

GG2

Tema 1-1: Introducción a las técnicas usadas en la calidad de servicio (QoS)

José Manuel Arco Rodríguez

Índice

- ♦ Introducción a la QoS
- ♦ Necesidades del tráfico de usuario
- ♦ Técnicas utilizadas en QoS
 - Control de admisión de conexiones
 - Encaminamiento con QoS
 - Conformado
 - Función Policía
 - Control de congestión
 - Control de flujo
 - Planificadores de tráfico
- ♦ Señalización

Tráfico generado

- ♦ El tráfico de las aplicaciones multimedia se puede considerar que está formado por 3 componentes: vídeo, audio y datos
- ♦ Los flujos multimedia pueden ser caracterizados por los siguientes factores:
 - La variación de la velocidad de emisión
 - Constante, un CD o DVD o con codificaciones de audio o vídeo a velocidad constante
 - Variable, codificadores que aprovechan los silencios para no enviar tráfico
 - Interacción entre usuarios, tráfico será en tiempo real o no (videoconferencia, TV o radio por Internet)

Tráfico generado

- ♦ Simetría en el tráfico en cada sentido de transmisión, e.g.* las conexiones domesticas a Internet tiene mucha más velocidad de bajada que de subida. Esto se puede aprovechar para optimizar recursos reservando más BW en los interfaces dirección de ida al abonado
- ♦ Hay aplicaciones que pueden necesitar los tres tipos de datos
 - ♦ Videoconferencia con pizarra electrónica
 - ♦ Juegos en línea
- ♦ e.g.* significa por ejemplo
- ♦ i.e. significa, quiere decir

Requerimiento del tráfico generado

- ♦ Vídeo: Requiere mucho ancho de banda sin comprimir (hasta 48.000 Mbps con resolución 4K (8000x4000 x 120 fps*8))
- ♦ Comprimidos
 - ♦ TDT (720 líneas) HD (1080 líneas) de 2 a 8 Mbps
 - ♦ 4k de 25 a 800 Mbps
 - ♦ Retardos aceptables de 150 mseg
- ♦ Audio: Requiere retardo bajo, máximo 150 mseg, BW 0,064 Mbps *
- ♦ Datos: sin grandes requerimientos

♦ * Para apreciar mas las diferencias, comparar los BW como si 1 Mbps equivaliera a 1 €

