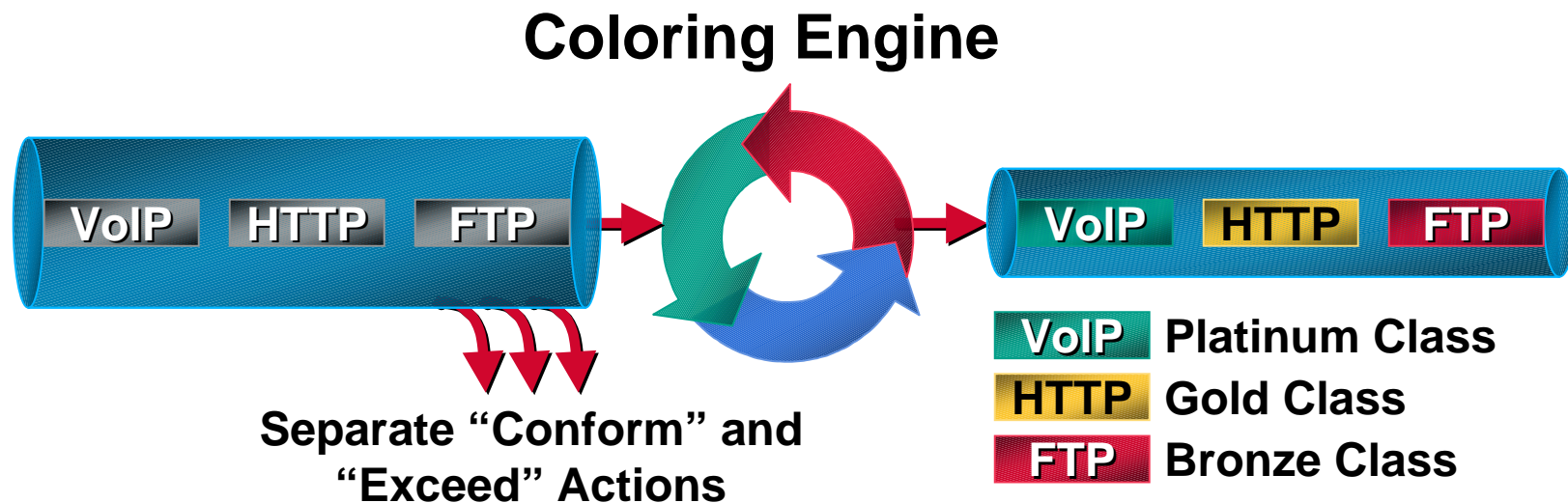
Cisco.com

What Is Coloring?



- Use this information to define QoS policies

Color the Packets



- **Color closer to the application**
- **Set the DSCP (Diffserv Code Point) at the edge of network**
- **Avoid host application-based coloring**

Step 3: Define Policies for the Classes

- **Minimum bandwidth guarantee**
This is the minimum guaranteed bandwidth to the class all the time
- **Give priority to the class**
Class is treated in a strict priority manner
- **Maximum bandwidth limits**
This is the maximum amount of bandwidth class will ever get
- **Congestion management**

Minimum Bandwidth Guarantee/ Priority for a Class

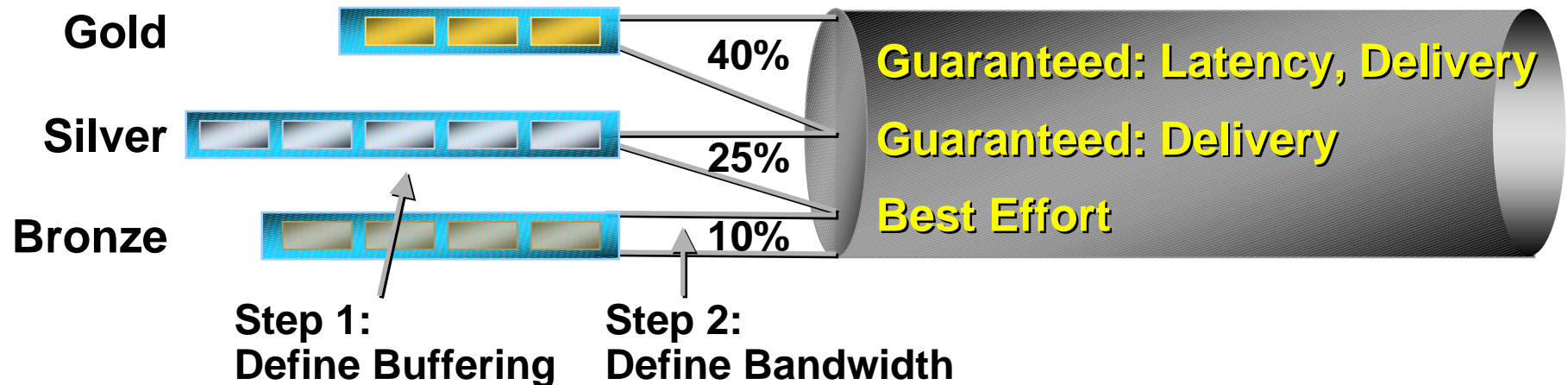
“

Policy required:

**Make sure my platinum class gets a
priority treatment and gold class gets
a minimum bandwidth guarantee**

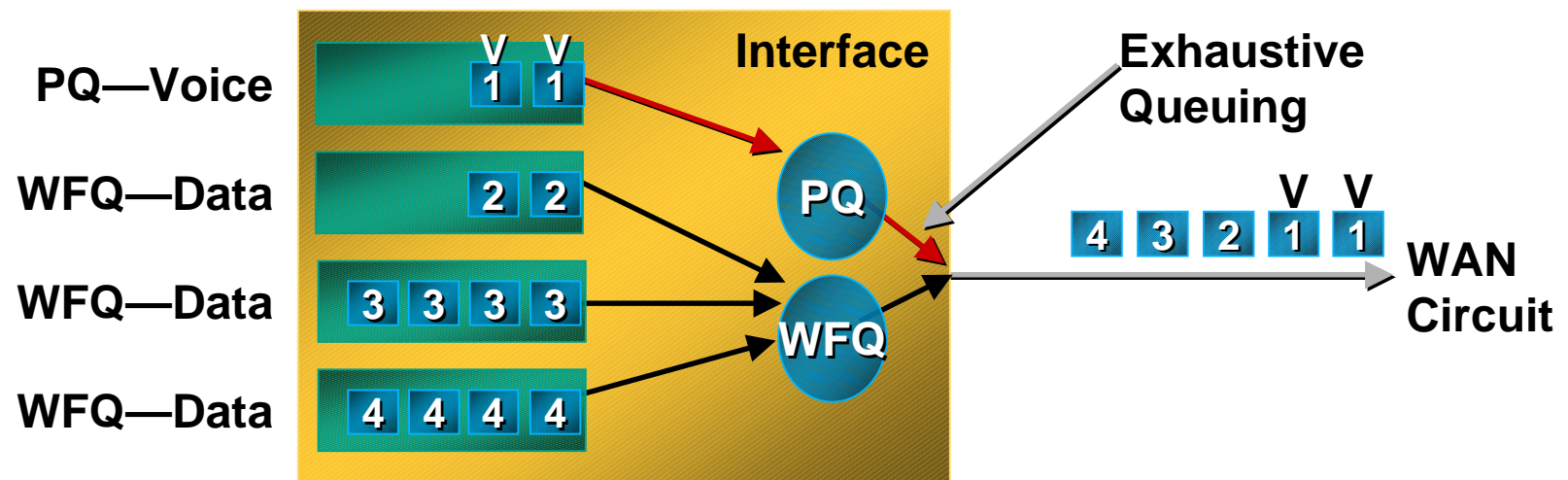
”

Scheduling

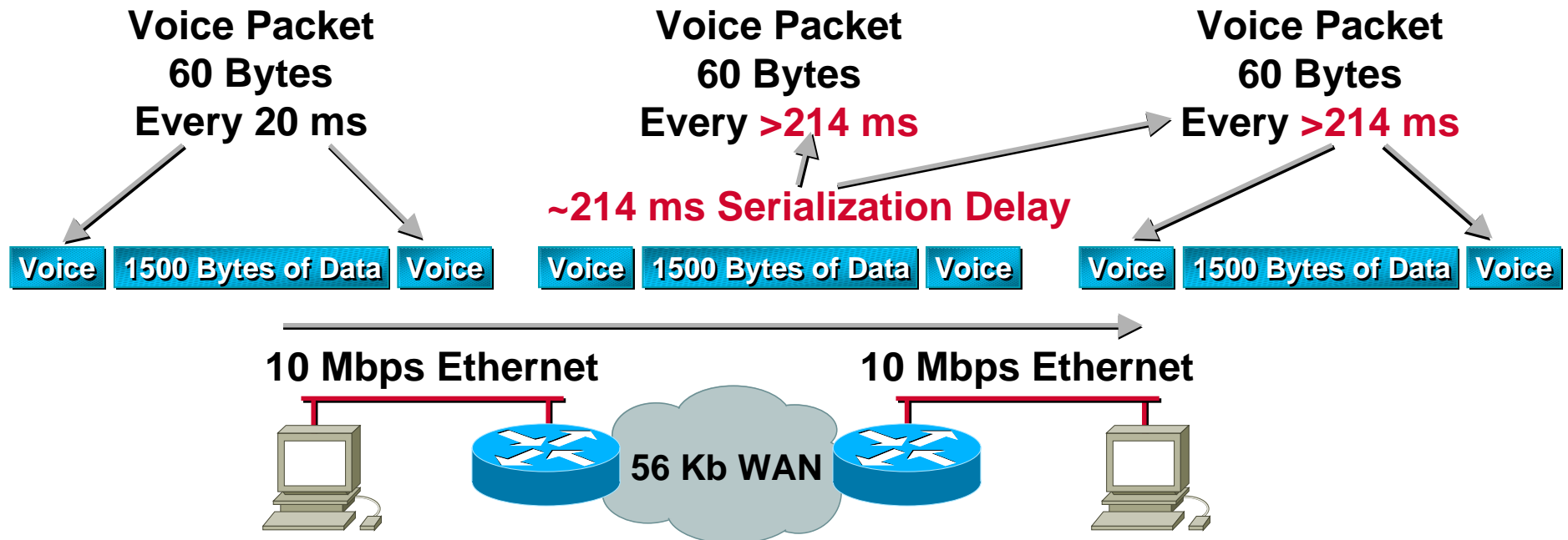


- Weights guarantee minimum bandwidth
- Buffering controls latency
- Unused capacity is shared amongst the other classes
- Each queue can be separately configured for QoS
- Benefits:
 - Maximize transport of paying traffic
 - No loss of service class guarantees
 - No wasted bandwidth as with PVCs

Low Latency Queuing



Large Packets “Freeze Out” Voice



- Large packets can cause playback buffer underrun, resulting in slight voice degradation
- Jitter or playback buffer can accommodate some delay/delay variation

Fragmentation Recommendations

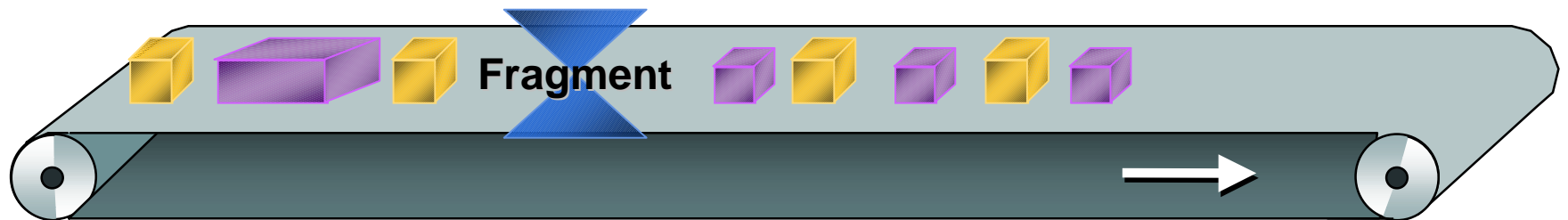
Assuming 10 ms Max Blocking Delay “Rules of Thumb”

**10 ms/Time for 1
Byte at BW =
Fragment Size**

| Link Speed | Frag Size |
|------------|------------|
| 56kbps | 70 Bytes |
| 64kbps | 80 Bytes |
| 128kbps | 160 Bytes |
| 256kbps | 320 Bytes |
| 512kbps | 640 Bytes |
| 768kbps | 1000 Bytes |
| 1536kbs | 2000 Bytes |

