

**CAPÍTULO IV:****SIMULACIÓN EN EL ANÁLISIS DE CAPACIDAD DE SISTEMAS DE COMUNICACIONES****4.1 Capacidad de sistemas de comunicaciones y sus análisis basado en la Teoría de Colas**

**4.1.1 Introducción.** Se tratará el concepto de red en el ámbito general así como sus parámetros más importantes, tales como la cadencia eficaz, el retardo de tránsito, la tasa de fallos o la disponibilidad.

**4.1.2 Concepto.** Una red es un conjunto de medios para proporcionar servicios de telecomunicación entre cierto número de ubicaciones. Una ubicación (fija o móvil) es conocida como punto de terminación de red o simplemente "ptr". Así pues, se podría ver una red como algo abstracto que ofrece un determinado servicio en puntos de terminación de red.

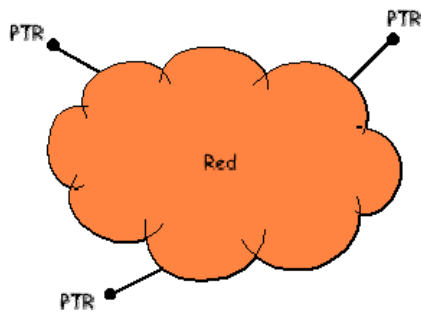


Figura 4.1 Red de comunicaciones

Dentro de esta especie de 'nube' que se acaba de dibujar existen normalmente recursos de transmisión y recursos de conmutación. Los recursos de transmisión más utilizados son los de tipo punto a punto dedicados y la conmutación se produce en nodos.

Asociado a una red hay un operador, nombre que recibe quien gestiona u opera la red; es el encargado de reparar, extraer medidas, mantener la red, y a veces sacar un beneficio económico por la explotación de los servicios.

**4.1.3 Parámetros.** Los parámetros más importantes que caracterizan una red son:

**a) Cadencia Efectiva (Cef).** También denominado caudal. Es la cantidad de bps (bits por segundo) que se pueden introducir a la red en el punto de terminación de red (ptr), es decir, el ritmo al cual la red acepta información. La definición sólo habla de lo que ocurre en un extremo de la red, y no de la cantidad de bits que van de un extremo a otro

de la red en un segundo. Por tanto, es importante no confundir que la cadencia sea 9.600 bps con que 9.600 bits atraviesen la red en un segundo.

Además, es necesario señalar que la capacidad nominal del enlace ( $C$ ) y la cadencia no son lo mismo. ' $C$ ' es toda la capacidad que brinda el enlace y como hay recursos compartidos en la red (enlaces y nodos), ocurre que  $C_{ef} < C$ .  $C_{ef}$  no es un valor determinista, puesto que depende del estado de la red, y por tanto es muy difícil de predecir.

**b) El Retardo de tránsito ( $T$ ).** Es el tiempo que transcurre desde que la red recoge un bit en el punto de terminación de red origen hasta que se recibe en el ptr destino. Este tiempo  $T$  siempre será mayor que el tiempo de propagación de la señal. En principio  $C_{ef}$  y  $T$  son dos magnitudes independientes; una tubería puede ser ancha y corta ( $C_{ef}$  alto y  $T$  bajo), o larga y estrecha ( $C_{ef}$  bajo y  $T$  alto). Suele ocurrir que si  $T$  es alto fuerza a que  $C_{ef}$  sea pequeño, por las razones que se exponen a continuación.

Al producto  $C_{ef} * T$  se le llama Memoria de la red, y expresa el número de bits en tránsito, pues es la cantidad de información que ha salido del origen, pero no ha llegado a destino, luego está en la red.

Para saber si un bit ha llegado bien y no hay que retransmitirlo, es necesario esperar un tiempo  $2 * T$  (o bien  $T_1 + T_2$  si los trayectos son asimétricos o también conocido como Retardo de ida y vuelta, luego en el origen se han de almacenar al menos  $C_{ef} * 2T$  bits para el caso en que se haya de retransmitir, y esto implica un uso de memoria muy grande si  $T$  y  $C_{ef}$  son muy altos.

**c) La tasa de fallos.** Se caracteriza por medio de la **Probabilidad de Error** en bit ( $P_e$ ), esto es, la probabilidad de que un bit no llegue correctamente a su destino. Los fallos pueden ser debidos a pérdidas, corrupción, duplicación y desórdenes en bits o paquetes. Muchos de éstos son debidos a que el software de comunicaciones no puede responder ante todas las situaciones posibles, pues suele trabajar sobre complejos sistemas distribuidos.

El uso de códigos reduce la tasa de fallos, pero no puede hacer nada si el sistema está indisponible, por ejemplo, si se caen los enlaces que conectan un nodo con el resto, dicho nodo está incomunicado.

**d) La Disponibilidad del Servicio.** Viene determinada por el tanto por ciento del tiempo en que el servicio está funcionando (disponible). 100% es el límite ideal al que se debe intentar llegar.

**e) La Cobertura.** Corresponde al área de alcance del servicio que proporciona la red. Es, en otras palabras, una enumeración de los puntos de terminación de red, dónde es posible usar la red.

