



Calidad de Servicio (QoS)

Agosto'2002

Agenda

Cisco.com

- **Qué es ?**
- **Porqué importa ?**
- **Cómo se define ?**
- **Donde y como se aplica**
- **Servicios a los que afecta**
- **End to End QoS y Policy Management**
- **Un ejemplo**
- **P&R**

Qué es ?

Cisco.com

- **Respuesta práctica:**
 - **“(Des)igualdad administrada”**

Qué es ?

Cisco.com

- **Respuesta formal**
 - **Conjunto de técnicas para controlar:**
 - El retardo y su variación (jitter)
 - El uso del ancho de banda
 - El descarte de paquetes
 - **para los flujos de información en una red.**

Agenda

Cisco.com

- Qué es ?
- **Porqué importa ?**
- Cómo se define ?
- Donde y como se aplica
- Servicios a los que afecta
- End to End QoS y Policy Management
- Un ejemplo
- P&R

Porqué Importa ?

Cisco.com

Antecedente

- **La tendencia a unificar, en una misma infraestructura de red, todos los servicios de comunicación hace indispensable el manejo de Calidad de Servicio.**

Porque importa ?

Cisco.com

- **Una implementación real de QoS permite la utilización controlada y administrada de la infraestructura de red, otorgando:**
 - **Diferenciación de los recursos de red utilizados por diversos servicios de comunicación (tipos de flujo)**
 - **Control sobre el “nivel de servicio” global de la red**



CustQoS200205

© 2002, Cisco Systems, Inc. All rights reserved.



Porqué importa ?

Cisco.com

- **El ancho de banda, por sí mismo, no es suficiente**
 - **El ancho de banda es sólo uno de los atributos de un canal de comunicación. Que pasa con:**
 - El retardo ?
 - La tasa de errores ?
 - El costo ?

Porqué importa ?

Cisco.com

- El ancho de banda, por sí mismo, no es suficiente
 - **Respecto del costo del ancho de banda como solución a la congestión**
 - Debo caracterizar con más precisión la situación
 - ◆ Congestión permanente v/s transitoria
 - ◆ Proporción de tráfico relevante
 - ◆ Posibilidad de postergar alguna porción del tráfico

Porqué importa ?

Cisco.com

- El ancho de banda, por sí mismo, no es suficiente

Porción Postergable del tráfico

Porción relevante del tráfico	Duración de la Congestión	Porción Postergable del tráfico	
		Menor	Mayor
Mayor	Permanente	+ Ancho de banda	+ Ancho de banda Modulación
	Transitoria	+ Ancho de banda Modulación	Modulación
Menor	Permanente	Modulación	Modulación
	Transitoria	Modulación	Modulación

Porqué importa ?

Cisco.com

- El ancho de banda, por sí mismo, no es suficiente
 - Entonces, si simplemente se invierte en más ancho de banda puro, lo más probable es que se lo malgaste:
 - Resolviendo un problema transitorio con una herramienta permanente
 - Subutilizando la inversión
 - Generando una solución frágil y efímera, contrapuesta al su propio objetivo: cursar más tráfico.
 - Por esto se requieren técnicas de clasificación, priorización y despacho elaboradas !

Agenda

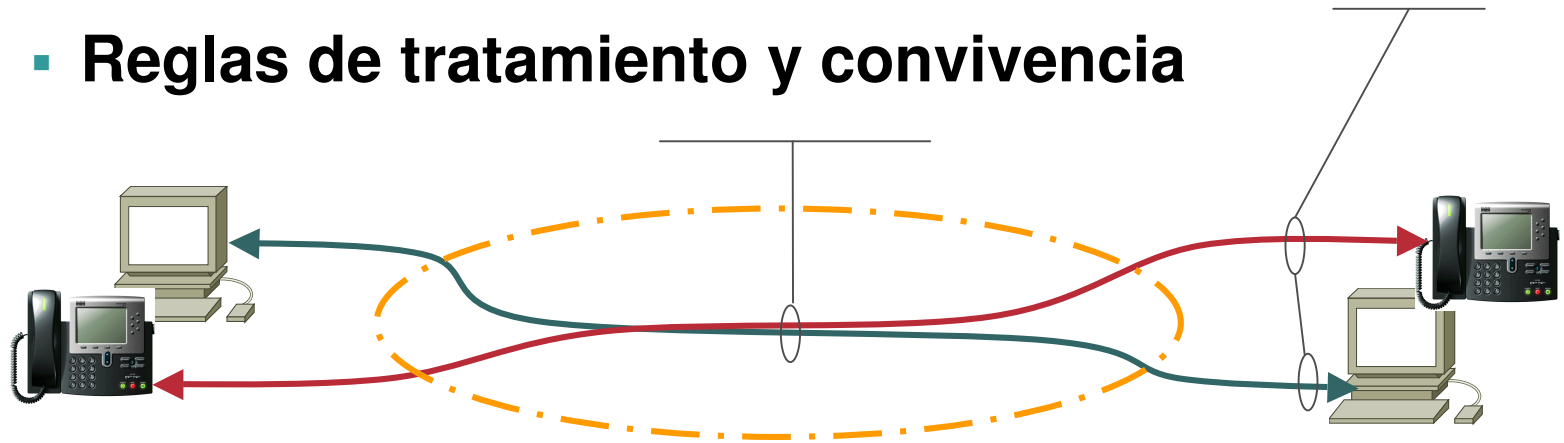
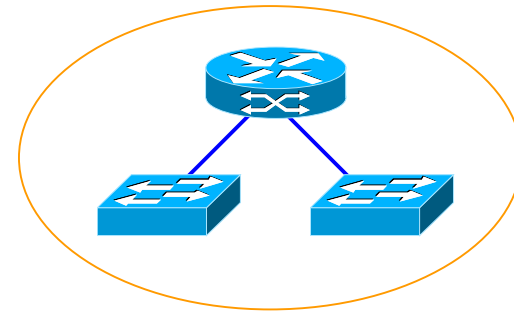
Cisco.com

- Qué es ?
- Porqué importa ?
- **Cómo se define ?**
- Donde y como se aplica
- Servicios a los que afecta
- End to End QoS y Policy Management
- Un ejemplo
- P&R

Cómo se define ?

- **Elementos de QoS**

- Recursos de red
- Flujos de información y sus requerimientos
- Reglas de tratamiento y convivencia



Cómo se define ?

Cisco.com

- **Recursos y atributos**
 - **De los enlaces**
 - Ancho de Banda y retardo de serialización
 - Retardo de transmisión
 - **De los gateways (Routers y Switches)**
 - Retardo de procesamiento (forwarding)
 - Capacidad de almacenamiento de tránsito (buffer)

Cómo se define ?

Cisco.com

- **Flujos y sus requerimientos**
 - **Requerimientos según el servicio al que obedecen**
 - Ancho de banda
 - Variación del ancho de banda
 - Retardo total
 - Variación del retardo
 - Sensibilidad al descarte o pérdida de información
 - **Identificación del flujo**
 - Identificación explícita (ATM)
 - Mediante otros atributos (Origen-Destino, ...)
 - Marca en L2 y/o L3 (agregaciones)

Cómo se define ?

Tipificación de los requerimientos según el flujo

	Voz	FTP	ERP y Mission-Critical
Ancho de banda	Moderado a bajo	Moderado a alto	Baja
Sensibilidad al descarte aleatorio	Baja	Alta	Moderada a alta
Sensibilidad al retardo	Alta	Baja	Moderada a alta
Sensibilidad al jitter	Alta	Baja	Moderada

Cómo se define ?

Cisco.com

- **Reglas**

- **De admisión**

- Como determinar si es o no posible transportar un flujo específico (dado sus requerimientos) en un instante determinado sobre la red

- **De “convivencia”**

- Como resolver la competencia entre flujos por los recursos de red
 - Priorización, disciplinas de encolamiento y/o servicio

- **Manejo centralizado de las reglas**

- Escalabilidad, administrabilidad y homogeneidad

Cómo se define ? - Modelo OSI

Cisco.com

- **Un ingrediente adicional que regula la relación entre estos 3 elementos es el modelo de capas OSI**
 - Los enlaces corresponden a elementos de capa 1 y 2
 - Los Switches; capa 2 (y más)
 - Los Routers; capa 3 (y más)
 - Los flujos involucran a todas las capas

Cómo se define ? - Modelo OSI

Cisco.com

- **Y ?**
 - **Que como la calidad de servicio de un flujo involucra a todas las capas, se requiere intercambiar información de control entre ellas para asignar los recursos que cada una maneja.**
 - **Este fue el primer enfoque que se siguió y resultó demasiado costoso e ineficiente.**