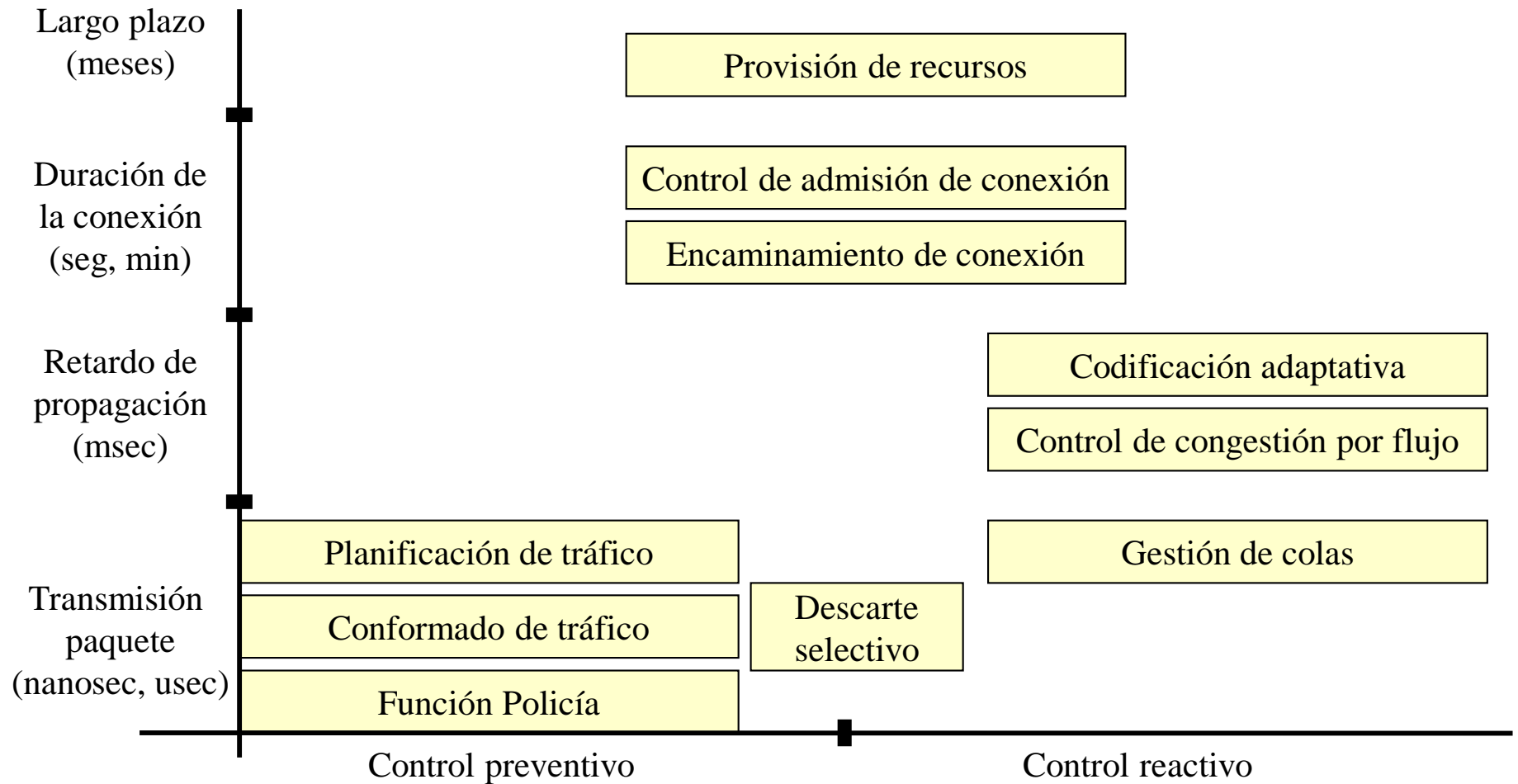
Índice

- ♦ Introducción a la QoS
- ♦ Necesidades del tráfico de usuario
- ♦ Técnicas utilizadas en QoS
 - Control de admisión de conexiones
 - Encaminamiento con QoS
 - Conformado
 - Función Policía
 - Control de congestión
 - Control de flujo
 - Planificadores de tráfico
- ♦ Señalización

Técnicas de control de trafico en Qos

- ♦ En la siguiente diapositiva se muestran las técnicas más importantes en función de:
 - Si son consideradas reactivas a la congestión, reaccionan cuando aparece, o preventivas, reaccionan antes de que aparezca la congestión
 - Escala de tiempos en donde actúan
- ♦ Estos tiempos puede ser muy cortos, el tiempo de transmisión un paquete (nano sec) o a largo plazo, varios meses
- ♦ Algunas técnicas pueden considerarse que se comportan tando de forma reactiva como preventiva

Técnicas de control de trafico en Qos



Provisionado de recursos

- ◆ El objetivo es hacer un dimensionado de la red para que el nivel de rechazo de las solicitudes o conexiones de QoS de los usuarios, sea muy bajo
 - Cuando un usuario pide una QoS y la red no tiene recursos para atenderla, la red rechaza la solicitud
- ◆ Incluye un buen dimensionado de la red en la topología, número de enlaces y su ancho de banda, número de conmutadores, etc
 - Un infraprovisionado puede hacer perder clientes al operador
 - Un sobreprovisionado puede encarecer el servicio al cliente
- ◆ El dimensionado es complicado ya que cambia el número de usuarios y el tráfico de las aplicaciones
 - En la práctica para evitar problemas se tiende a sobredimensionar

