

William Stallings

Comunicaciones y Redes de Computadores

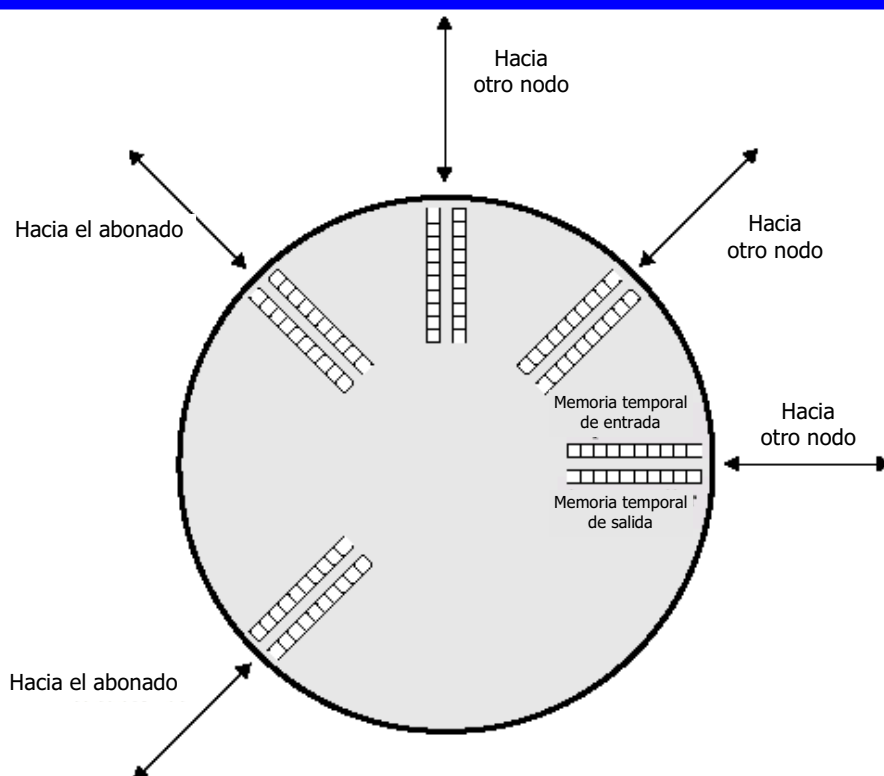
Capítulo 12

Congestión en redes de datos

¿Qué es la congestión?

- La congestión es lo que ocurre cuando los paquetes enviados por la red se aproximan al límite de esta.
- El control de la congestión intenta mantener el número de paquetes por debajo del nivel para el que decaen dramáticamente las prestaciones.
- La red de datos es una red de colas.
- Generalmente, utilizar el 80% de la capacidad es excesivo.
- La capacidad limitada de la cola significa que los datos se pueden perder.