**Fragmentation Not Needed if
Max Frame Size Is 1500 Bytes**

Link Fragmentation and Interleave(LFI)



- **Fragment large packets and interleave with voice packets over WAN links**
- **Reassemble at other end of link**
- **Reduces voice delay and jitter**

RTP Header Compression

- Header is 2x size of voice data (40 vs 20 bytes)
- RTP Header Compression(CRTP) reduces header to 2–4 bytes
- Used hop-by-hop on slow links



Maximum Rate Limiting

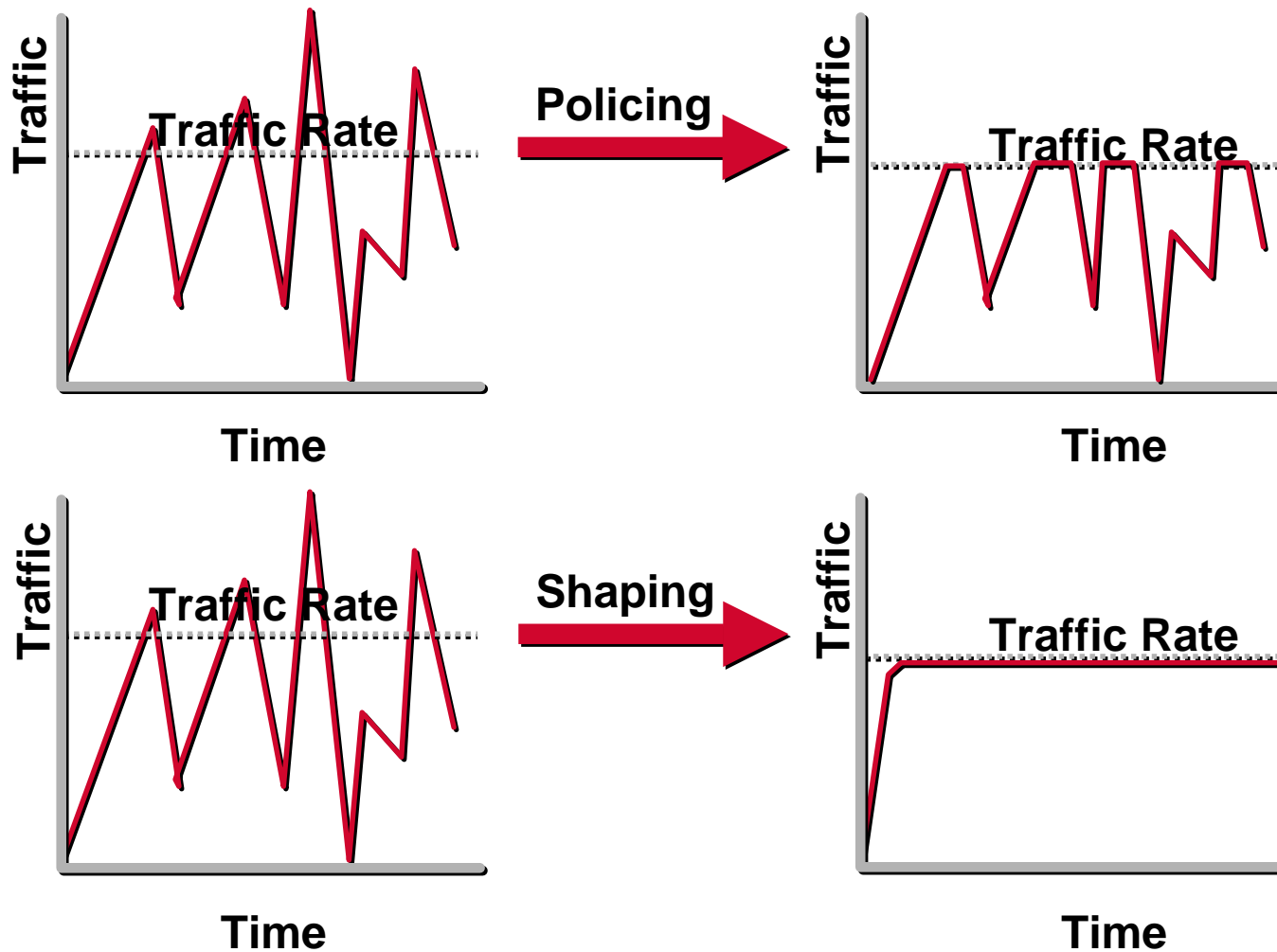
“

Policy required:

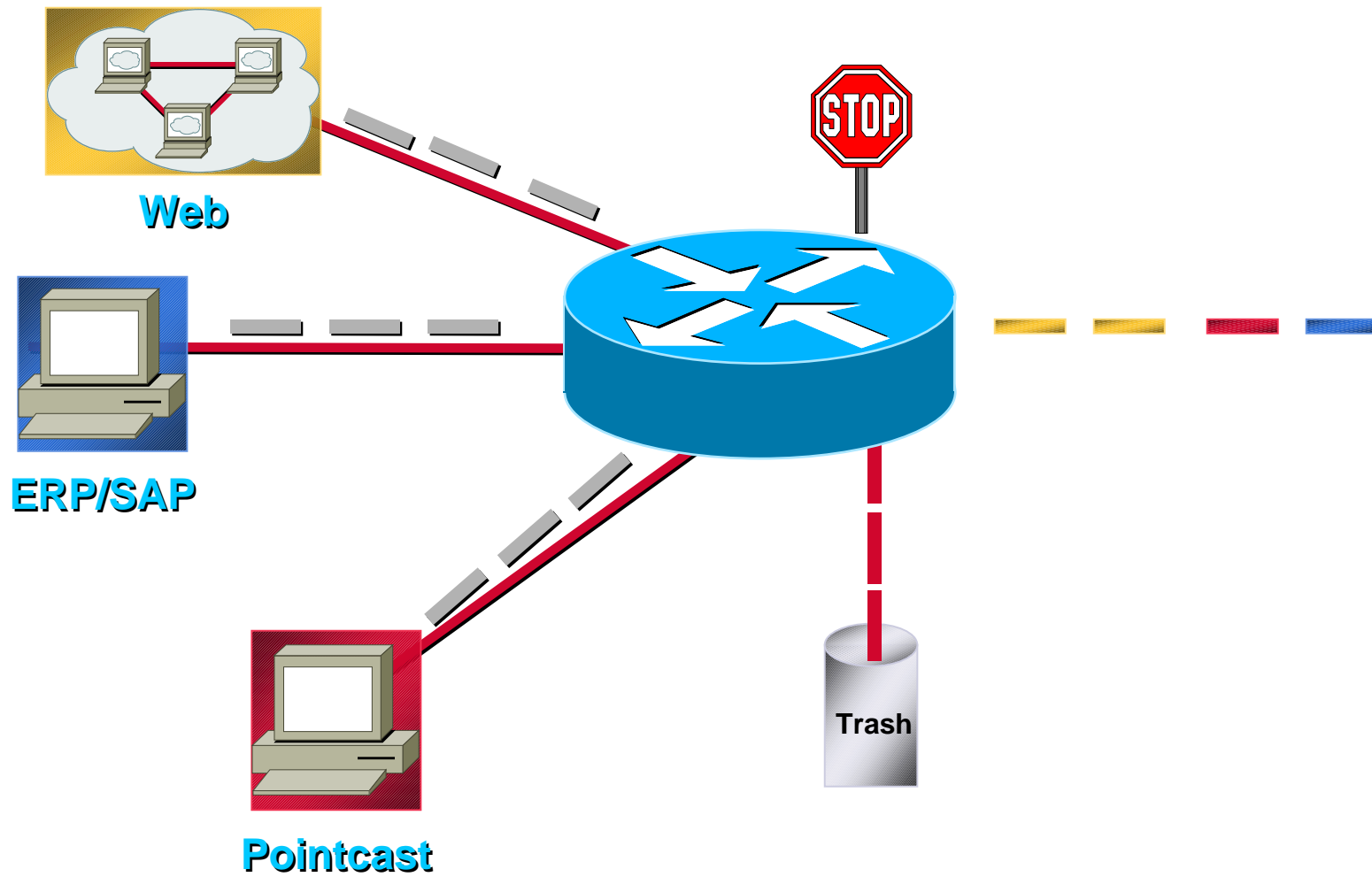
**Make sure my bronze traffic does
not get more than x kbps of
bandwidth at any time**

”

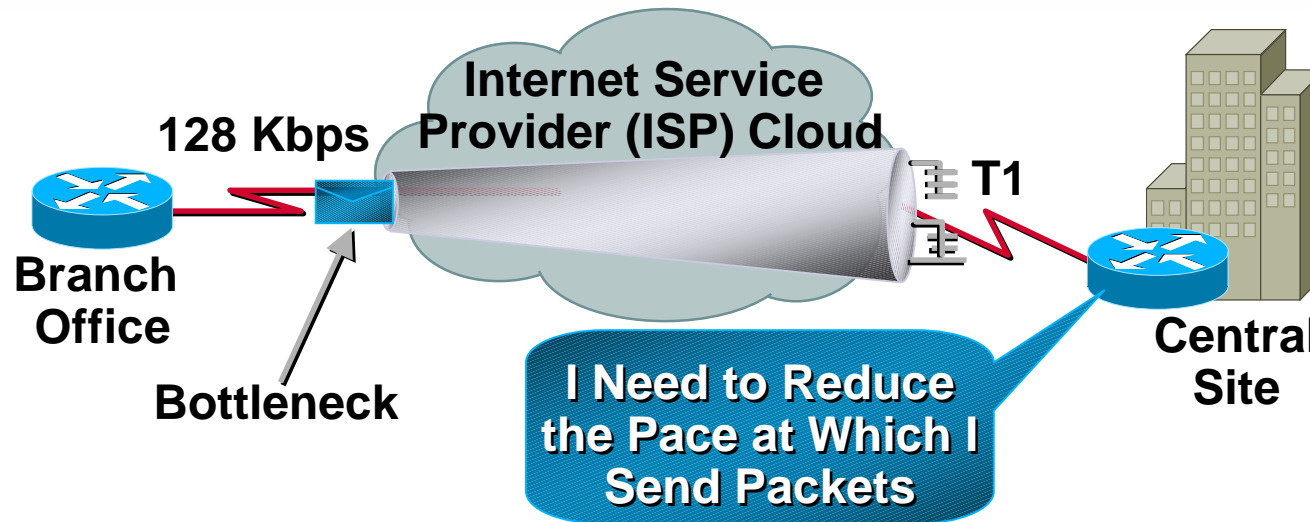
Traffic Policing vs. Shaping



Policer



Shaper



- Reduces outbound traffic flow to avoid congestion(via buffering)
- Eliminates bottlenecks in topologies with data rate mismatch
- Provides mechanism to partition interfaces to match far-end requirements