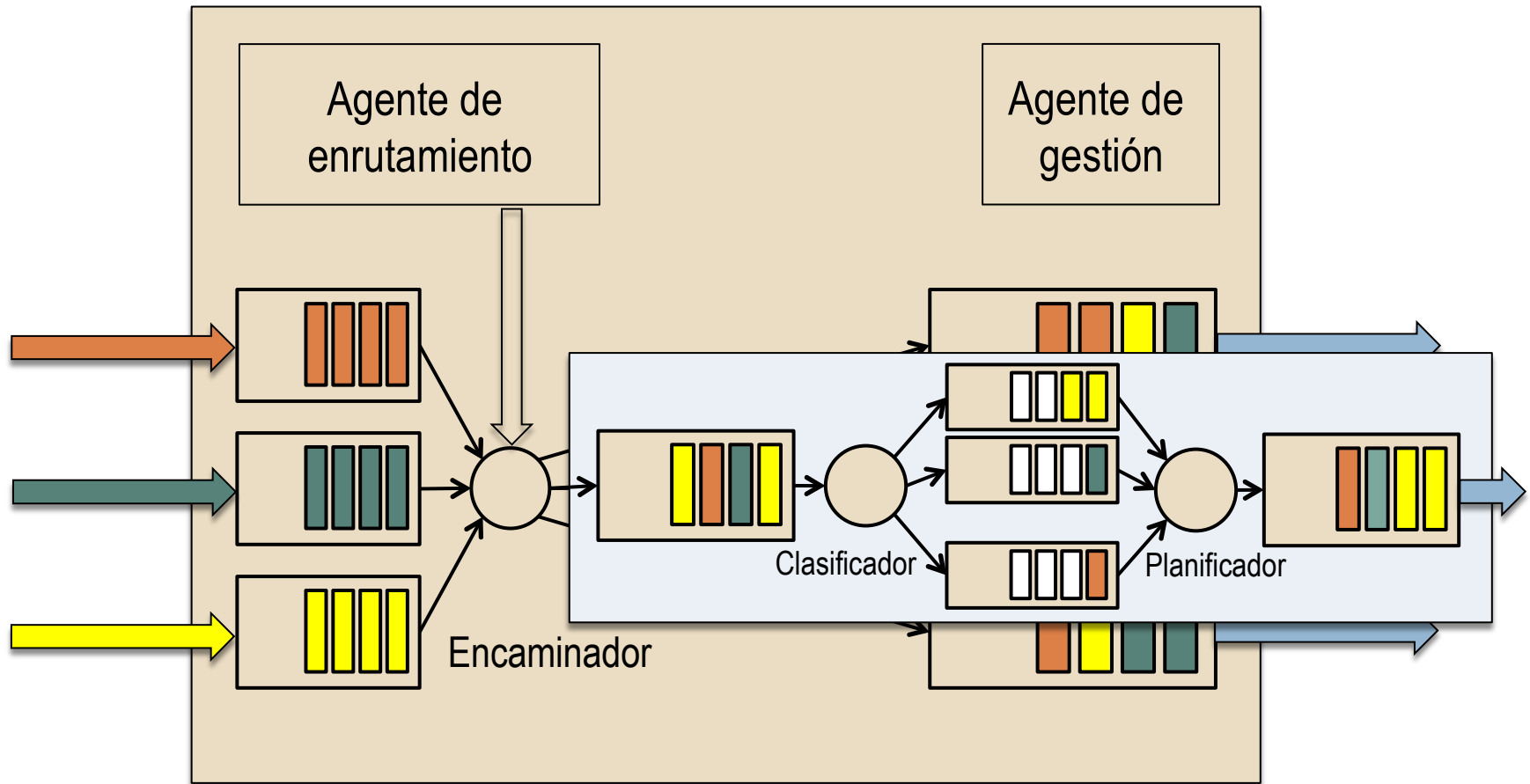
Índice

- ♦ Introducción a la QoS
- ♦ Necesidades del tráfico de usuario
- ♦ Técnicas utilizadas en QoS
 - Control de admisión de conexiones
 - Encaminamiento con QoS
 - Conformado
 - Función Policía
 - Control de congestión
 - Control de flujo
 - Gestores de colas
 - Planificadores de tráfico
- ♦ Señalización

Secuencia de operaciones sobre un flujo

- ♦ En el lado cliente:
 - Conformado de tráfico
- ♦ En los nodos de red (al menos en los de acceso):
 - Clasificación: aislamiento de flujos
 - Control de admisión
- ♦ Gestión activa de las colas
- ♦ Función policía: según el contrato, marcado y/o conformado.
- ♦ Enrutamiento: selección de la interfaz de salida.
- ♦ Clasificación: por requisitos de QoS.
- ♦ Planificación: por prioridades de las distintas clases de QoS.

Secuencia de operaciones sobre un flujo



Clasificación

- ◆ Objetivo: identificar los flujos, conexiones o clases de tráfico
- ◆ En redes conectivas está implícito en el identificador de conexión
- ◆ En redes no conectivas hay muchas posibilidades en base a campos de las cabeceras de red y de transporte
- ◆ Se suele tomar algunos o todos de los 5 campos que identifican un flujo IP
 - Dirección IP origen y destino
 - Aunque no son parte de la cabecera IP los puertos siempre van a continuación de la misma, sin fragmentación
 - También hay que acceder al campo *protocolo* para ver si es un puerto TCP o UDP
- ◆ También en base al campo IP ToS (Type of Service)

Control de admisión de conexiones (CAC)

- ◆ El CAC comprueba que haya suficientes recursos para soportar la nueva conexión con la QoS solicitada y que no afecte a las QoS de las conexiones ya establecidas
- ◆ La conexión se acepta o rechaza en función de lo anterior
- ◆ En redes conectivas, se realiza en la fase de establecimiento de la conexión en los diferentes nodos de la red
- ◆ En redes no conectivas, al solicitar un servicio, eg. al llamar por un teléfono IP

Algoritmos de CAC

- ◆ El CAC no está normalizado, cada operador implementa el que considera más conveniente
- ◆ Hay tres enfoques basados en evaluar alguno de los parámetros siguientes:
 - Ancho de banda (mas intuitivo)
 - Tasa de errores
 - Retardo
- ◆ El CAC basado en ancho de banda; determina el ancho de banda requerido para su QoS y comprueba que las suma de BW reservado y el nuevo, no supera la V_{tx}
 - Hay dos alternativas multiplexión determinista o estadística