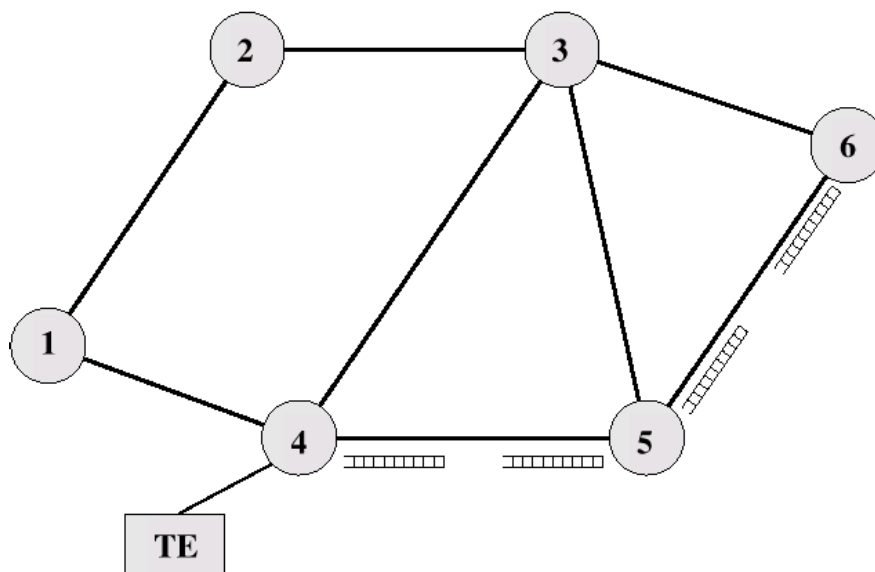
Colas de un nodo



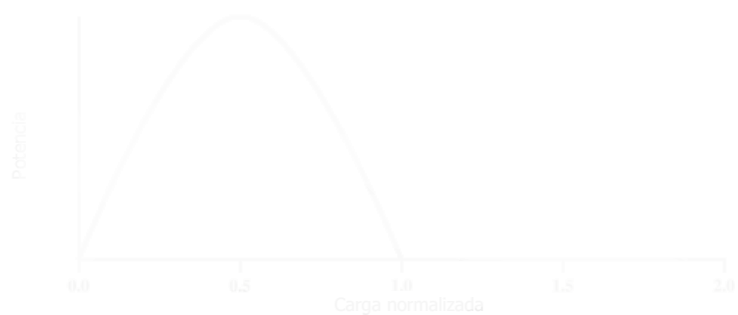
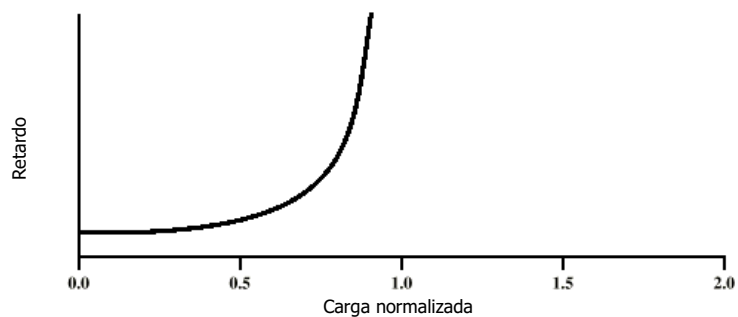
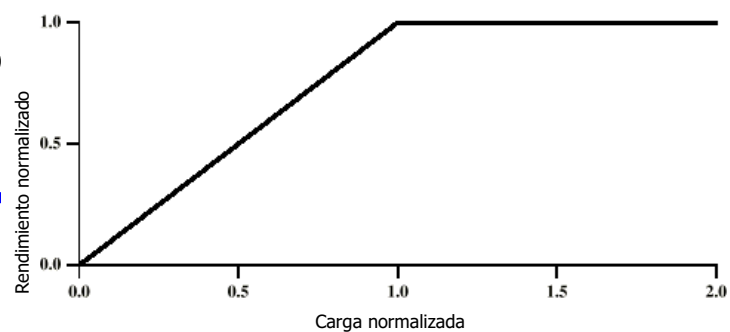
Efectos de la congestión

- Los paquetes se almacenan en la memoria temporal de entrada.
- Se selecciona la ruta .
- El paquete pasa a la memoria temporal de salida.
- Los paquetes en la cola se transmiten lo más rápido posible:
 - Multiplexación por división en el tiempo estadística.
- Si los paquetes llegan demasiado rápido las memorias temporales se llenan.
- Los paquetes pueden ser rechazados.
- Se puede usar el control de flujo:
 - La congestión puede propagarse por la red.

Interacción de las colas



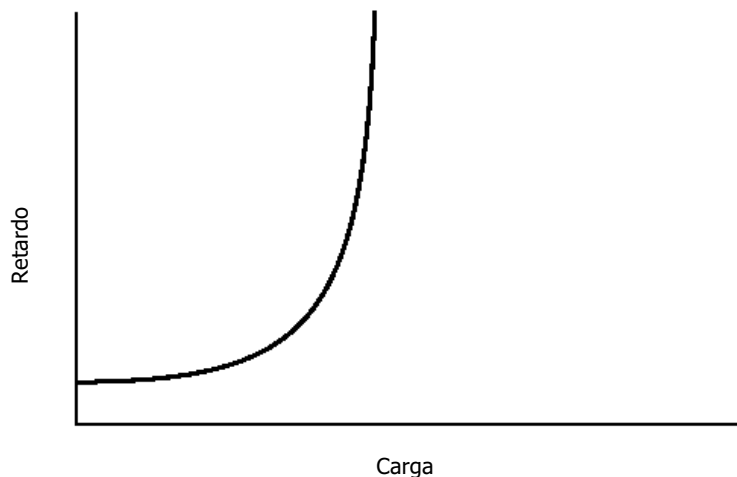
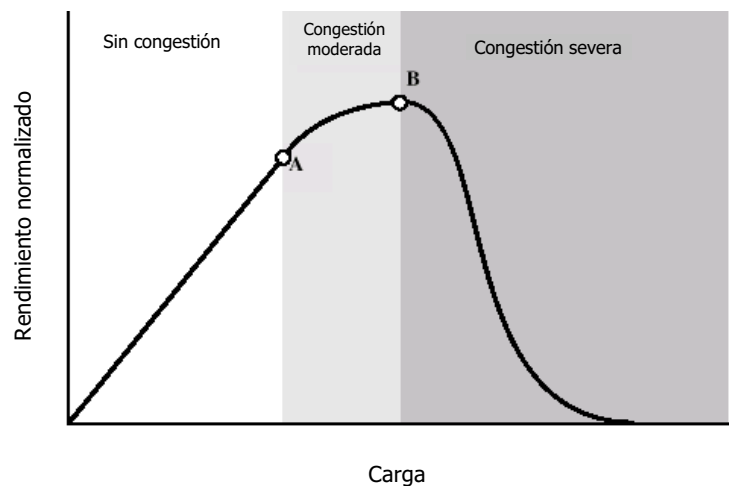
Funcionamiento ideal