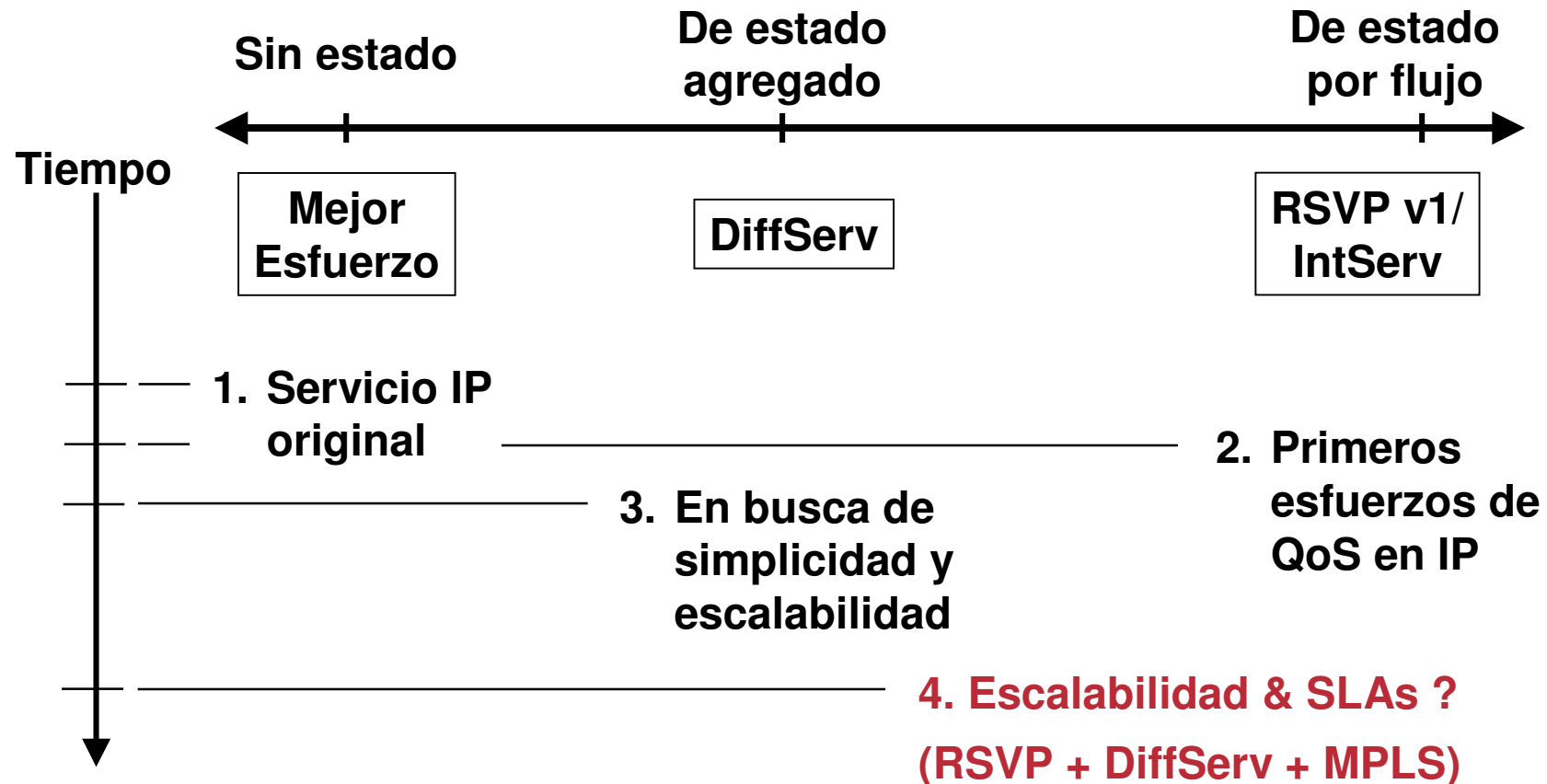
Cómo se define ?

Cisco.com

- **Modelos de Servicio de red**
 - De “mejor esfuerzo”
 - Modelo Integrado: IntServ (RSVP)
 - Modelo diferenciado: DiffServ (DSCP)

Cómo se define ? - El espectro de modelos de QoS IP

Cisco.com



Cómo se define ? - Solución intermedia

Cisco.com

- **El modelo de servicio diferenciado: DiffServ**
 - Se centra en IP (capa 3), ya que es lo mínimo para llegar desde el origen hasta el destino del flujo (End to End) al igual que sus antecesores (IntServ)
 - Se ofrecen *categorías* de servicio de transporte identificadas por “*Puntos de servicio*” (DSCP)
 - Los flujos se asocian a estas categorías de manera *agregada* (múltiples flujos por categoría)

Cómo se define ? - DiffServ

- **Pros**

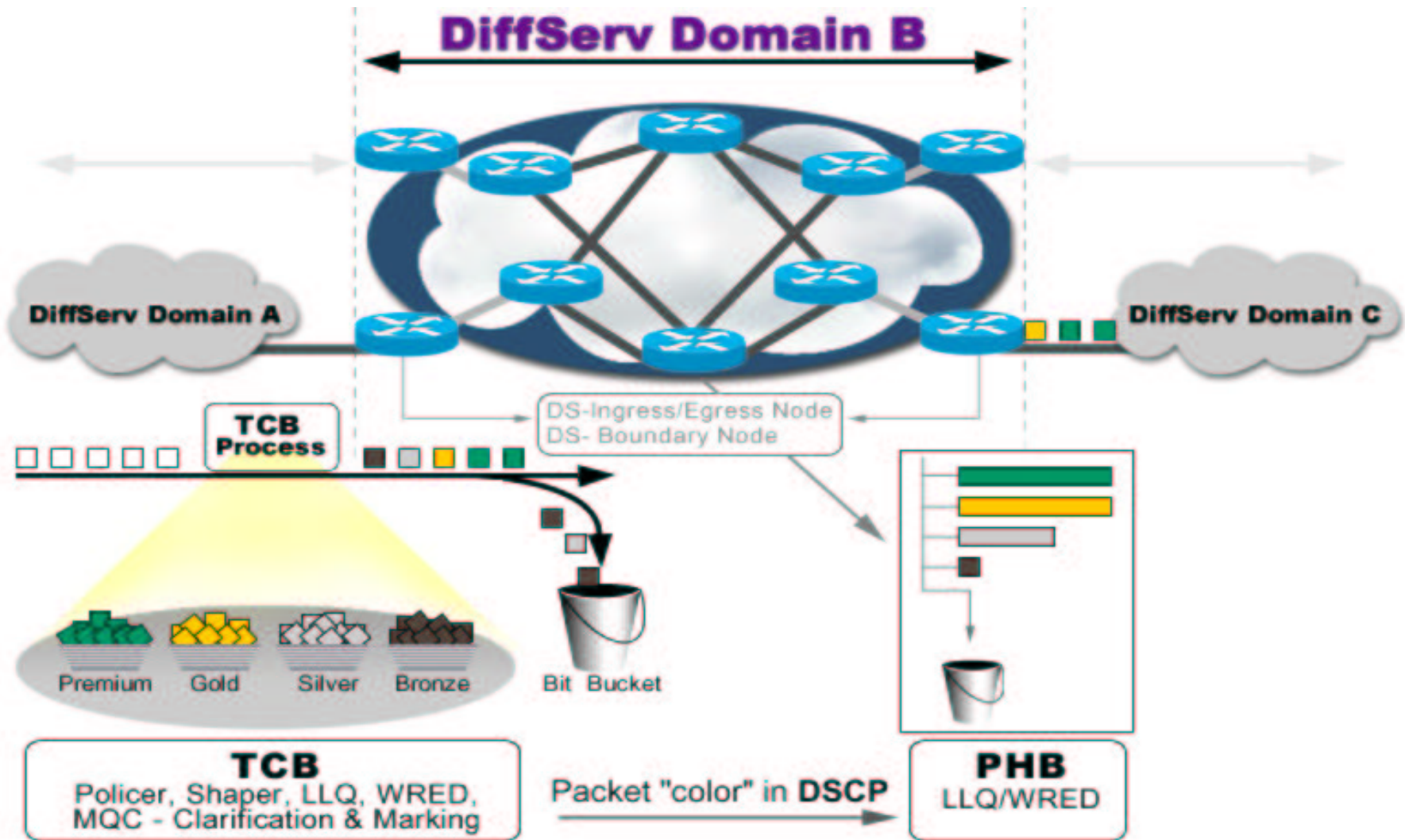
- Se simplifica radicalmente la interacción entre las aplicaciones y la red
- Se provee de una manera de identificar y así tratar adecuadamente a cada categoría (y a los flujos que la utilizan)
- Perdura desde el origen hasta el destino

- **Contras**

- Si no se diseña con cautela, una o más categorías pueden ver deteriorado su nivel de servicio

Cómo se define ? - Arquitectura DiffServ (RFC-2475)

Cisco.com



Cómo se define ? - DiffServ PHB

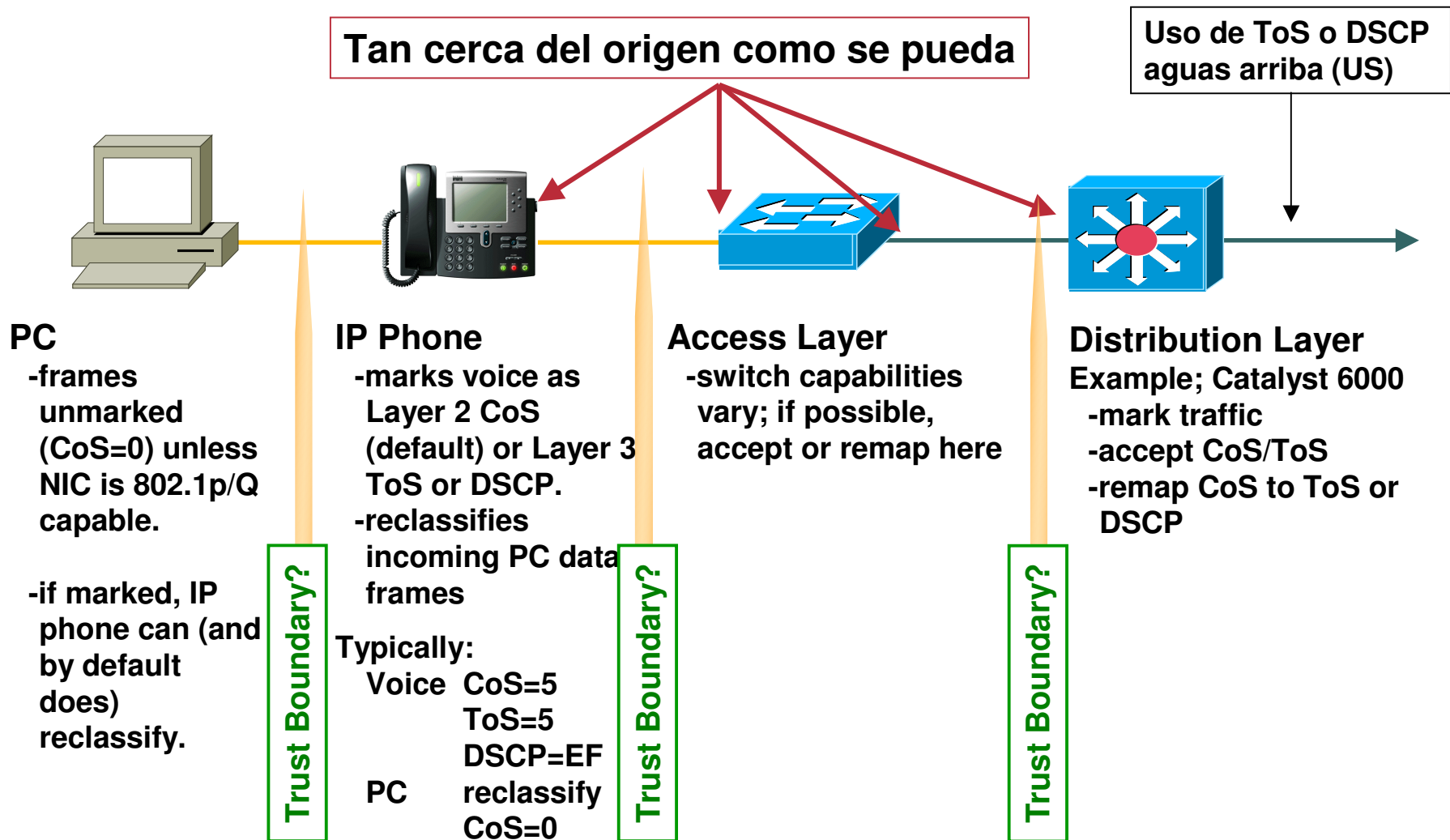
- **Uno de los elementos centrales de DiffServ es la siguiente definición de “Comportamientos por tramo” (PHBs)**
 - **Expedited Forwarding (EF): RFC2598**
 - Cola dedicada de Baja Latencia
 - *Comparable a “A/B dedicado” de IntServ (RSVP)*
 - **Assured Forwarding (AF): RFC2597**
 - n colas $\times$ m preferencias de descarte
 - *Comparable a “Carga Controlada” de IntServ*
 - **Class Selector: Compatible con IP Precedence (RFC1122)**
 - **Default: Mejor Esfuerzo o Buena Voluntad (best effort)**

Cómo se define ? - DiffServ DSCP

- Cada uno de los PHBs definidos se identifica mediante un grupo de códigos de punto de servicio diferenciado: DSCP (un número).
- Este código o MARCA DSCP se transporta en el encabezado IP
- Los diversos PHB se implementan mediante disciplinas de encolamiento y despacho (scheduling)

Cómo se define ? - Donde se marcará el tráfico ?