Congestion Avoidance

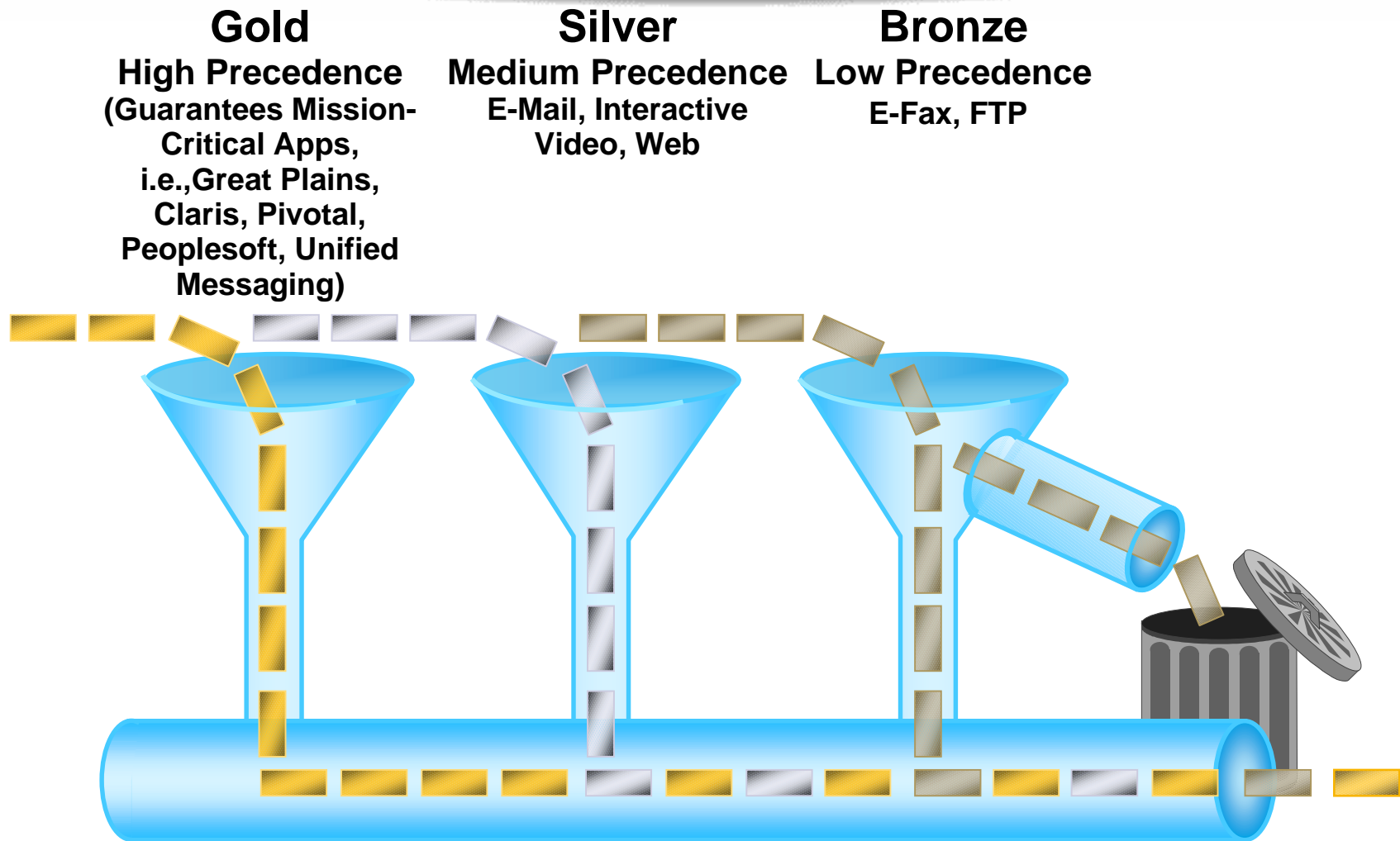
“

Policy required:

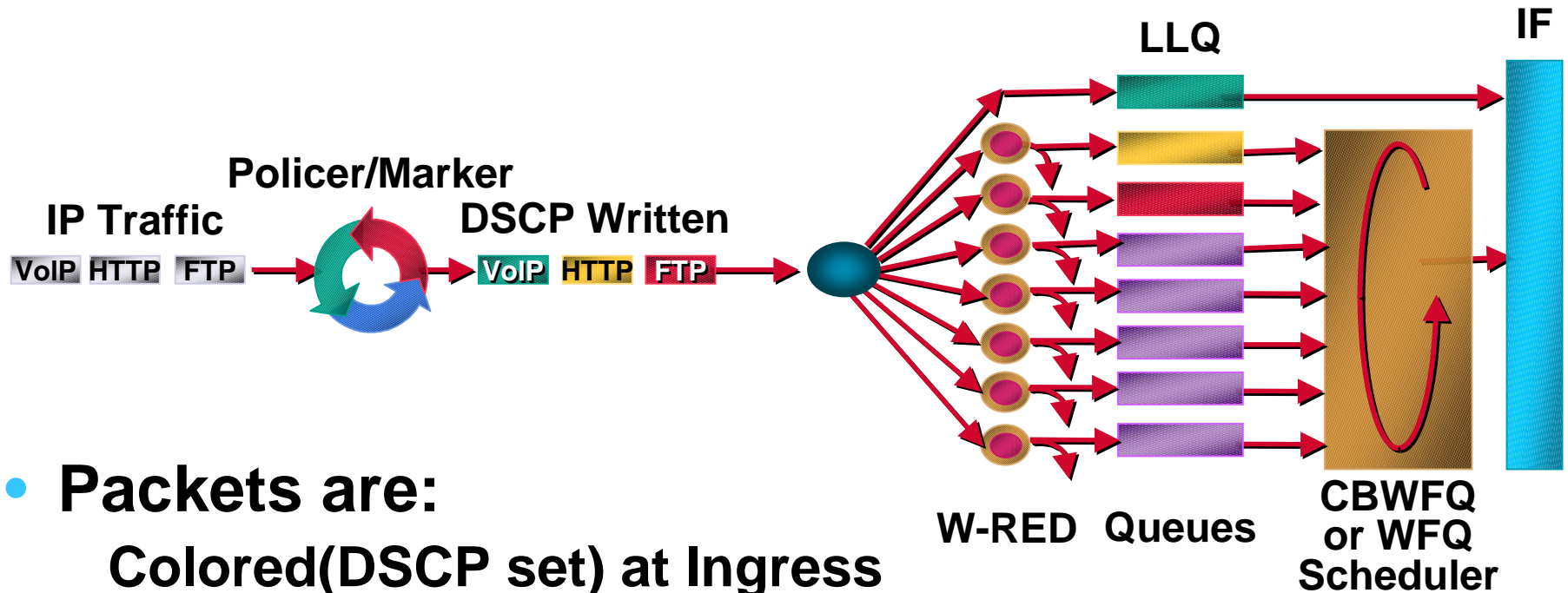
**Make sure my bronze or silver traffic
gets dropped when there is
congestion and not gold traffic**

”