CAC multiplexión determinista

- ◆ El ancho de banda reservado a una conexión es el máximo que pueda necesitar
- ◆ Ventaja, retardos y tasa de errores reducidos al mínimo
- ◆ Inconveniente, se infrautiliza la red, se reserva la máxima velocidad a una conexión cuando la mayor parte del tiempo no se transmite al máximo (o no se transmite)
- ◆ **Ganancia estadística**, analizando el tráfico conjunto de miles de usuarios, se ve que por el efecto citado, el tráfico medio emitido es muy inferior al máximo posible
- ◆

CAC de multiplexión estadística

- ◆ Aprovecha la ganancia estadística
- ◆ El ancho de banda reservado a una conexión (capacidad equivalente, **Ce**) está comprendido entre el máximo y el medio
- ◆ La suma de las velocidades máximas de las conexiones puede superar a la velocidad del enlace, no así la suma de los Ce
- ◆ La ganancia en ancho de banda aumenta cuando el Ce se aproxima a la velocidad media y disminuye al acercarse a la de pico
- ◆ Ejemplo para ATM

Servicio	Capacidad equivalente
Constant Bit Rate	$Ce = V_{max}$
Variable Bit Rate	$V_{media} \leq Ce \leq V_{max}$
Available Bit Rate	$Ce = V_{min}$
Unspecified Bit Rate	No aplicable, $Ce=0$

Índice

- ♦ Introducción a la QoS
- ♦ Necesidades del tráfico de usuario
- ♦ Técnicas utilizadas en QoS
 - Control de admisión de conexiones
 - Encaminamiento con QoS
 - Conformado
 - Función Policía
 - Control de congestión
 - Control de flujo
 - Gestores de colas
 - Planificadores de tráfico
- ♦ Señalización

Encaminamiento QoS en redes conectivas

- ◆ Objetivo, abrir la conexión por el camino que tenga los recursos necesarios para dar la QoS
 - Si la aplicación usa varios tipos de tráfico (con distintas necesidades) cada uno debería ir por una ruta diferente
- ◆ Objetivo, abrir cada conexión por el camino que tenga los recursos requeridos
- ◆ En ATM, el protocolo PNNI (Private Network-to-Network Interface), se usa un protocolo distribuido que calcula la ruta más corta a un destino que ADEMÁS cumple con la QoS requerida

Encaminamiento QoS en redes NO conectivas

- ♦ En IP no es posible
 - En OSPF el tráfico va SOLO por una ruta, la de menor coste
 - Recordar $\text{Coste} = 10^8 / \text{BW}$; 100 Mbps=1 10 Mbps=10
 - Si todos los enlaces son iguales, la de menos saltos=menor coste
 - No diseñado para QoS
- ♦ Se puede solucionar instalando MPLS-TE (Traffic Engineering)
- ♦ Más adelante se estudiará con más detalle, la idea es que el nodo de entrada:
 - Tiene un mapa de la topología y recursos de la red
 - Calcula el camino con menor coste y que cumple con la QoS
 - Establece un camino por la ruta calculada

Índice

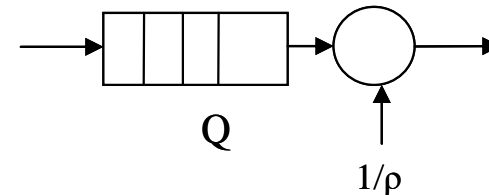
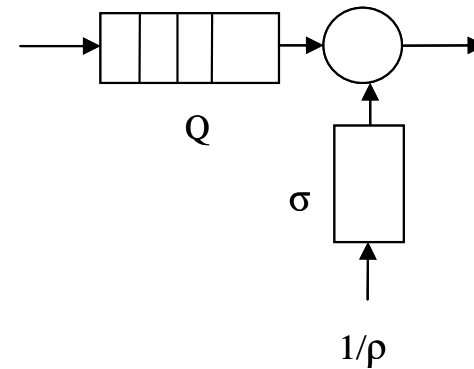
- ♦ Introducción a la QoS
- ♦ Necesidades del tráfico de usuario
- ♦ Técnicas utilizadas en QoS
 - Control de admisión de conexiones
 - Encaminamiento con QoS
 - Conformado
 - Función Policía
 - Control de congestión
 - Control de flujo
 - Planificadores de tráfico
- ♦ Señalización

Conformado del tráfico (Shaping)

- ◆ El conformado se hace en el equipo del usuario emisor durante la transmisión de datos
- ◆ El conformado controla que el tráfico generado no incumpla lo declarado, para evitar que actúe la función policía
- ◆ Cuando actúa retiene los paquetes hasta que cumplan las condiciones
- ◆ **Modifica, cambia, regula o aplana** el patrón de emisión
- ◆ Los algoritmos más empleados son variantes del Token Bucket (leer pg 5 de la práctica 2 y Kurose pg 624)
- ◆ Similar a un peaje en una autopista