Cisco.com

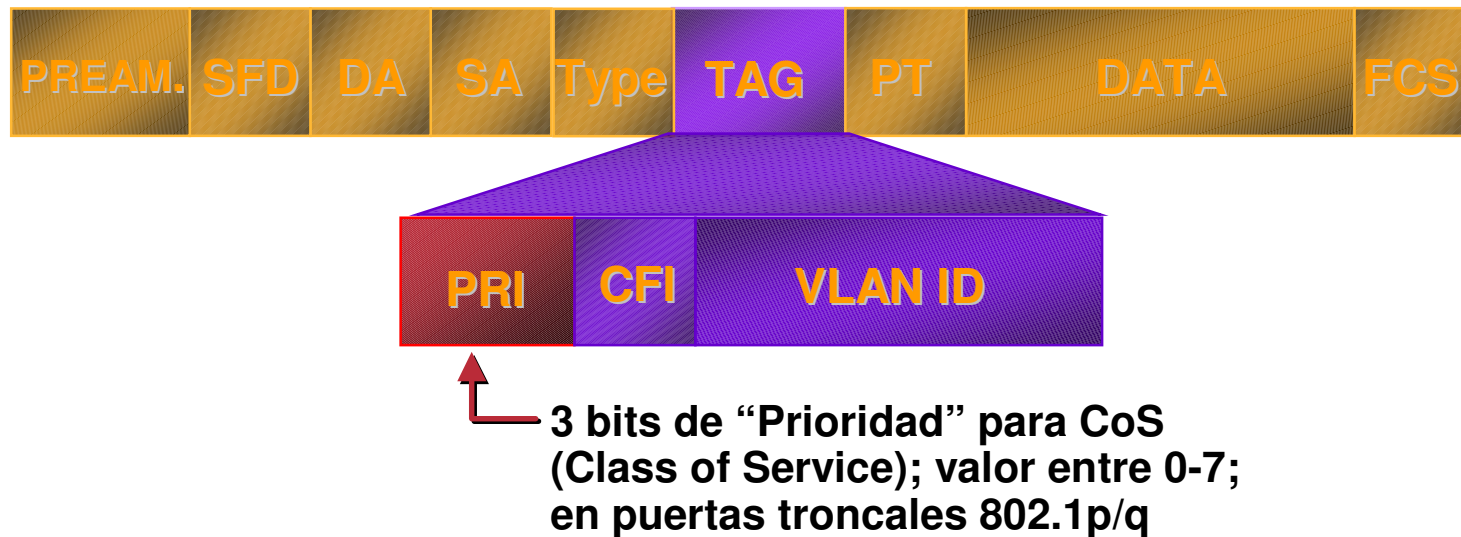


Cómo se define ? - Class of Service en capa 2

Cisco.com

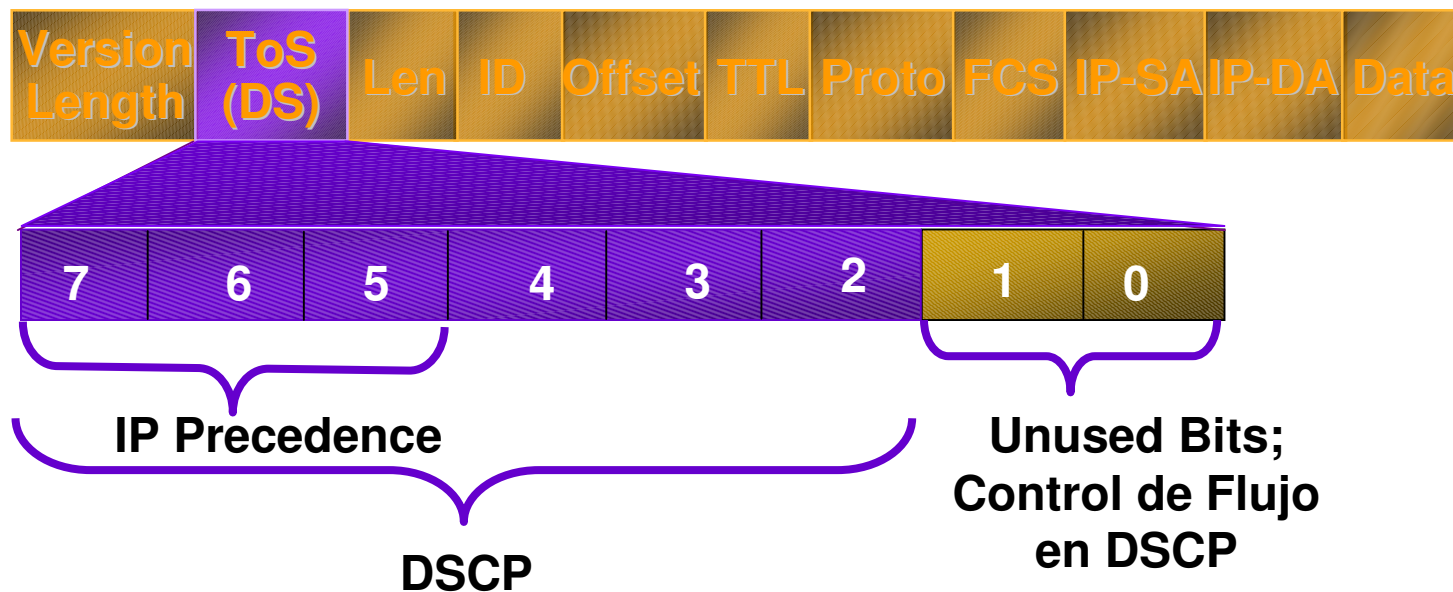
802.1p:

- Especificación del IEEE
- Focalizada en el soporte de QoS en LANs y en puertos 802.1Q (troncales)
- Soporta 8 Clases de Servicio
- Cesa donde termina el segmento LAN; es decir en el primer ROUTER



Cómo se define ? - Type of Service en capa 3

Capa 3, IPv4



**IPv4 Standard: 3 bits de “ToS” llamados IP Precedence;
DiffServ: 6 bits de DS llamados Códigos de Punto DS
(DSCP) más 2 para control de flujo**

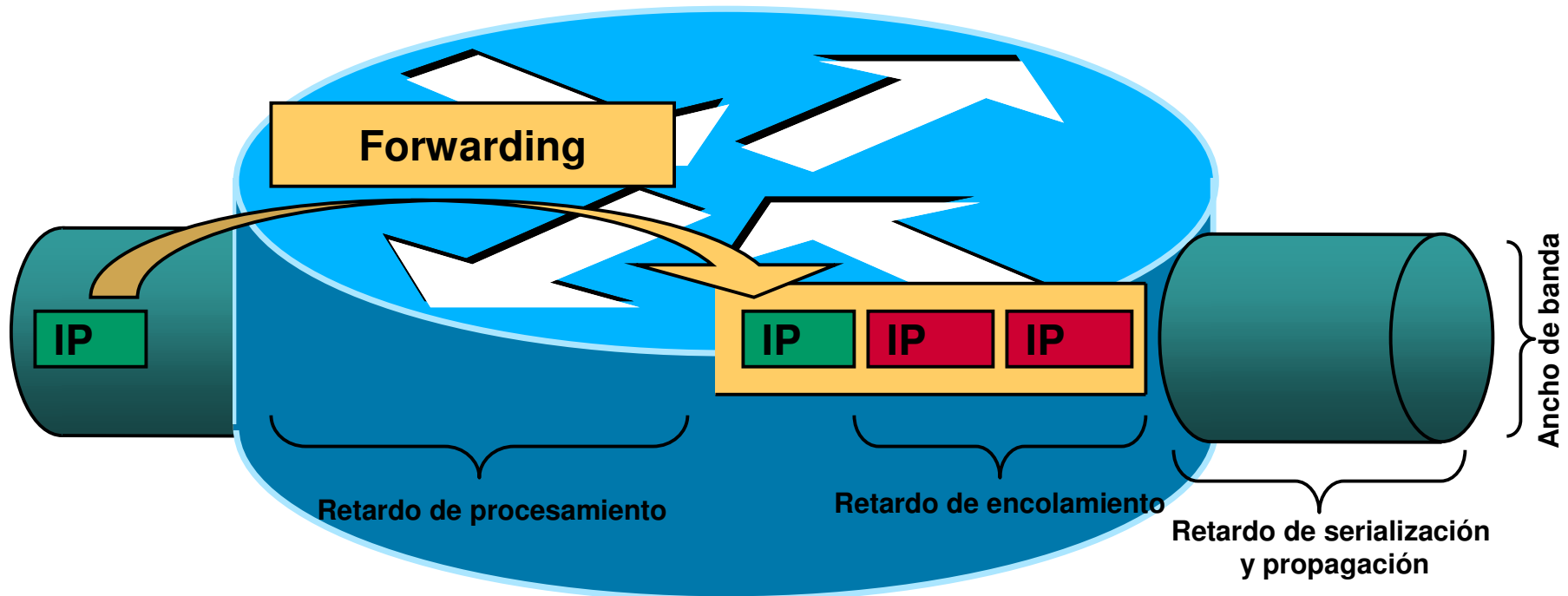
Cómo se define ? - Capas y Marcas

Cisco.com

- Hemos visto que se provéen mecanismos de marcado en capa 2 y 3
- *Ambos* son IMPRESCINDIBLES para lograr calidad de servicio de punta a cabo.
- Además, se deben poder establecer y/o reconocer dichas marcas en los todos los Routers y Switches