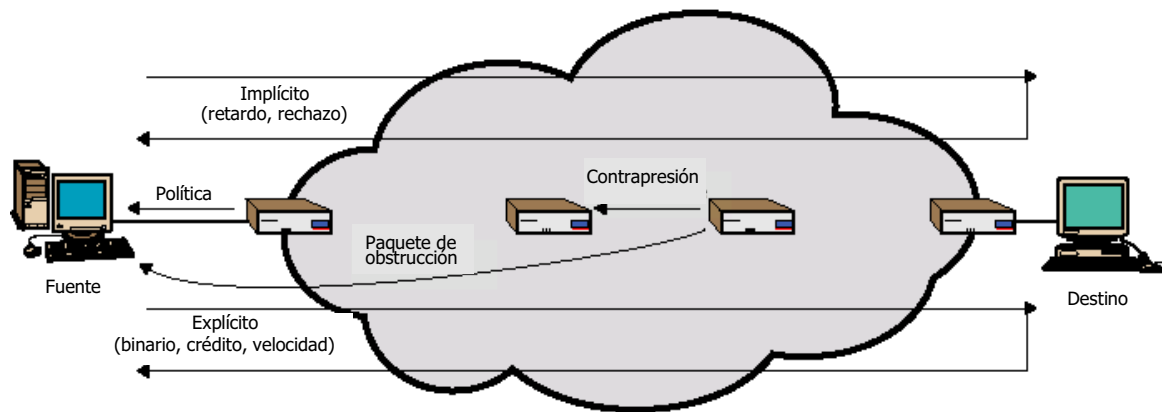
Funcionamiento real

- El funcionamiento ideal supone infinitas memorias temporales y ningún coste asociado a la transmisión de los paquetes ni al control de congestión.
- Las memorias temporales son limitadas.
- El control de congestión consume capacidad de la red debido al intercambio de señales de control.

Efectos de la congestión



Mecanismos de control de congestión



Contrapresión

- Si el nodo se congestiona puede frenar total o parcialmente el flujo de paquetes.
- También puede tener que llegar a frenar el tráfico sobre sus líneas de entrada.
- El bloqueo se propaga hacia atrás en dirección a los sistemas emisores.
- Se puede aplicar de forma selectiva a las conexiones lógicas, generalmente para aquellas con más tráfico.
- Se emplea en redes orientadas a conexión que permiten control de flujo a nivel de enlace (por ejemplo: X.25)
- No se emplea en redes de retransmisión de tramas ni en redes ATM.
- Últimamente se ha probado en redes internet IP.

Paquetes de obstrucción

- Paquete de control:
 - | Generado por un nodo congestionado.
 - | Transmitido hacia un nodo origen.
 - | Ejemplo: paquete Ralentización del Emisor usado en ICMP:
 - | Tanto desde un dispositivo de encaminamiento como desde un sistema final destino.
 - | El emisor frena los envíos hasta que no reciba más mensajes de ralentización.
 - | Mensaje enviado para cada uno de los paquetes rechazados o anticipados.
- Técnica relativamente burda.

Señalización implícita de congestión

- El retardo de transmisión puede aumentar con la congestión.
- Se rechazan paquetes.
- El origen puede detectar esto como una indicación implícita de congestión.
- Efectiva para configuraciones no orientadas a conexión (datagrama):
 - | Ejemplo: redes internet IP.
 - | TCP incluye mecanismos de control de congestión y de flujo (véase el Capítulo 17).
- Empleada en LAPF en redes de retransmisión de tramas.

Señalización explícita de congestión

- La red avisa a los sistemas finales acerca del incremento de la congestión.
- Los sistemas finales toman medidas para reducir la carga de entrada a la red.
- Hacia atrás:
 - | La prevención de congestión se aplica en el sentido opuesto al que se recibe la notificación.
- Hacia adelante:
 - | La prevención de congestión se aplica en el mismo sentido en que se recibe la notificación.

Categorías de señalización explícita

- Binarias:
 - | Un bit en un paquete de datos indica la congestión.
- Basadas en crédito:
 - | Indican cuántos paquetes de origen se pueden enviar.
 - | Se emplean en el control de flujo extremo a extremo.
- Basadas en velocidad:
 - | Proporcionan un límite explícito de velocidad.
 - | Ejemplo: ATM.

Gestión de tráfico

- Idoneidad.
- Calidad de servicio:
 - | Puede requerir tratamientos diferentes para las diversas conexiones.
- Reservas:
 - | Ejemplo: ATM.
 - | Acuerdo de tráfico entre el usuario y la red.

Control de congestión en redes de conmutación de paquetes

- Envío de un paquete de control hacia todos o algunos nodos emisores:
 - | Necesita tráfico adicional mientras dure la congestión.
- Consideración de la información de encaminamiento:
 - | Puede reaccionar muy rápidamente.
- Uso de paquetes de prueba extremo a extremo:
 - | Introduce datos suplementarios en la red.
- Añadir información de congestión a los paquetes que atraviesan los nodos:
 - | En la misma dirección o en dirección contraria a la de la congestión.

Gestión de tráfico en ATM

- Alta velocidad, pequeño tamaño de celda, reducido número de bits suplementarios.
- Campo actualmente en estado de desarrollo.
- Requisitos:
 - La mayor parte del tráfico no está sujeto a control de flujo.
 - Realimentación lenta debido al reducido tiempo de transmisión en comparación con los retardos de propagación.
 - Amplia variedad de aplicaciones.
 - Diferentes patrones de tráfico.
 - Diferentes servicios de red.
 - Las elevadas velocidades en conmutación y transmisión aumentan la inestabilidad.