Algoritmos de conformado

- ◆ Token Bucket genérico
- ◆ Realiza conformado con ráfaga
- ◆ Token Bucket
- ◆ Realiza conformado a velocidad constante



- ◆ En teoría el conformado puede tener pérdidas, si se llena el buffer Q , pero si se evita bajando la velocidad de la fuente de tráfico que está en el mismo ordenador

Índice

- ♦ Introducción a la QoS
- ♦ Necesidades del tráfico de usuario
- ♦ Técnicas utilizadas en QoS
 - Control de admisión de conexiones
 - Encaminamiento con QoS
 - Conformado
 - Función Policía
 - Control de congestión
 - Control de flujo
 - Planificadores de tráfico
- ♦ Señalización

Función policía (*Policing*)

- ◆ También llamada *vigilancia* o *limitación*
- ◆ Se lleva a cabo en el nodo de acceso a la red del operador durante la transmisión de datos
- ◆ La función policía vigila que el tráfico generado está dentro de lo declarado
- ◆ La vigilancia, puede **recortar** el tráfico en exceso
- ◆ En caso de incumplimiento hay tres opciones:
 - Descartar paquete
 - Marcarlos como descartables, posteriormente la red, en caso de congestión, puede descartas
 - Reclasificar a otro servicio menos prioritario

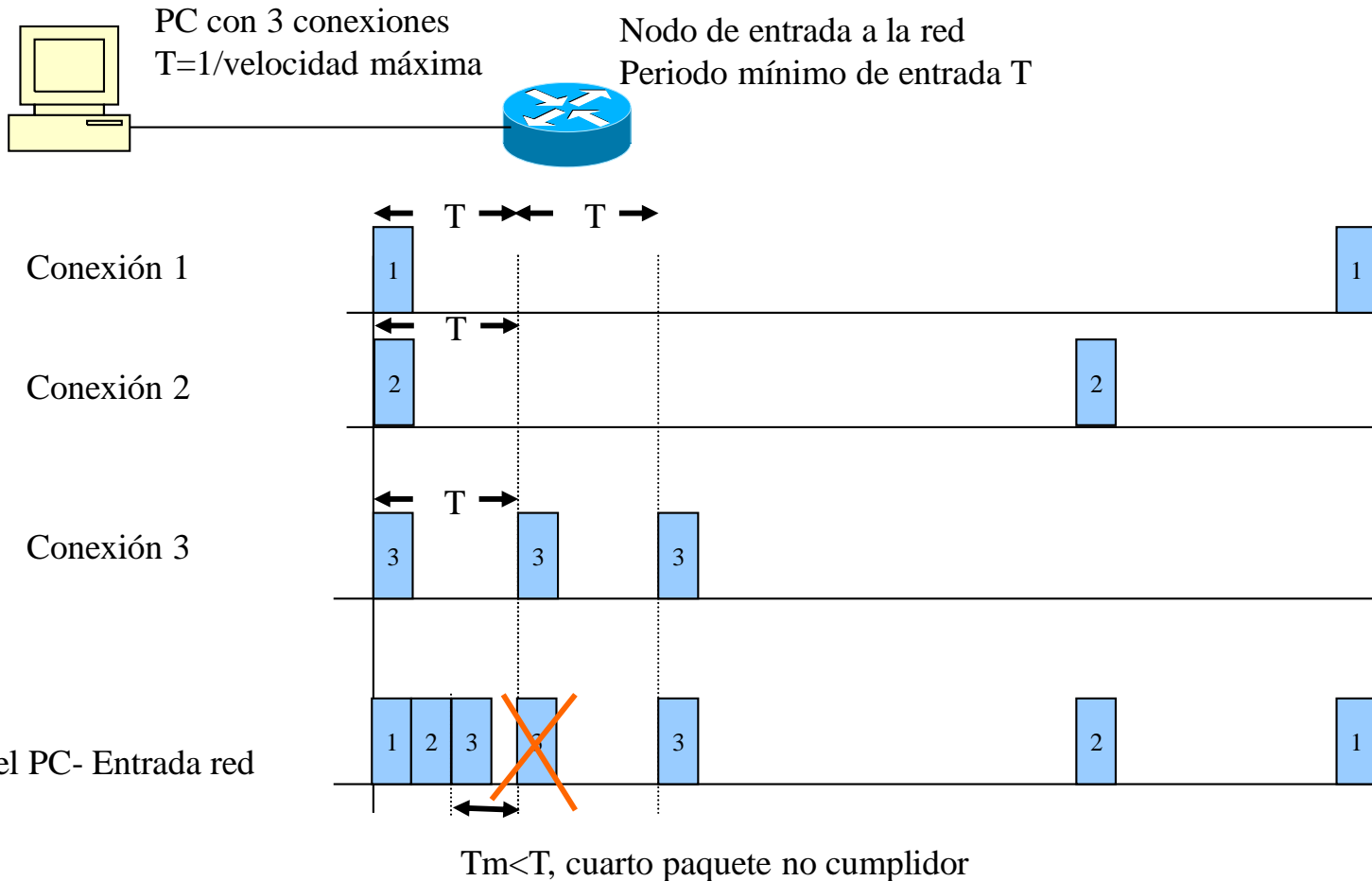
Función policía (police)