Cómo se define ? - Orígenes del retardo

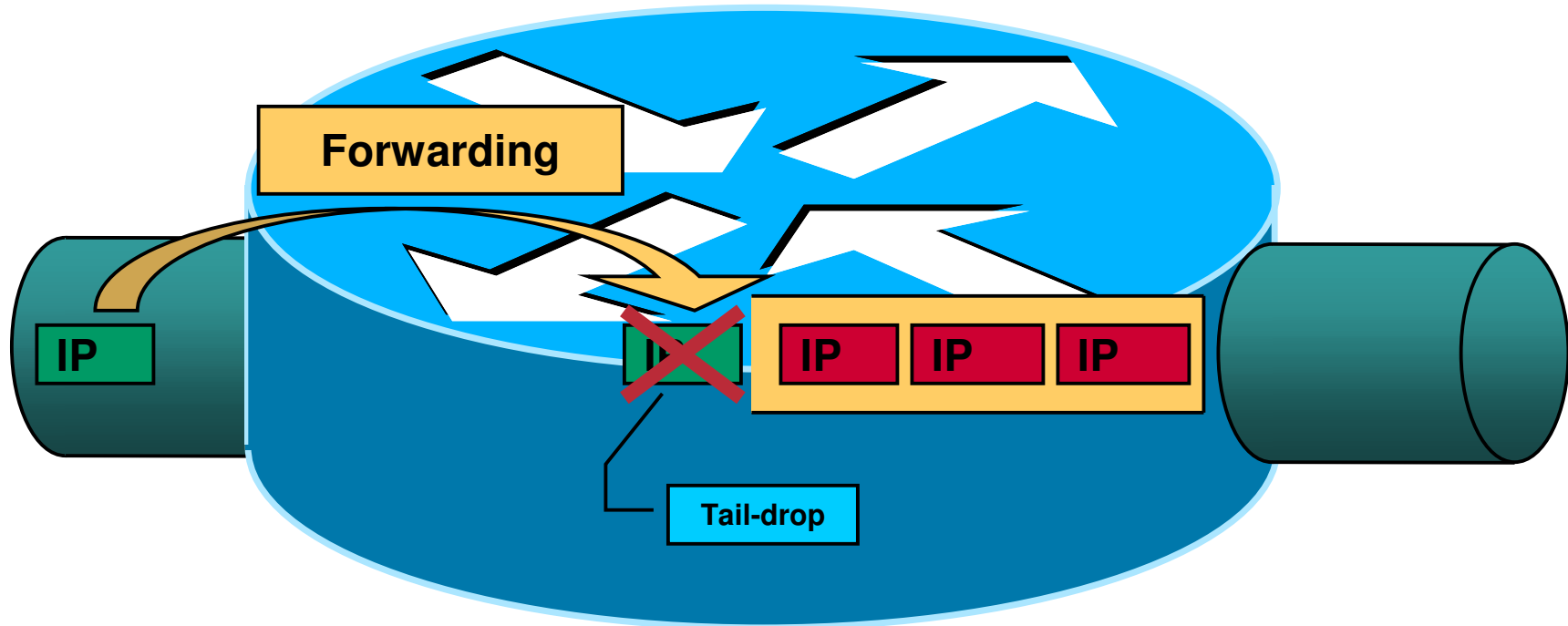
Cisco.com



- **Retardo de procesamiento** es lo que le toma al router para sacar un paquete de la interfase de ingreso, decidir cual será la interfase de egreso y ponerlo en la cola de salida de esta última.
- **Retardo de Encolamiento** es el tiempo que los paquetes permanecen en la cola de salida de una interfase.
- **Retardo de propagación y serialización** es lo que toma para vertir en el medio físico el paquete y para que llegue íntegro hasta el destino.

Cómo se define ? - Descarte de paquetes

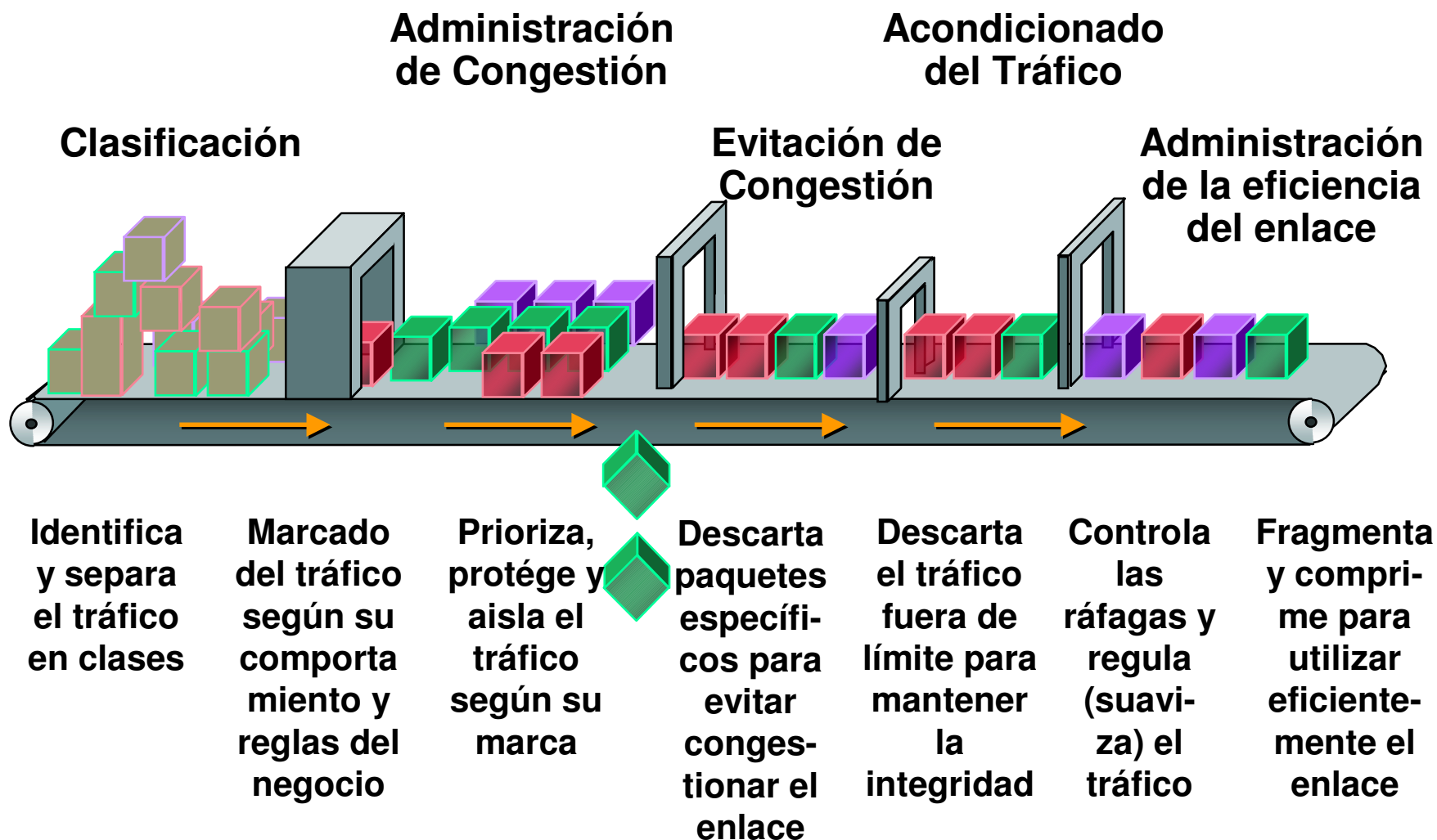
Cisco.com



- Los **Tail-drops** (descarte por la cola - en el sentido de rabo) suceden al estar la cola de salida llena. Son los más comunes y suceden cuando un ENLACE está congestionado.
- Existen muchos otros tipos de descarte menos comunes y que pueden requerir mejoras al hardware: descarte a la entrada, ignore, overrun, no buffer, ... Estos se deben frecuentemente a congestión del ROUTER.

Cómo se define ? - Mecanismos de QoS

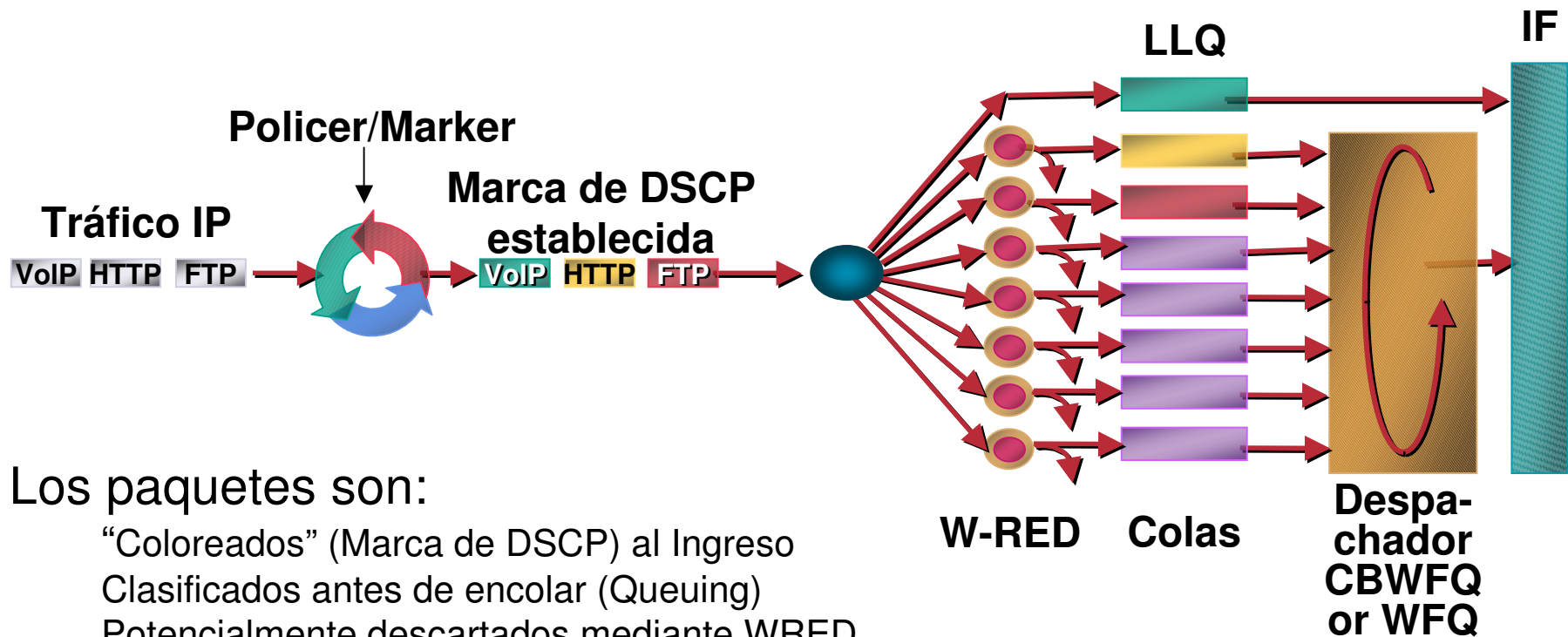
Cisco.com



Cómo se define ?