Weighted Random Early Detection

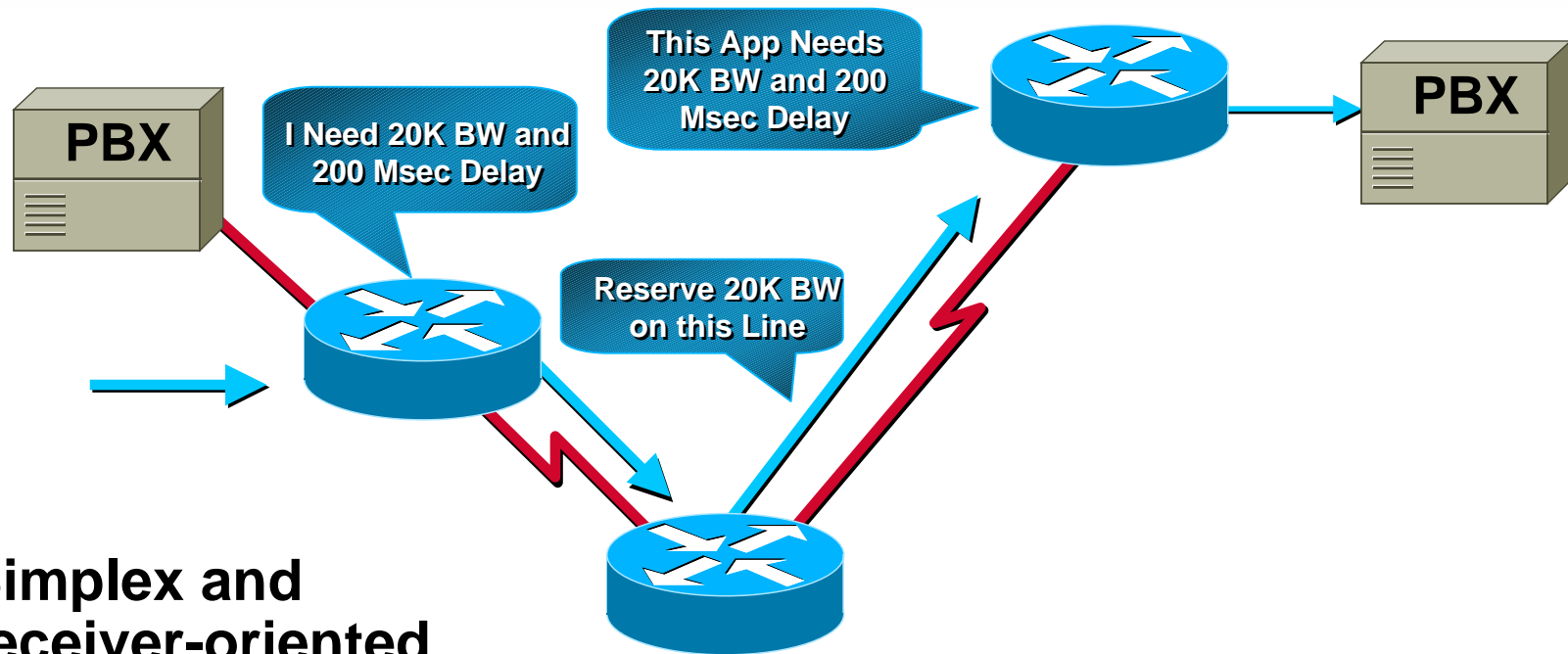


Putting it All Together



- **Packets are:**
 - Colored(DSCP set) at Ingress**
 - Classified and potentially discarded by W-RED (Congestion Management)**
 - Assigned to the appropriate outgoing queue**
 - Scheduled for transmission by CBWFQ**

Integrated Services



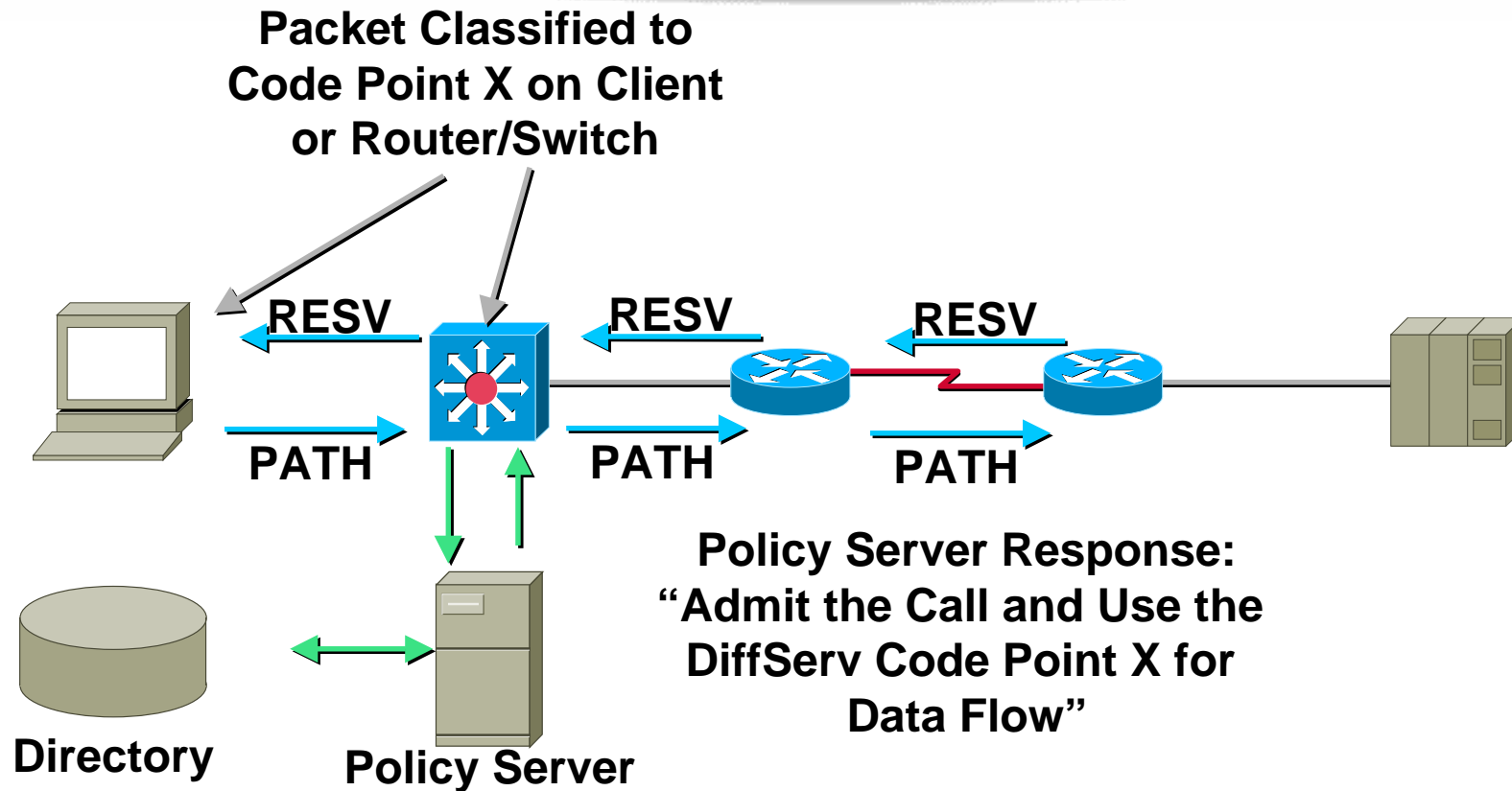
- Simplex and receiver-oriented
- RSVP QoS services