- ◆ Esta es la forma que tiene la red de comprobar el cumplimiento del contrato por parte de los usuarios
- ◆ Se pueden usar para controlar la tasa máxima o media de un emisor (o ambas con dos etapas en cascada)
- ◆ Este algoritmo debe trabajar a la velocidad de la línea para no introducir retardo y **NO modifica** el patrón del tráfico cumplidor
- ◆ Similar a un radar en una autopista

Tolerancia en la medición

- ♦ El trabajo de la función policía, consiste en comprobar que los paquetes no se reciben más juntos del periodo de la máxima velocidad $T=1/V_{\max}$
 - Si llegan más juntos implica que se están transmitiendo a mas velocidad de la máxima
- ♦ Un tráfico conformado correctamente por una aplicación, puede ser eliminado en el nodo de entrada a la red por la función policía
- ♦ Esto es debido al funcionamiento de la torre de protocolos del ordenador del usuario
 - Si varias conexiones conformadas en el nivel de aplicación se multiplexan en el nivel físico (figura siguiente transparencia)
 - También al insertar los paquetes en una portadora pueden sufrir un desplazamiento temporal que cambia el conformado

Tolerancia en la medición



Tolerancia en la medición

- ♦ Para evitar este problema se hace la medición del periodo mínimo con una tolerancia
- ♦ La tolerancia puede expresarse como
 1. Cantidad en tiempo (L) que indica el adelanto máximo con el que se puede presentar un mensaje
 2. Cantidad de tráfico extra (σ_i) (ie. el tiempo que lleva ser transmitido) que se transmite como cumplidor

Tolerancia en la medición

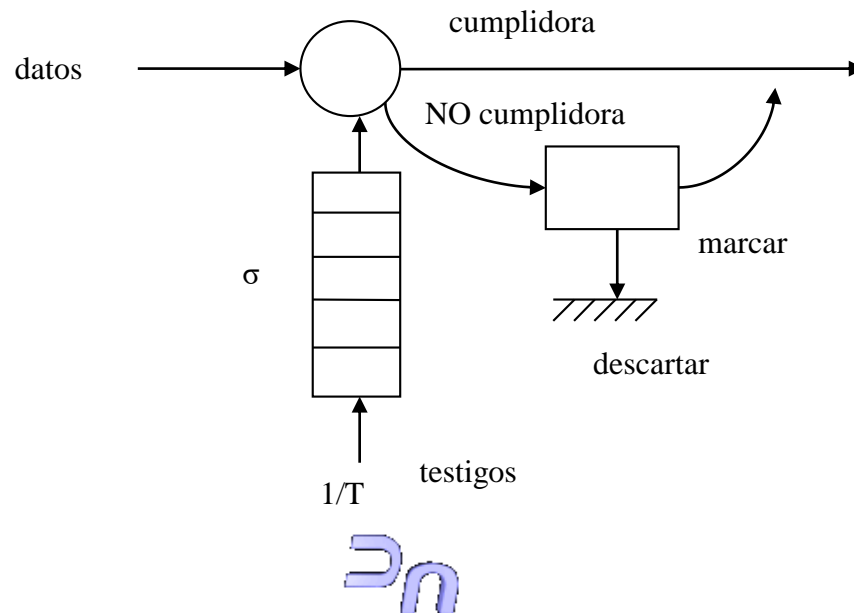
- ♦ Esta tolerancia aplicada de forma aislada para vigilar cada paquete, implicaría que de forma permanente cada paquete es válido si se recibe con un adelanto T (cada $T-L$), con lo que se supera la V_{max}
- ♦ Para evitarlo se hace una medición global que va acumulando los adelantos consecutivos
- ♦ La tolerancia puede considerarse como:
 1. El adelanto máximo acumulado con el que se puede presentar un mensaje
 2. Ráfaga de tráfico extra que se transmite como cumplidora

Algoritmo del cubo agujereado (Leaky Bucket)

- ♦ Hay varios algoritmos para implementar la vigilancia
- ♦ En este cada paquete que entra es líquido que cae en el cubo
- ♦ El cubo con capacidad σ tiene un agujero y se drena con el caudal $(1/T)$ a vigilar
- ♦ Si no boza, el paquete pasa, en caso contrario, el paquete sufre la acción policía
- ♦ De esta forma se miden los adelantos acumulados (tolerancia)
- ♦ Algunos autores consideran al Leaky Bucket un caso particular del Token Bucket con $Q=0$

Cubo agujereado (Cont.)