Cisco.com

- **Los elementos que participan**



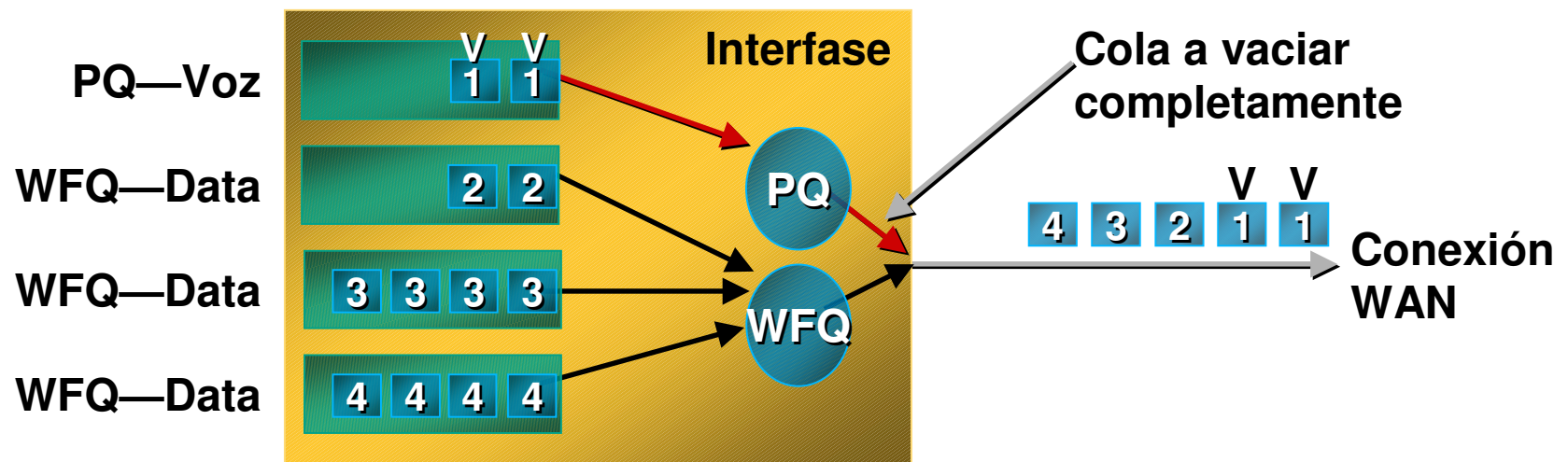
Los paquetes son:

- “Coloreados” (Marca de DSCP) al Ingreso
- Clasificados antes de encolar (Queuing)
- Potencialmente descartados mediante WRED
- Puestos en colas segun su marca de DSCP
- Programados para salir (scheduled) por CBWFQ (y/o LLQ)

Cómo se define ? - Low Latency Queuing

Cisco.com

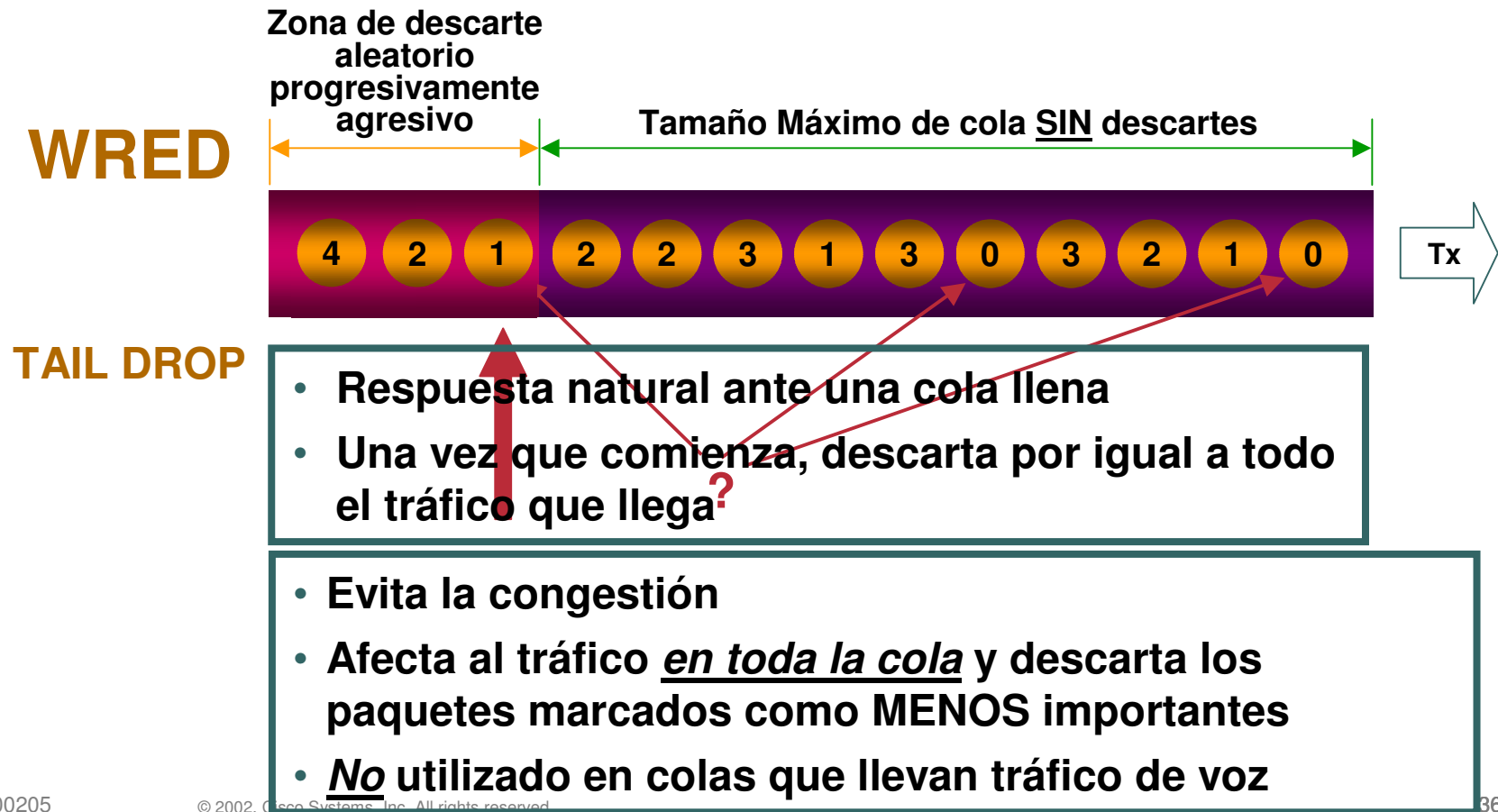
- Mecanismo de encolamiento y despacho de baja latencia



Cómo se define ? - Evitación de congestión

Cisco.com

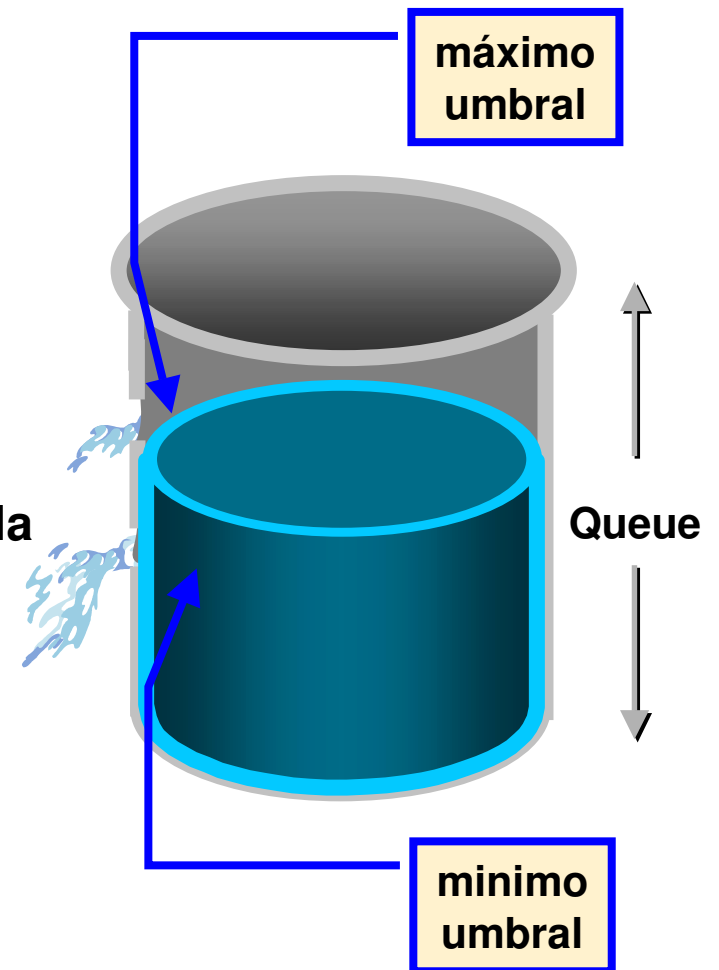
- **WRED: Descarte aleatorio temprano ponderado (weighted)**



Cómo se define ? - Random Early Detection (RED)