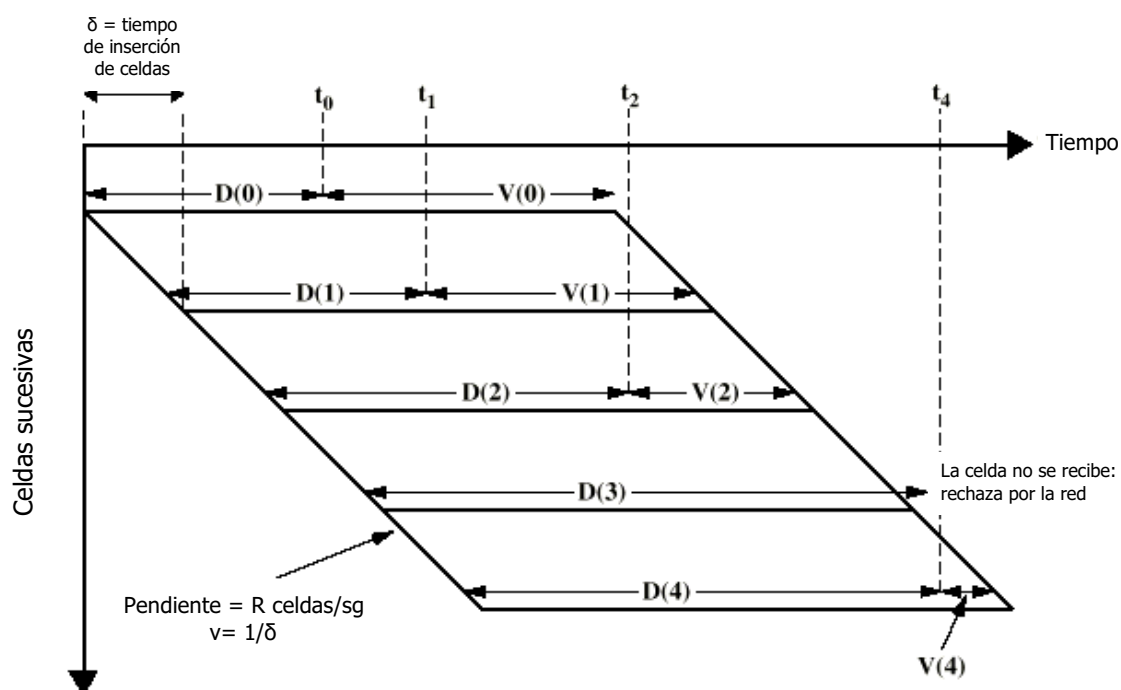
Efectos de latencia/velocidad

- Transferencia de celdas ATM a 150 Mbps.
- $2,8 \times 10^{-6}$ segundos para insertar una sola celda.
- El tiempo para atravesar la red depende de los retardos de propagación y de conmutadores.
- Supongamos una propagación a una velocidad igual a dos tercios la de la luz.
- Si el origen y el destino se encuentran situados en las costas opuestas de EE.UU., el retardo de propagación será de 48×10^{-3} segundos.
- Con el control implícito de la congestión, cuando la notificación de la celda perdida alcance la fuente, ya se habrán transmitido $7,2 \times 10^6$ bits.
- Esta técnica no es adecuada para ATM.

Variación del retardo de celdas

- Para una red ATM de voz o vídeo, los datos son como una secuencia de celdas.
- El retardo en la red debe ser pequeño.
- La velocidad de envío debe ser constante.
- Inevitablemente existe alguna variabilidad en el tránsito.
- Retardo en el envío de celdas hacia el usuario. De esta manera, se puede mantener una velocidad constante de bits a la aplicación.