- ♦ El modelo para tráfico es el siguiente
- ♦ Un paquete debe obtener un testigo de la cola de testigos, antes de entrar en la red
- ♦ Una vez que entra el paquete, el testigo es eliminado

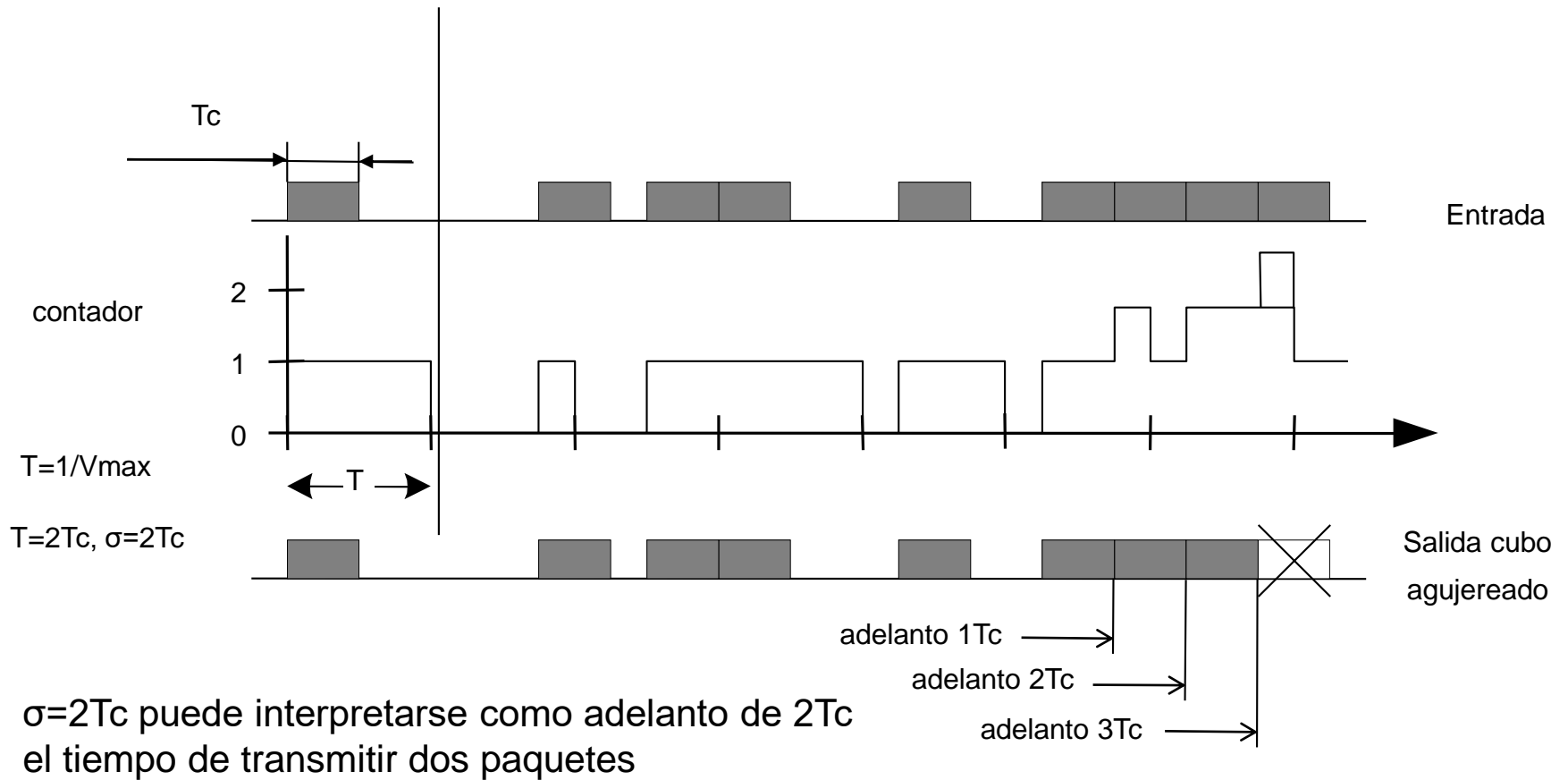


Cubo agujereado (Cont.)