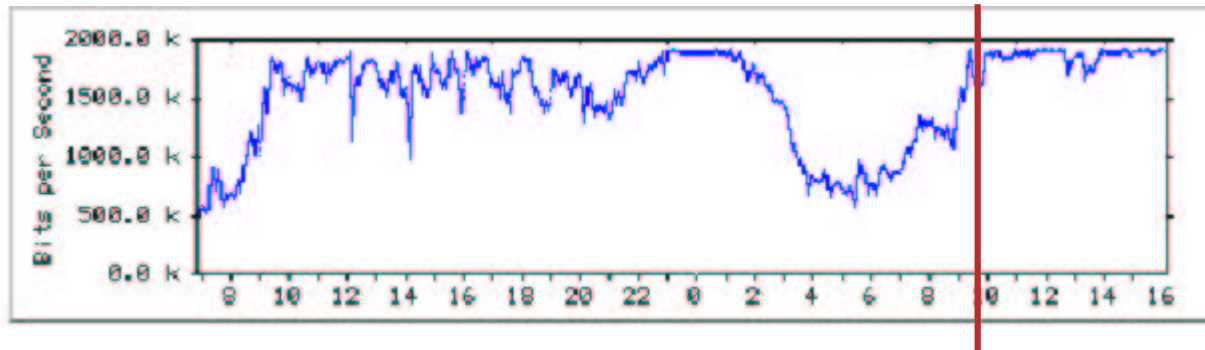
Cisco.com

- Monitorea la profundidad de cola para detectar congestión
- Distribuye el descarte sobre todas los flujos
- Permite que TCP se ajuste antes de insistir
- Descartes aleatorios para desincronizar las sesiones TCP y controlar su tasa
- Reduce el encolamiento promedio de largo plazo
- Descarta aleatoriamente los paquetes en toda la cola
- Aumenta la tasa de descarte con la profundidad de la cola
- **TODOS los paquetes pueden ser objeto de descarte**



Cómo se define ? - Efecto del RED

Cisco.com



Courtesy of Sean Doran, Ebone

RED Enabled

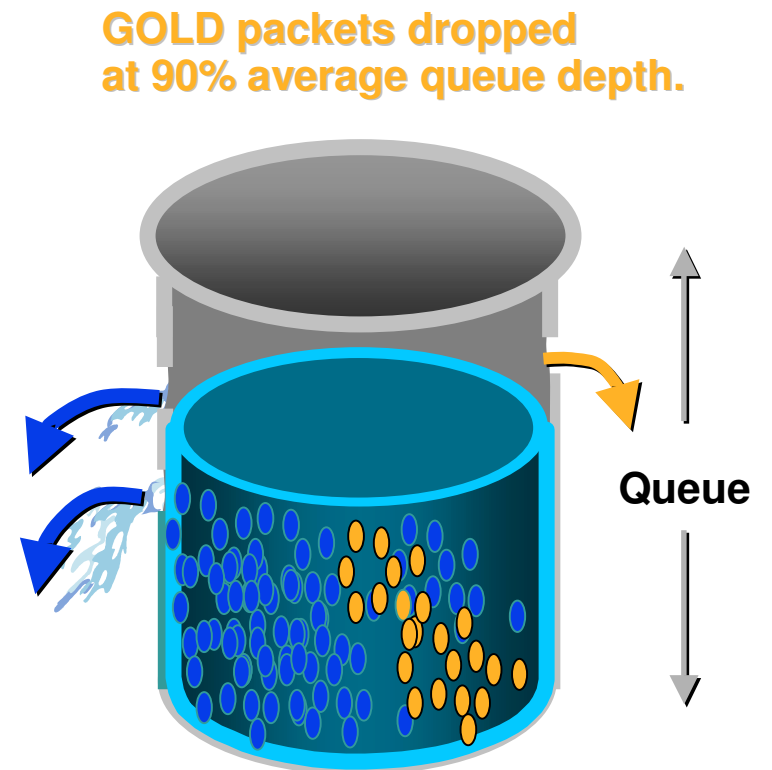
- **Un día completo, bajo el 100% de capacidad**
 - Encolamiento simple FIFO con tail drop
- **A las 10:00 del segundo día, capacidad al 100%**
 - Random early detection (RED) habilitado

Cómo se define ? - Weighted RED

Cisco.com

- **Weighted Random Early Detection (WRED):**
 - Implementación Cisco de RED
 - Si se requiere descartar, se descarta el tráfico de menor prioridad PRIMERO

**DiffServ
compliant
12.1(5)T**



BLUE packets start dropping at a 50% average queue depth. Drop rate increases as average queue depth increases.

Agenda

Cisco.com

- Qué es ?
- Porqué importa ?
- Cómo se define ?
- **Donde y como se aplica**
- Servicios a los que afecta
- End to End QoS y Policy Management
- Un ejemplo
- P&R

Donde y como se aplica

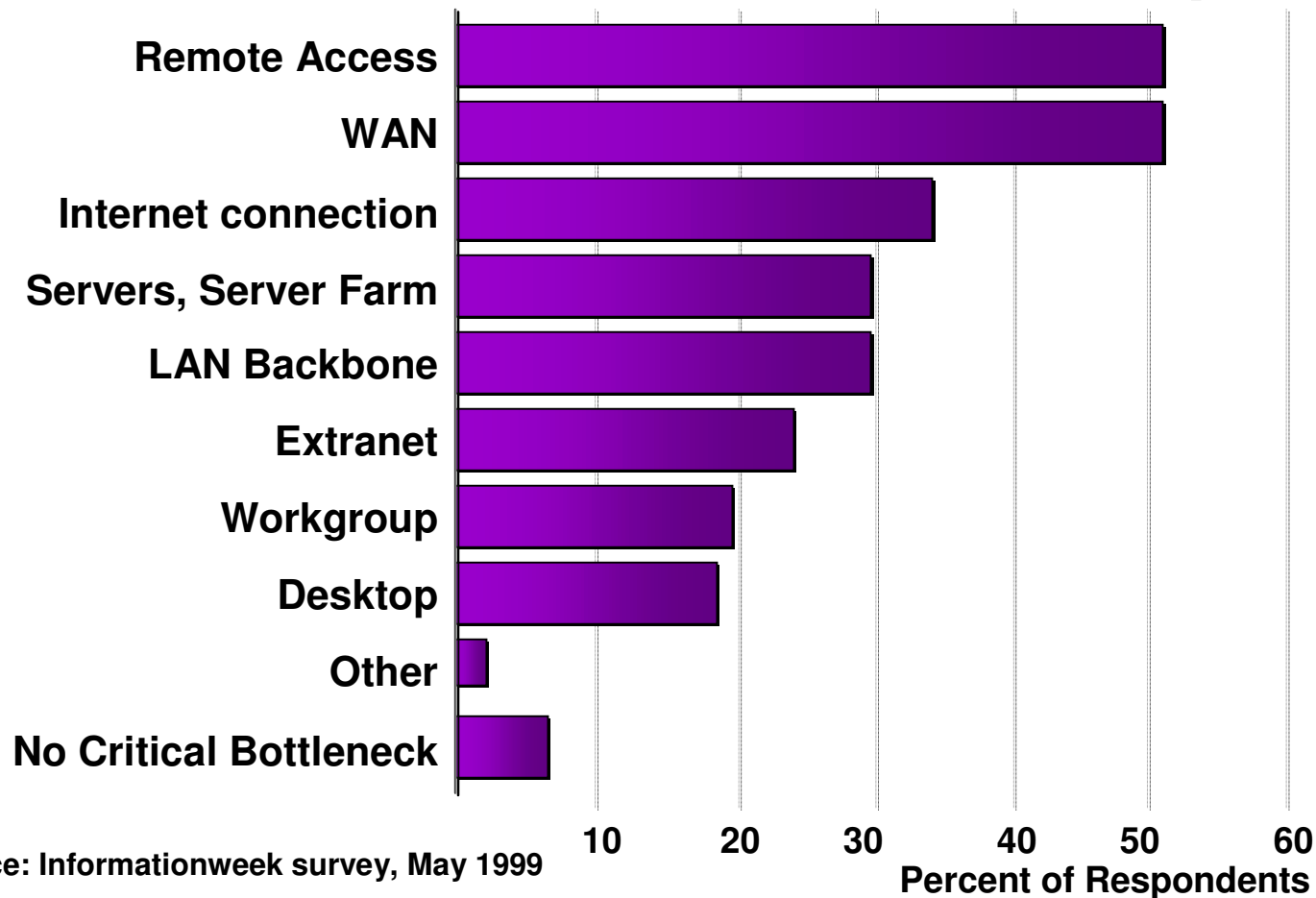
Cisco.com

- **En todo y cada enlace**
- **En todo y cada gateway (L2 y L3)**

Donde y como se aplica - Puntos de congestión en la red

Cisco.com

Donde están los cuellos de botella más críticos del ancho de banda de su red corporativa ?



Source: Informationweek survey, May 1999

- **Enlaces**
 - **Según la tecnología de cada caso:**
 - **Permiten diferenciar (marcar) el tratamiento de las unidades de información transmitidas (frames o paquetes)**
 - ◆ **Frame Relay**
 - ◆ **ATM**
 - ◆ **Ethernet 802.1p/q**
 - ◆ **Otros (SRP, ...)**

- **Gateways**
 - **Routers**
 - Permiten diferenciar el tratamiento a nivel de red
 - ◆ IP: ToS / DSCP
 - Implementan puntos de aplicación de reglas (Policy enforcement)
 - ◆ Marcado
 - ◆ Admisión
 - ◆ Despacho (Disciplinas de encolamiento)

Donde y como se aplica

Cisco.com

- Gateways

- Routers

- Implementan mecanismos de regulación y optimización
 - ◆ Fragmentación
 - ◆ Entrelace de paquetes (Interleaving)
 - ◆ Compresión
 - ◆ Regulación de la tasa de despacho (shaping)
 - Implementan mecanismos de admisión y descarte
 - ◆ Descarte aleatorio temprano (RED) de salida ante congestión
 - ◆ Descarte a la entrada para imponer la tasa de entrada (CAR y Policing)