Tiempo de ensamblado de celdas CBR



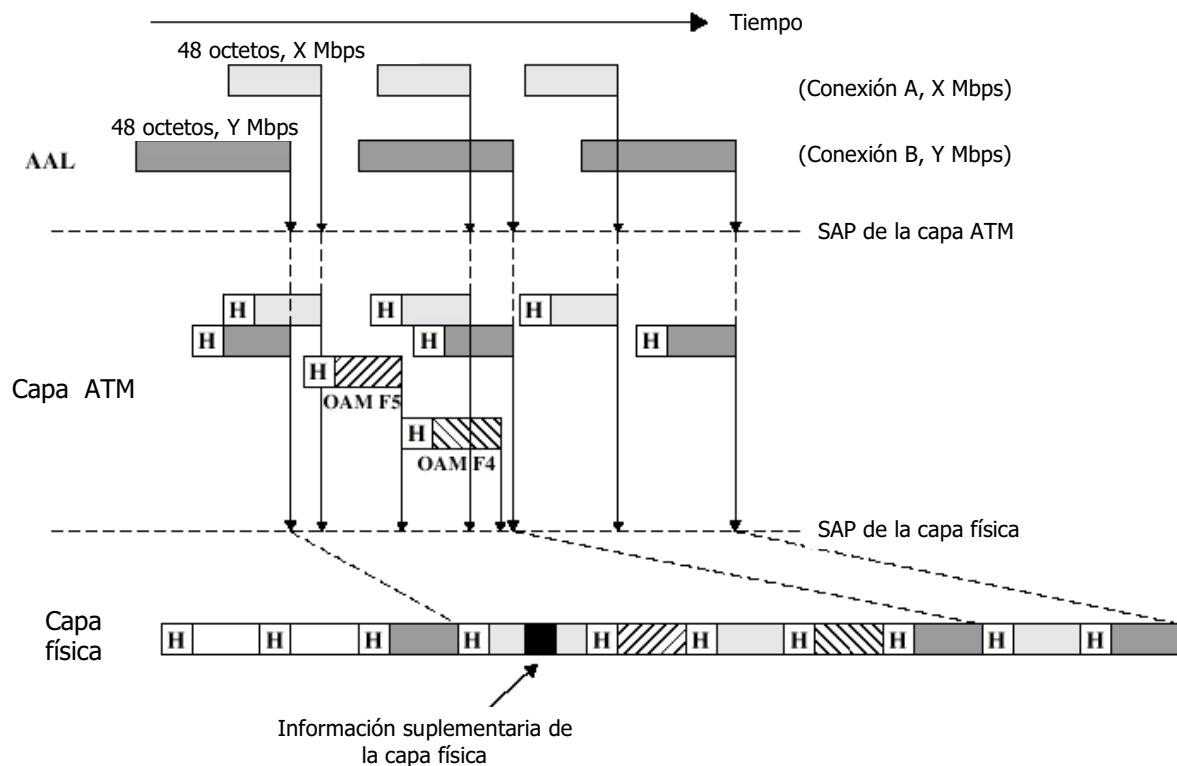
Contribución de la red a la variación del retardo de celdas

- Redes de conmutación de paquetes:
 - | Retardos producidos por la puesta en cola.
 - | Tiempo para llevar a cabo el encaminamiento.
- Retransmisión de tramas:
 - | Igual que las anteriores, pero en menor medida.
- Redes ATM:
 - | Variaciones inferiores a las de retransmisión de tramas.
 - | El protocolo ATM está diseñado para minimizar el proceso suplementario en los nodos de conmutación intermedios.
 - | Los conmutadores ATM ofrecen un rendimiento extremadamente alto.
 - | El único retardo perceptible es la congestión.
 - | Rechaza las cargas que producen congestión.

Variación del retardo de celdas en la UNI

- La aplicación transmite datos a una velocidad constante.
- El procesamiento a tres capas de ATM produce retardo:
 - | Celdas mezcladas que proceden de diferentes conexiones.
 - | Mezcla de celdas de operación y mantenimiento.
 - | Si se emplean tramas de jerarquía digital síncrona, éstas se insertan en la capa física.
 - | Estos retardos no se pueden predecir.

Orígenes de la variación del retardo de celdas



Control de tráfico y de congestión

- El control del tráfico y de la congestión en la capa ATM debería permitir un número suficiente de clases de calidad de servicio (QoS) para todos los servicios de red posibles.
- No debería depender de los protocolos AAL específicos del servicio de red, ni de protocolos de capa superior específicos de la aplicación.
- Debería minimizar la complejidad de la red y de los sistemas finales.

Niveles de tiempo

- Tiempo de inserción de celdas.
 - Tiempo de propagación de ida y vuelta.
 - Duración de la conexión.
 - Término de larga duración.
-
- Determinación de si se puede dar cabida a una nueva conexión.
 - Acuerdo con el abonado acerca de los parámetros de prestaciones.

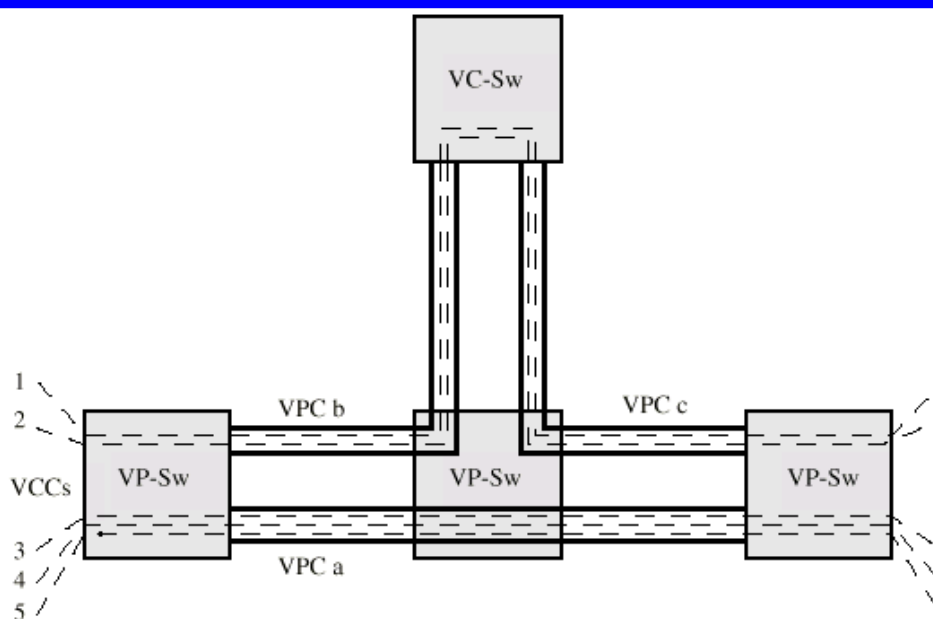
Técnicas de gestión de tráfico y de control de congestión

- Gestión de recursos haciendo uso de caminos virtuales.
- Control de admisión de conexiones.
- Control de los parámetros de uso.
- Rechazo selectivo de celdas.
- Adaptación del tráfico.