Guaranteed service—mathematically provable bounds on end-to-end datagram queuing delay/bandwidth

Controlled service—approximate QoS from an unloaded network for delay/bandwidth

- RSVP provides the policy to WFQ

Policy



**RSVP(Quantative) Is Used for the Control Path Flow;
Data Path Uses an Aggregate as Identified by the DSCP;
RSVP Is Used to Signal the Data Path Aggregate**

Complete QoS Management

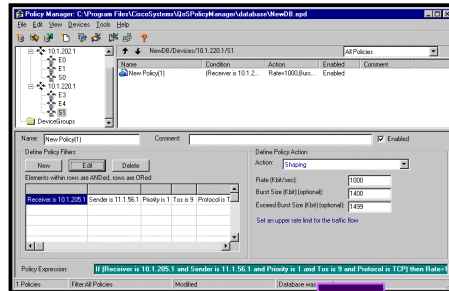
CONFIGURE

TRENDING

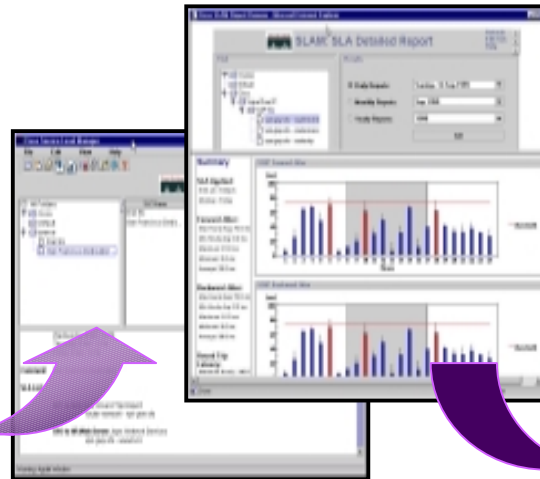
MONITORING

Network Wide

Device



QPM



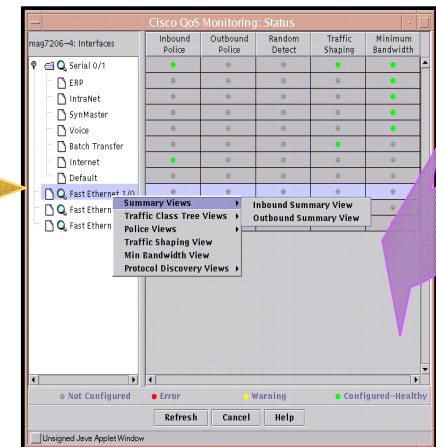
SLAM



IPM



QDM



QDM

Cisco.com

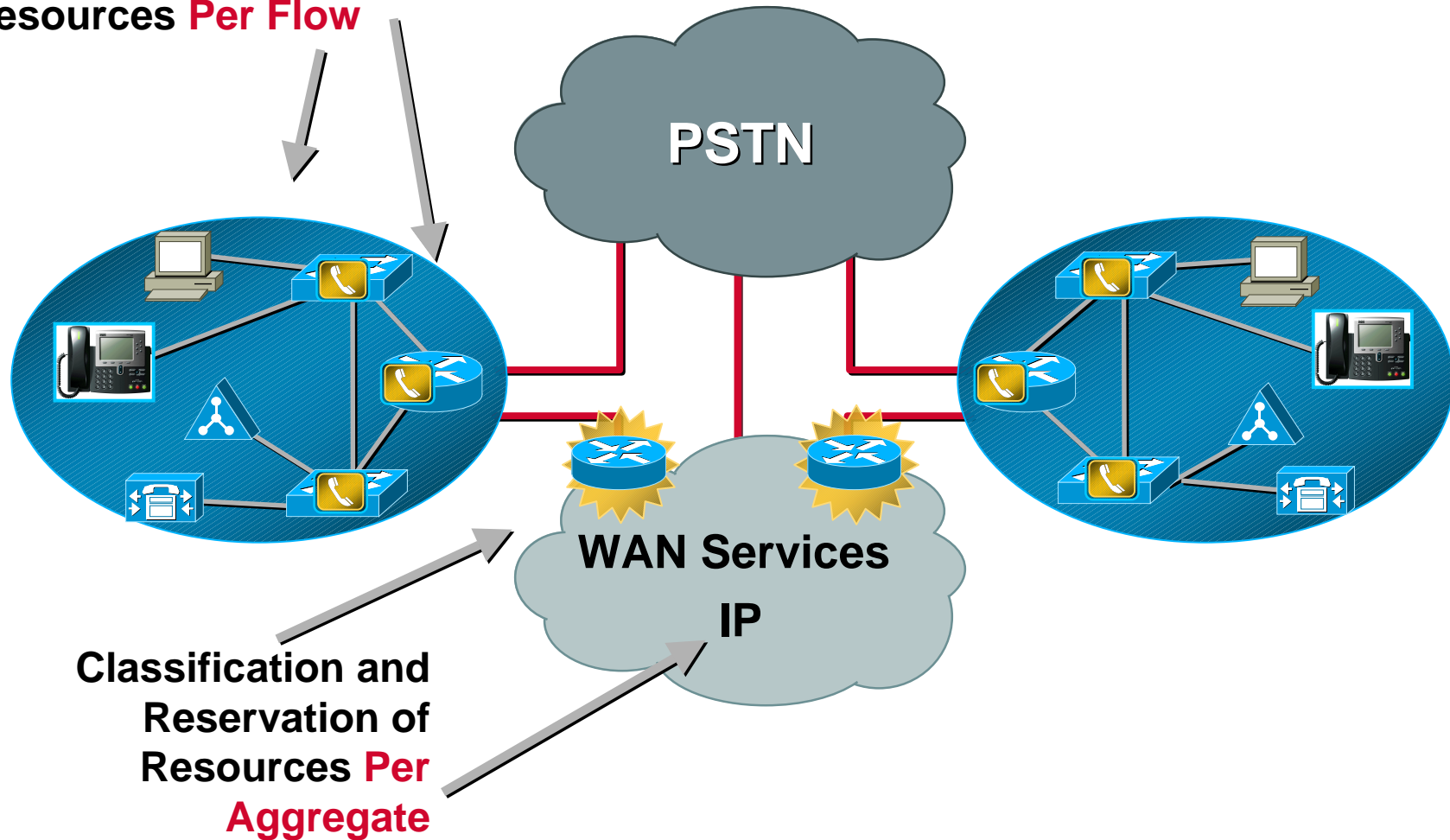
Agenda

- **Why Traffic Management Is Important?**
- **What Is QoS?**
- **How to Deploy QoS for Traffic Management?**
- **What Are Some of QoS Enabled Services?**