Donde y como se aplica

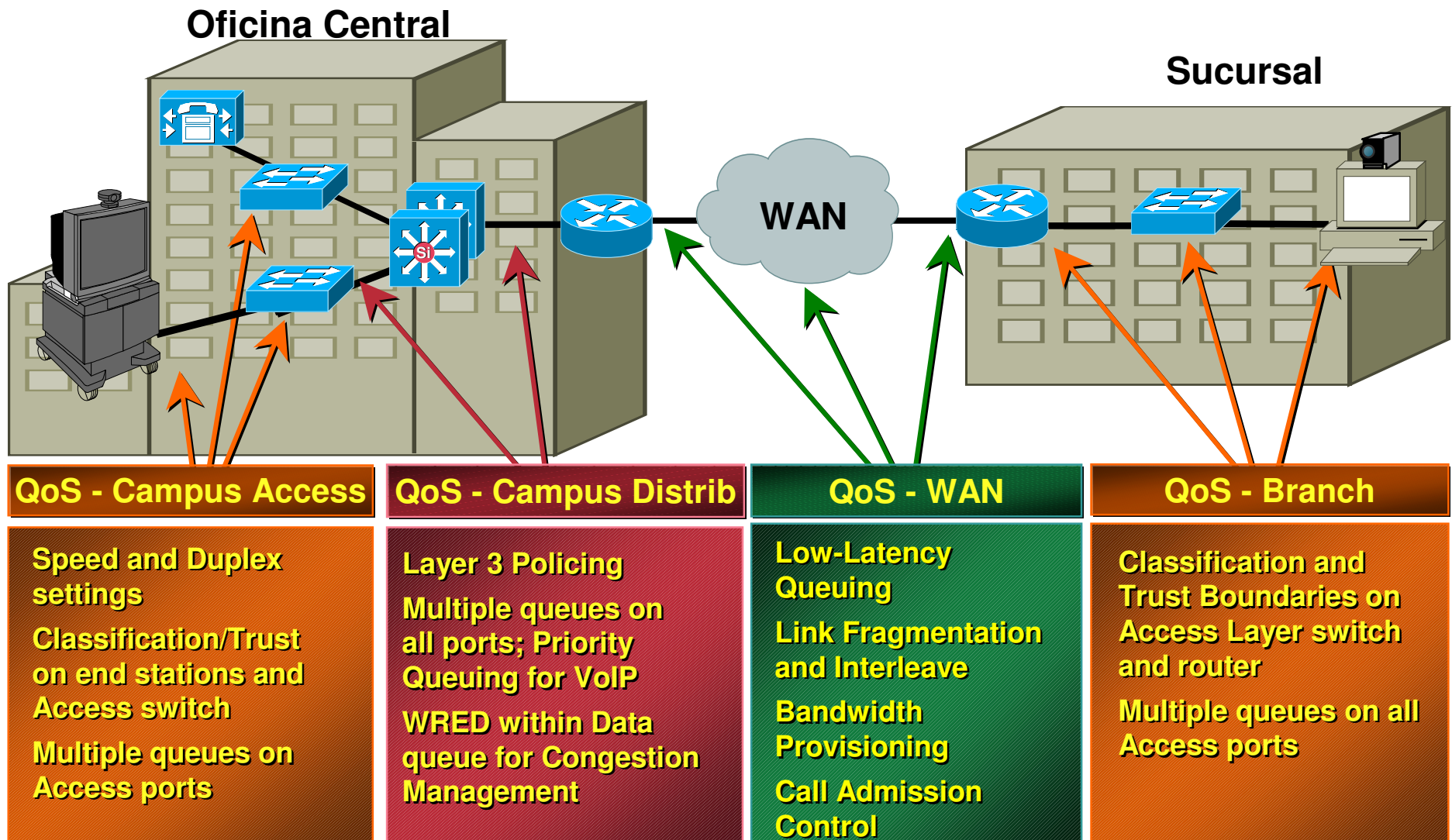
Cisco.com

- Gateways

- **Switches (L2/3/4) de Cisco Catalyst 2950 y 3550**
 - Implementan reglas a nivel de enlace de datos (capa 2) tanto como de capas superiores (3/4) sobrepasando las características tradicionales de un LAN switch.
 - Capaces de realizar
 - ◆ Marcado
 - ◆ Imponer reglas de Admisión y
 - ◆ Disciplinas de encolamiento y Despacho (**LLQ**, **RED**, ... !)
- **Incluso inspeccionan el código DSCP y otros atributos a nivel de TCP/IP (L4; Ports, Origen, ...)**

Donde y como se aplica En que parte de la red?

Cisco.com



Donde y como se aplica

Cisco.com

- El conjunto de los GW y enlaces dotados de los atributos mencionados conforman una red capaz de otorgar calidad de servicio de punta a cabo.

Agenda

Cisco.com

- Qué es ?
- Porqué importa ?
- Cómo se define ?
- Donde y como se aplica
- **Servicios a los que afecta**
- End to End QoS y Policy Management
- Un ejemplo
- P&R

Servicios a los que afecta

Cisco.com

- **Algunos servicios:**
 - **Telefonía IP (voz)**
 - **Distribución de video**
 - **Videoconferencia**
 - **Tratamiento de imágenes**
 - **Diseño y Manufactura asistida**
 - **Aplicaciones críticas (transaccionales)**
 - **Aplicaciones NO críticas**
 - **Niveles de servicio diversos por grupos de usuarios**

Servicios a los que afecta

Cisco.com

- **La red típica tendrá las siguientes aplicaciones:**
 - **Conversacional en tiempo real (i.e. Voz)**
 - **Clase de tiempo-real con streaming (i.e. Real-Video)**
 - **Clase interactiva (i.e. clientes Telnet, HTTP)**
 - **Background / Masivo (i.e. FTP, Dbase Synch)**

- **Desde el punto de vista corporativo**
 - **Habilita la convergencia de servicios sobre la red corporativa**
 - **Optimiza el uso de los recursos**
 - **Habilita el contar con y controlar el “nivel de servicio” de los transportes WAN**

- **Desde el punto de vista del proveedor de servicio**
 - **Habilita la diferenciación por “nivel de servicio” de la oferta de transporte WAN**
 - **Optimiza el uso de los recursos de la red de área extensa (Backbone WAN)**

Agenda

Cisco.com

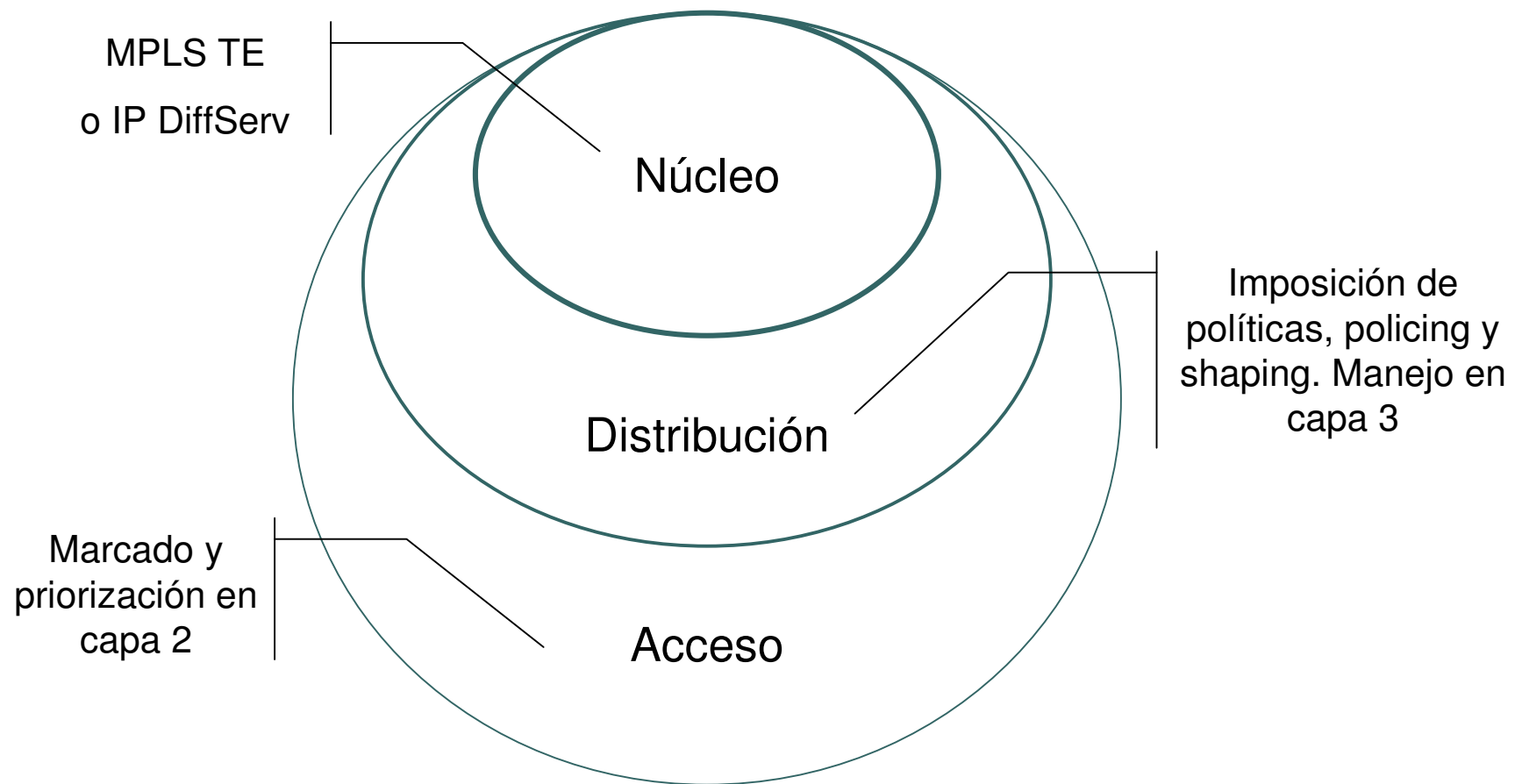
- Que es
- Por que importa
- Como se define
- Donde y como se aplica
- Servicios a los que afecta
- **End to End QoS y Policy Management**
- Un ejemplo
- P&R

E2E QoS y Policy Management

Cisco.com

- **Disponer de QoS es un asunto de toda la red**
 - **Desde el “punto” de conexión del usuario hasta el del servidor**