Gestión de recursos haciendo uso de caminos virtuales

- Separación de los flujos de tráfico según las características del servicio.
 - Aplicación usuario-usuario.
 - Aplicación del usuario a la red.
 - Aplicación red-red.
-
- Parámetros relacionados con la gestión de los recursos de red:
 - Tasa de pérdida de celdas.
 - Retardo de transferencia de celdas.
 - Variación del retardo de celdas.

Configuración de VCC y VPC



- VPC = conexión de camino virtual
- VCC = conexión de canal virtual
- VP-Sw = función de conmutación de camino virtual
- VC-Sw = función de conmutación de canal virtual

Diferentes VCC en la misma VPC

- Todas las VCC en la misma VPC deberían experimentar prestaciones de red similares.
- Opciones para reservar capacidad para la VPC:
 - | Demanda conjunta de pico.
 - | Multiplexación estadística.

Control de admisión de conexiones

- Primera línea de defensa.
- El usuario especifica las características de tráfico para la nueva conexión VCC o VPC seleccionando una calidad de servicio.
- La red acepta la conexión sólo si puede conseguir satisfacer la demanda.
- Contrato de tráfico:
 - | Velocidad de pico de celdas.
 - | Variación del retardo de celdas.
 - | Velocidad sostenible de celdas.
 - | Tolerancia a ráfagas.

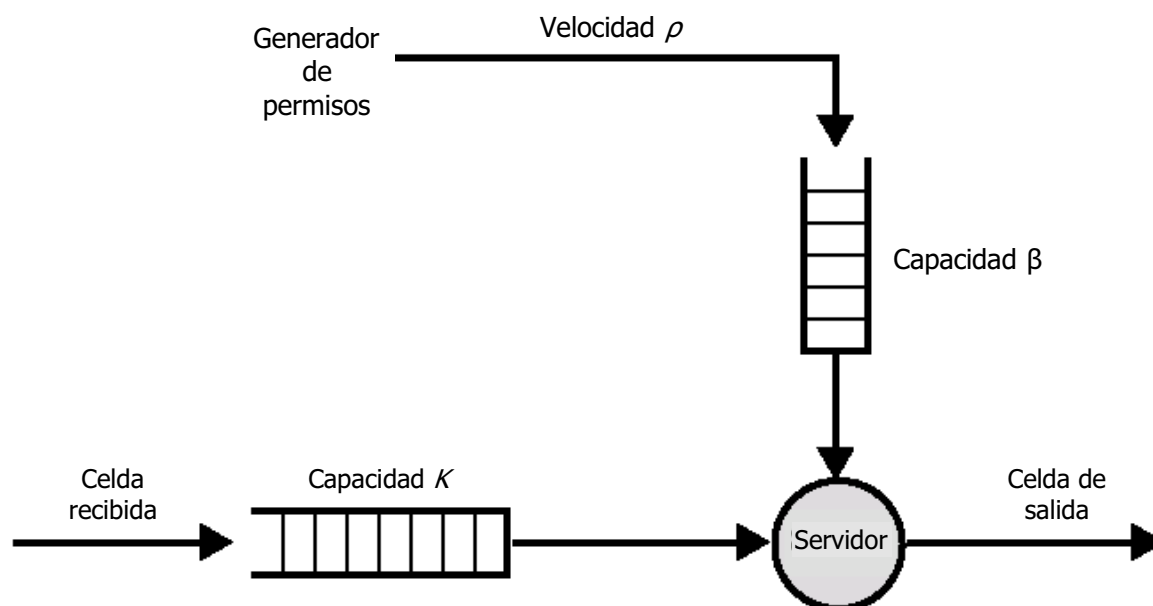
Control de los parámetros de uso

- Conexión supervisada para determinar si el tráfico está en concordancia con el contrato.
- Protección de los recursos de la red ante una sobrecarga en una conexión.
- Se emplea en VCC y VPC.
- Velocidad de pico de celdas y variación del retardo de celdas.
- Velocidad sostenible de celdas y tolerancia a la aparición de ráfagas.
- Rechazo de celdas que no cumplen con el contrato de tráfico.
- Política de tráfico.

Adaptación del tráfico

- Suaviza el flujo de tráfico y reduce la agrupación de celdas.
- Cubo de permisos.