- ♦ La implementación del cubo agujereado se hace con un registro contador con valor nunca inferior a cero que:
 - Se incrementa +1 cada vez que llega un paquete (Paq)
 - Se decrementa -1 cada T , periodo a vigilar
 - Al superar el límite superior σ , (tolerancia, adelanto acumulado)
 - Indica que el paquete es no cumplidor
 - No debe tocarse el valor del contador, al no ser un paquete válido no cuenta
- ♦ Recordar

- Registro (número positivo) $+1(\text{Paq})$
 $-1(T)$
 $\leq \sigma$

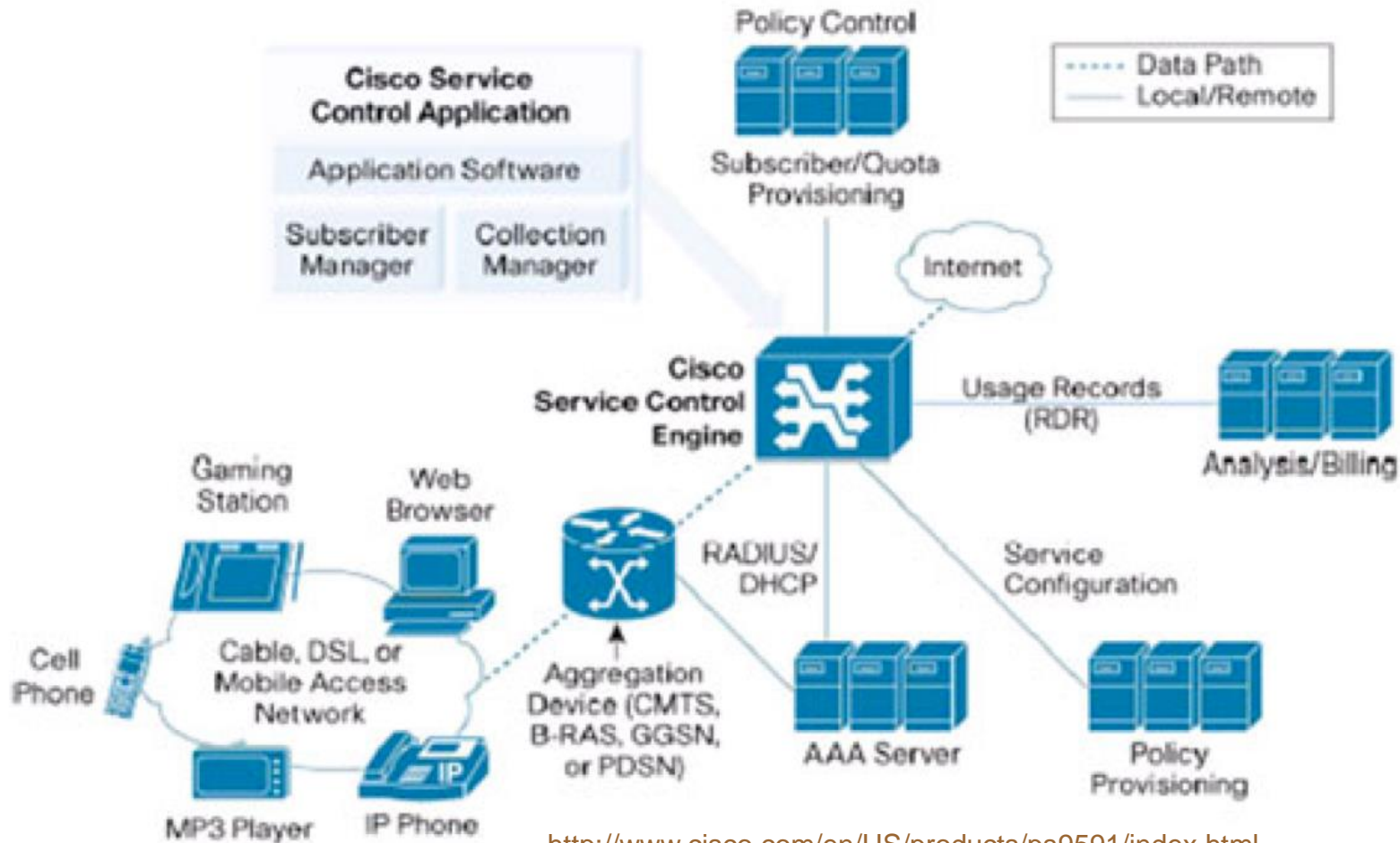
Cubo agujereado, ejemplo



Vigilancia realizada por operadores

- ◆ En el acceso doméstico (cableado) limitan la velocidad
 - Implementado en los equipos de las centrales locales
- ◆ En móviles limitan la cantidad de tráfico mensual, si se supera
 - Bajan la velocidad hasta un segundo límite o
 - Cobran por tráfico extra
- ◆ Se implementa con controlador externo, que cambia los parámetros de las sesiones en línea
- ◆ Implica una base de datos para seguir el tráfico de cada usuario
- ◆ Implementaciones
 - SCE (Service Control Engine) de Cisco
 - Resource Control Series (SRC Series) de Juniper

Cisco Service Control Engine



<http://www.cisco.com/en/US/products/ps9591/index.html>

Referencias

- ♦ Para ver las diferencias entre el conformado y la función policía, práctica 2 pg 7
- ♦ J. Chao, X. Guo, "Quality of Service Control in High-Speed Networks" Editorial Wiley. 2002, Capítulos 2, 3, 8