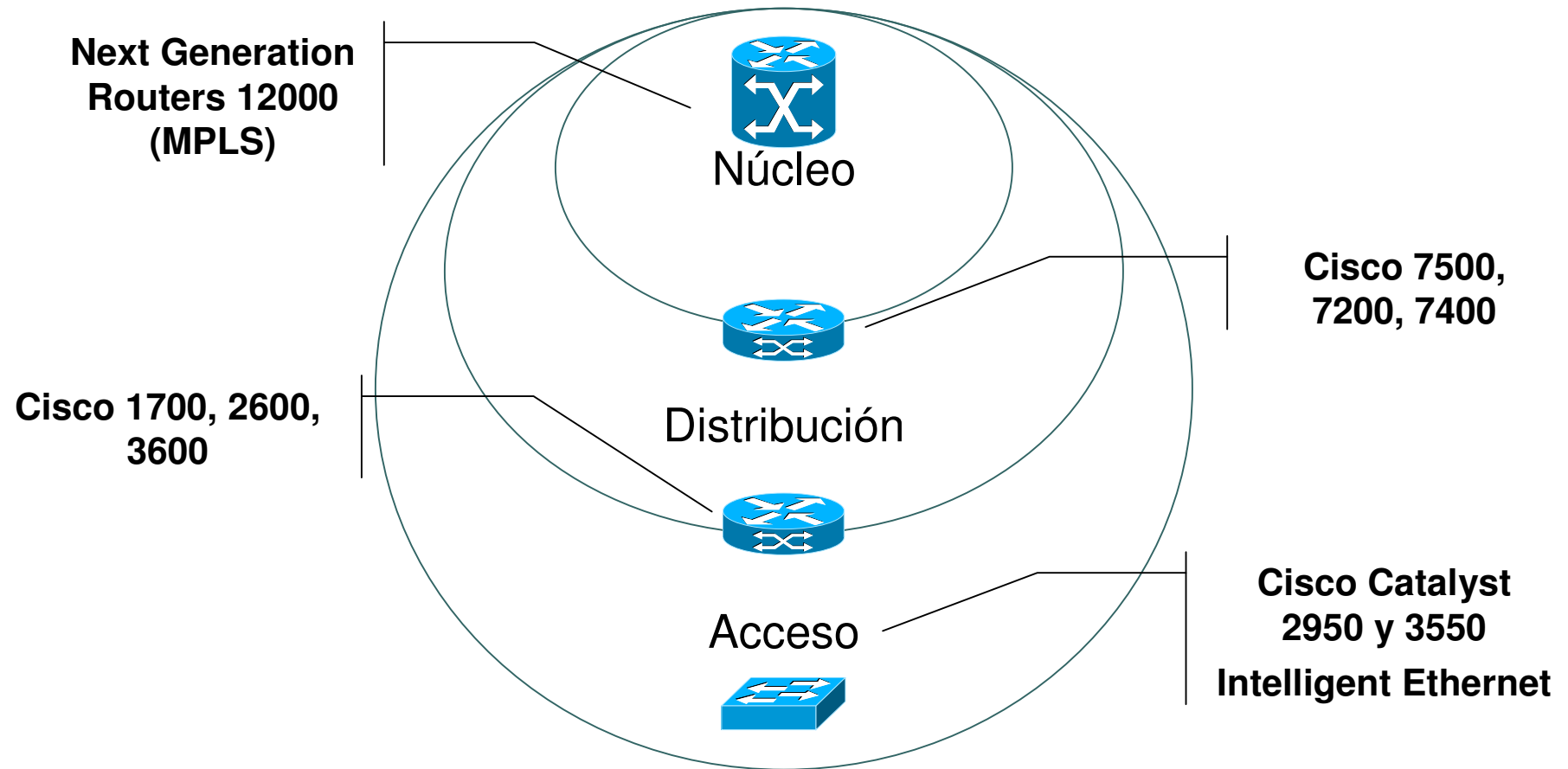
- **QoS en la arquitectura de red**



E2E QoS y Policy Management

Cisco.com

- QoS en la arquitectura de red



E2E QoS y Policy Management

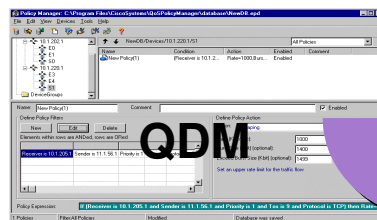
Cisco.com

CONFIGURE

VERIFY

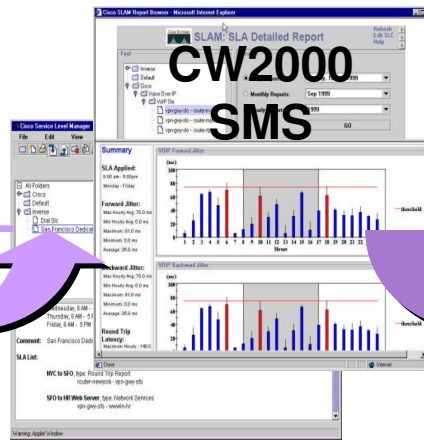
TROUBLESHOOT

Network Wide



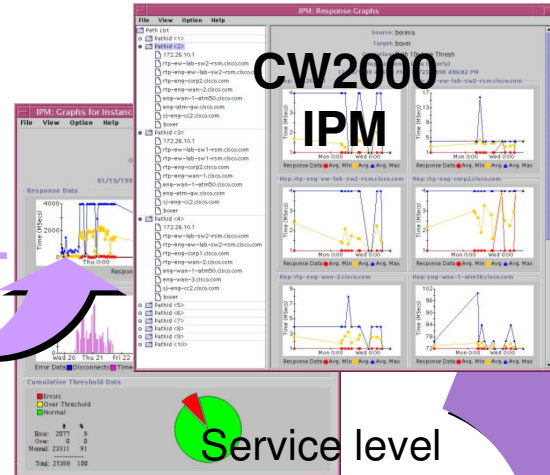
QoS network
policy configuration

XML



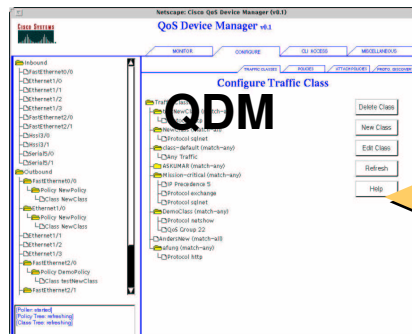
Network service
level verification

XML

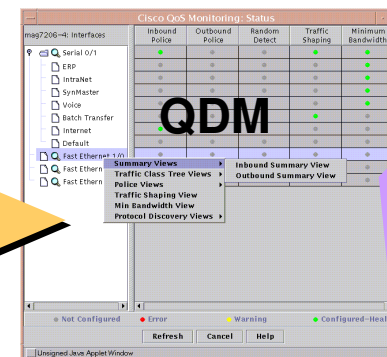


Service level
troubleshooting

Device



Per-device traffic
class configuration



Per-device traffic
class monitoring

Agenda

Cisco.com

- Que es
- Por que importa
- Como se define
- Donde y como se aplica
- Servicios a los que afecta
- End to End QoS y Policy Management
- **Un ejemplo**
- P&R

Ejemplo: Arquitectura de red corporativa jerárquica

Cisco.com

QoS Policy

High bandwidth

Policies:

Drop (congestion avoidance)

Consolidation point

Speed mismatches

Policies:

(re)Mark, Schedule, Police, Drop

Services:

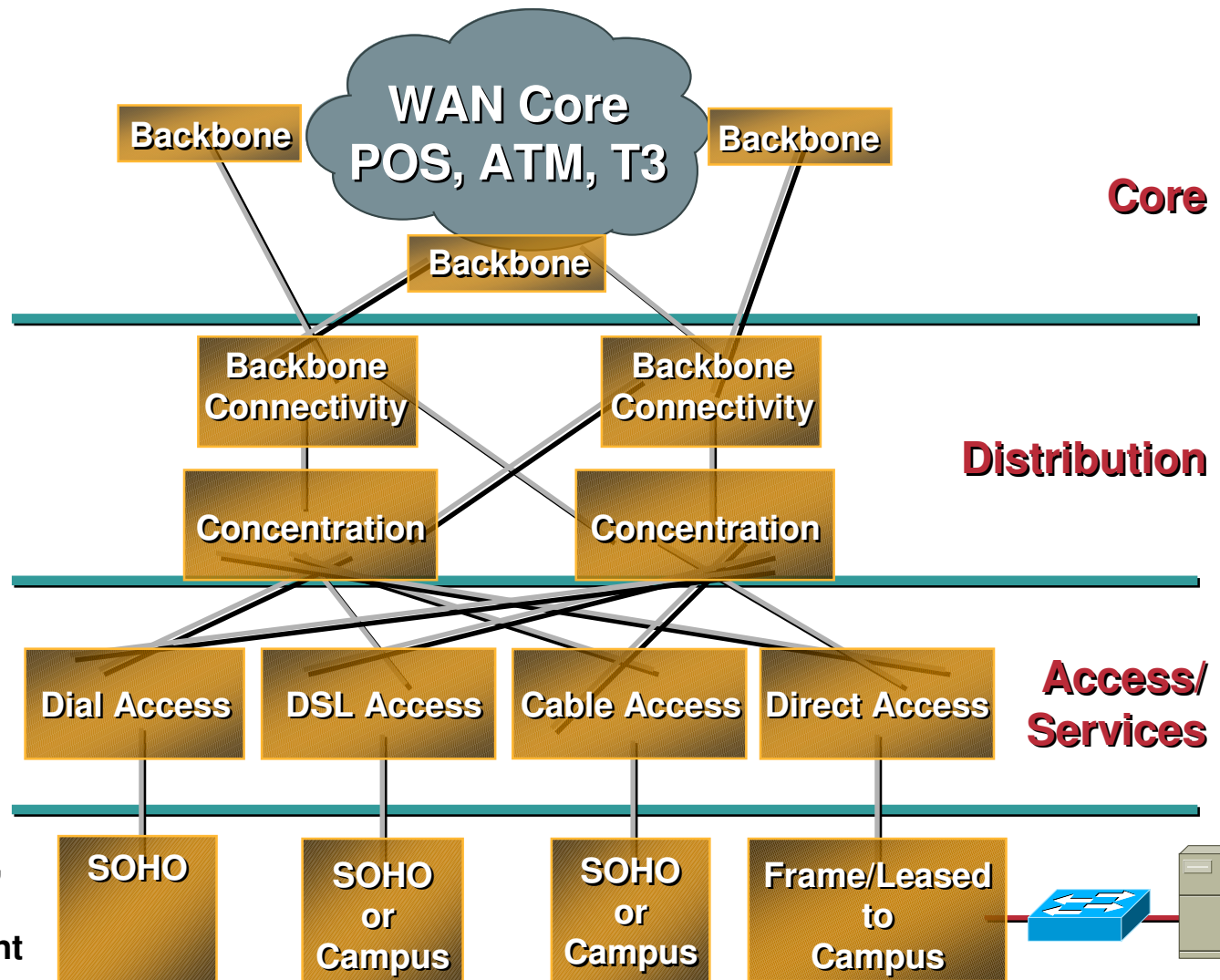
Caching, Firewall

Demarcation point
CPE/PE

Lower Bandwidth

Policies:

Mark, Call Admission, Schedule, Shape, Drop, Fragment



CISCO SYSTEMS