Cubo de permisos



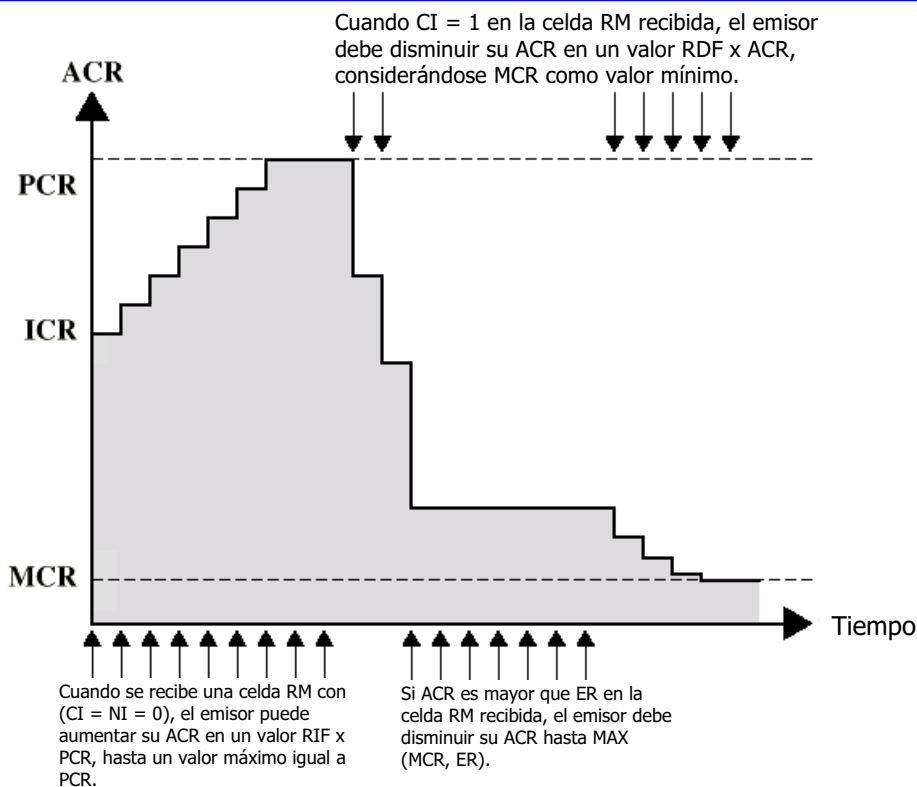
Gestión de tráfico ABR en ATM

- Algunas aplicaciones (acceso a Web, transferencia de ficheros) no tienen características de tráfico bien definidas.
- Mínimo esfuerzo:
 - Permite que las aplicaciones compartan la capacidad no utilizada.
 - Si la congestión aumenta, se pierden celdas.
- Control en bucle cerrado:
 - Las conexiones ABR comparten la capacidad disponible.
 - El compartimiento varía entre una velocidad de celdas mínima acordada (MCR) y la velocidad de pico de celdas (PCR).
 - El flujo ABR está limitado a la capacidad disponible por realimentación.
 - Las memorias temporales absorben el exceso de tráfico durante el retardo de la realimentación.
 - Baja tasa de pérdida de celdas.

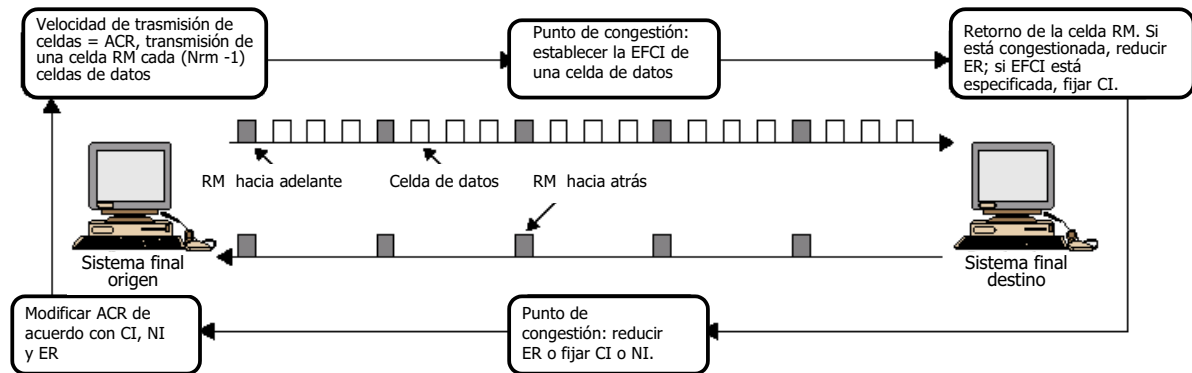
Mecanismos de realimentación

- Características de la velocidad de transmisión:
 - | Velocidad de celdas permitida.
 - | Velocidad de celdas mínima.
 - | Velocidad de pico de celdas.
 - | Velocidad de celdas inicial.
- Comienza con $ACR=ICR$.
- El emisor ajusta ACR de acuerdo con la realimentación desde la red:
 - | Celdas de gestión de recursos:
 - | Bit indicador de congestión.
 - | Bit de no incremento.
 - | Campo de velocidad explícita de celdas.

Variaciones en la velocidad de celdas permitida



Flujo de datos y de celdas



Realimentación de control de velocidad

- Marca EFCI: indicación explícita de congestión hacia adelante.
- Marca de velocidad relativa.
- Marca de velocidad explícita.

Control de congestión en retransmisión de tramas

- Minimización del rechazo de celdas.
- Mantenimiento de una calidad de servicio acordada.
- Minimización de la posibilidad del monopolio de un usuario final.
- Fácil de implementar:
 - Poco coste adicional en la red o en el usuario.
- Generación de mínimo tráfico adicional.
- Distribución adecuada de los recursos.
- Limitación en la expansión de la congestión.
- Funcionamiento efectivo independientemente del flujo de tráfico.
- Impacto mínimo en otros sistemas.
- Minimización de la varianza de la calidad del servicio.