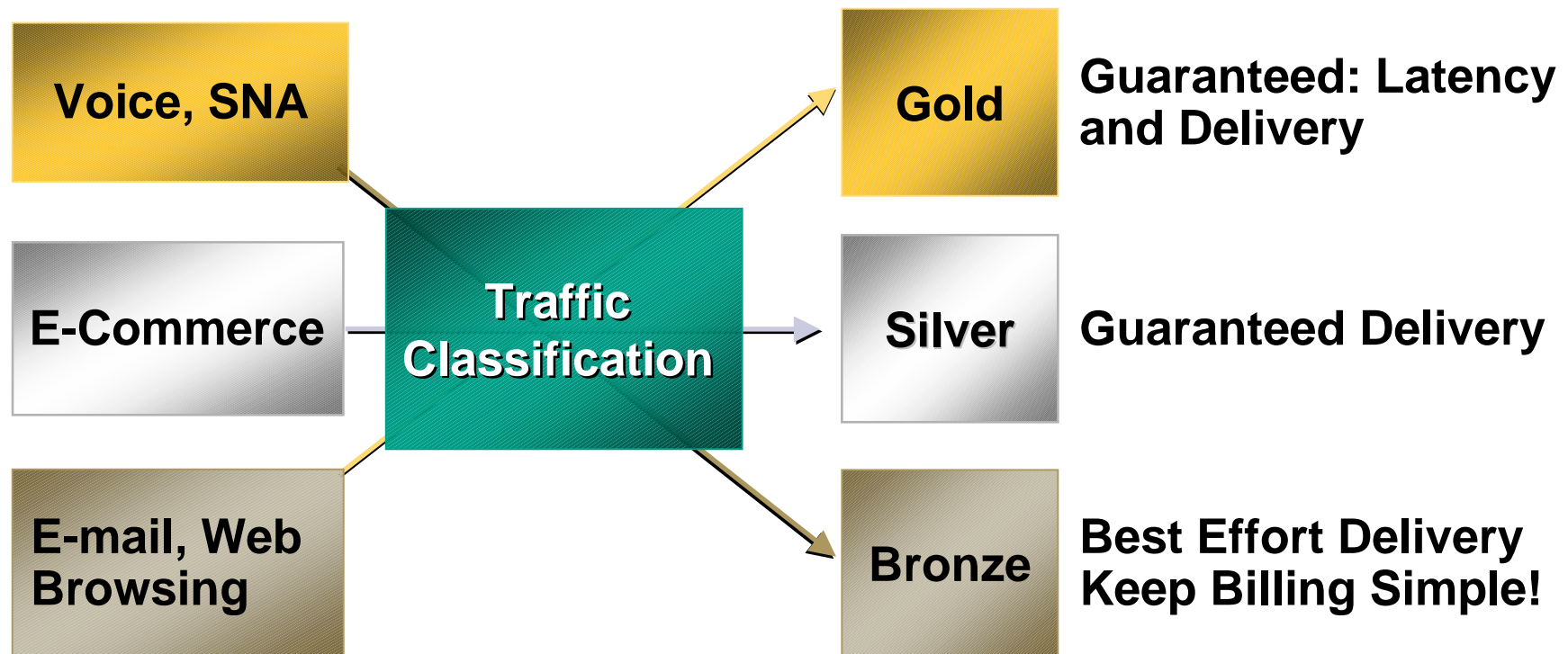


Example Application: QoS for Voice Over IP

Classification and
Reservation of
Resources **Per Flow**



Example Application: Premium IP



- **Deliver IP-Based Differentiated Services with Service Level Agreements(SLAs)**

Summary

“

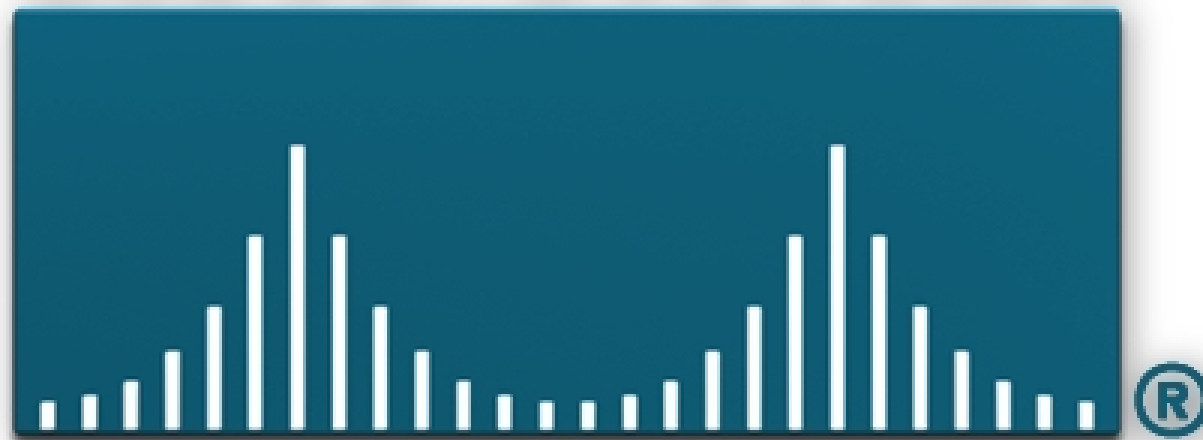
QoS Is Managed Unfairness

”



Introduction to Traffic Management and Quality of Service Technology

CISCO SYSTEMS



EMPOWERING THE INTERNET GENERATIONSM