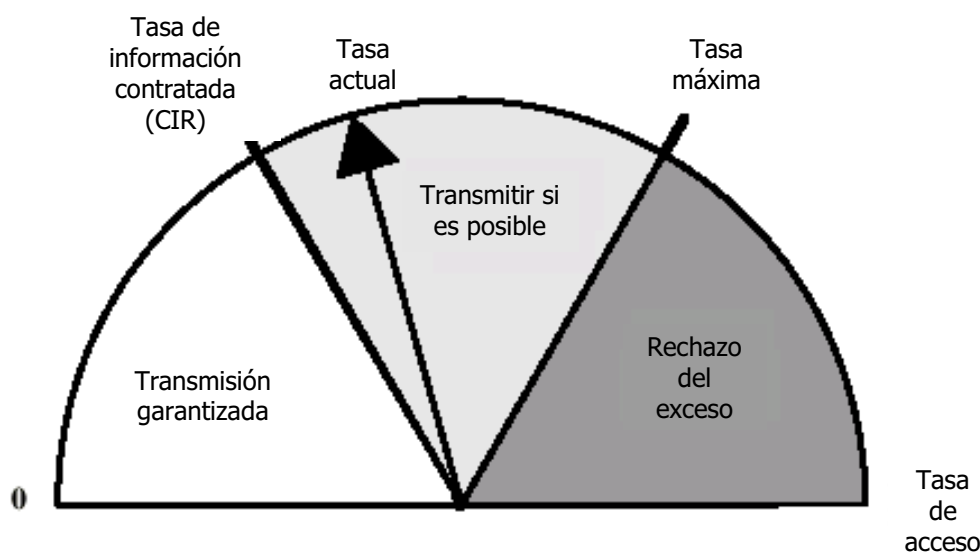
Técnicas de control de congestión

- Estrategia de rechazo.
- Prevención de congestión.
- Señalización explícita.
- Recuperación de congestión.
- Mecanismo de señalización implícita.

Gestión de la tasa de tráfico

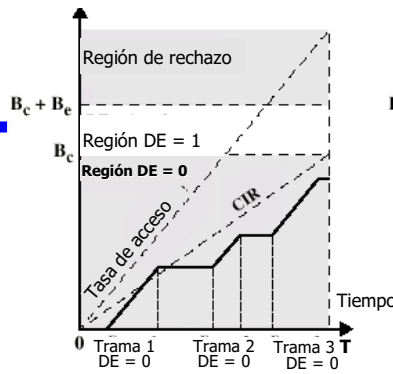
- Debe descartar tramas para combatir la congestión:
 - Arbitrariamente, independientemente del origen.
 - No importa la limitación. De esta manera, los sistemas finales transmiten tramas tan rápido como sea posible.
 - Tasa de información contratada (CIR):
 - Los datos transmitidos a velocidad superior a la CIR son susceptibles de ser rechazados.
 - No existe garantía de que se alcance la CIR.
 - La suma de las CIR no debe superar la velocidad de datos física.
- Tamaño de ráfaga contratado.
- Tamaño de ráfaga en exceso.

Funcionamiento de la CIR

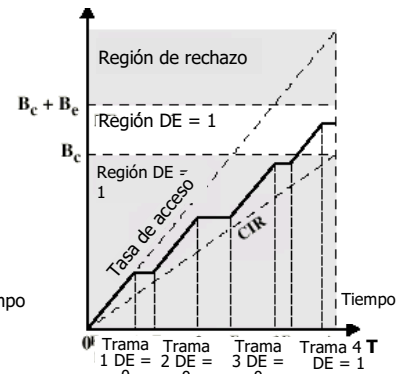


Relaciones entre los parámetros de congestión

Número de bits transmitidos

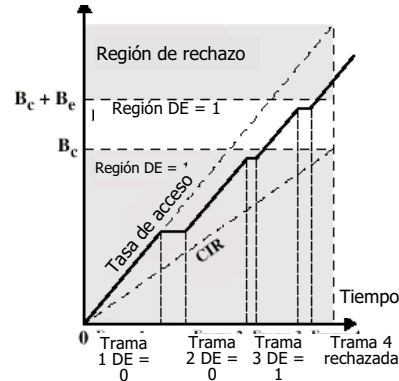


Número de bits transmitidos



(a) Todas las tramas respetan la CIR (b) Una trama marcada con DE

Número de bits transmitidos



(c) Una trama marcada con DE y otra rechazada

Señalización explícita

- La red alerta a los sistemas finales acerca del aumento de la congestión.
- Notificación explícita de congestión hacia atrás.
- Notificación explícita de congestión hacia delante.
- El gestor de tramas supervisa el comportamiento de sus colas.
- Puede notificar a algunas o a todas las conexiones lógicas.
- Respuesta del usuario:
 - Reduce la velocidad.

Lecturas recomendadas

- Stallings, W. *Comunicaciones y Redes de Computadores*, sexta edición. Madrid: Prentice Hall, 2000: Capítulo 12.