**4.1.4 Clasificación.** Son muchos los criterios en función de los cuales se pueden clasificar las redes. En este apartado se pondrá atención sólo en dos de ellos, que son:

**a) Su objetivo empresarial.** Las redes pueden ser **privadas** o **públicas**. Las primeras no buscan un beneficio económico sino una mejora en las herramientas de trabajo de una determinada empresa u organismo, mientras que las segundas sí persiguen ese beneficio. En otras palabras, sólo los usuarios de redes públicas se ven obligados a pagar por su utilización.

**b) Su cobertura.** En virtud de la cobertura se diferencian tres categorías:

**LAN (Local Area Network):** Denominadas en castellano redes de área local. Pueden abarcar una distancia de unos pocos metros (entorno de una habitación) o hasta cubrir un edificio, o como máximo unos pocos edificios cercanos entre sí (por ejemplo, el entorno de un campus universitario).

**WAN (Wide Area Network):** Es la red de mayor cobertura, llegando a cubrir el área de todo un país, un continente o incluso más.

**MAN (Metropolitan Area Network):** Abarcan un área intermedia entre las LAN y las WAN. Se habla por tanto de ciudades como cobertura.

## 4.2 Modelo de red de conmutación

**4.2.1 Introducción** En la sección anterior se han introducido los conceptos de red. En esta segunda parte se verá con algo de detalle las redes de conmutación y algunos modelos matemáticos que permitirán, jugando con sus parámetros de diseño, conducir su comportamiento hacia optimizaciones concretas.

Para simplificar el modelo se supone la existencia de mensajes, cuya longitud ( $b$ ) se modela mediante una variable aleatoria exponencial, como se vio en el capítulo 3, de parámetro ' $\mu$ ', cuya pdf. es:

$$\left. \begin{aligned} f(b) &= \mu \cdot e^{-\mu b} \cdot u(b) \\ F(b) &= (1 - e^{-\mu b}) \cdot u(b) \\ u(b) &= \begin{cases} 0, & b < 0 \\ 1, & b > 0 \end{cases} \end{aligned} \right\} \rightarrow \begin{cases} \text{media: } \bar{b} = 1/\mu & (\text{longitud media de mensajes})(\text{bits / mensaje}) \\ \text{varianza: } \sigma_b^2 = 1/\mu^2 \end{cases}$$

Así pues, nuestro punto de partida es una red de conmutación de mensajes. A cada nodo llegarán mensajes desde el exterior de la red; la tarea de la red será cursarlos,

es decir, hacerlos llegar a su destino; origen y destino serán nodos de la red, surgiendo así los conceptos de nodo origen y nodo destino. Se modela la llegada de mensajes exteriores con un determinado par origen y destino como procesos de Poisson. Los mensajes pasarán por nodos intermedios en su camino hacia el destino. Salvo que se diga lo contrario, la red dispondrá de un algoritmo de encaminamiento fijo, es decir, utilizará siempre un mismo camino entre cada par de nodos. Como los nodos pueden estar interconectados con otros nodos, y además puede existir *multiplexación* en los canales, estos *centros de conmutación* pueden recibir y procesar varios mensajes concurrentemente. Cuando un mensaje llega a un nodo, provenga de donde provenga, el nodo lo procesa (deduce por dónde tiene que reenviarlo) idealmente en tiempo cero, y entonces lo mete en la cola de salida correspondiente al enlace por el que va a salir el mensaje. En un nodo hay tantas colas como posibles canales de salida. La capacidad del servidor coincide con la capacidad del enlace de salida del nodo. A este procedimiento se le denomina almacenamiento y reenvío en los nodos.

**4.2.2 Parámetros de modelación.** Los parámetros que modelan el comportamiento de la red que se han descrito son los siguientes:

**$N$**  **Número de nodos de la red.** Se suponen etiquetados desde 1 hasta  $N$ .

**$M$**  **Número de canales de la red.** Están etiquetados desde 1 hasta  $M$ . Se dirá que se conoce la topología de la red si se conoce el número de nodos y su localización, así como el número de canales que unen los nodos entre sí.

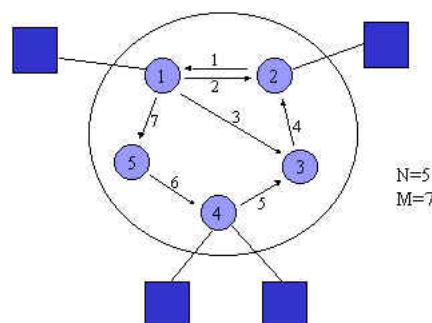


Figura 4.2 Nodos y canales de una red.

**$C_i$**  **Capacidad del canal  $i$ ,** es decir, valor máximo de la velocidad de transmisión de bits en el mismo, medida en bits por segundo (bps), o sus múltiplos, como por ejemplo: Kbps (1 Kbps = 1000 bps) y Mbps (1 Mbps =  $10^6$  bps). Los valores posibles de estas capacidades están determinados por las jerarquías digitales síncronas utilizadas en un

momento dado (en la actualidad, 9600 bps, 19200 bps, 64 Kbps, 2 Mbps, 34 Mbps, 155 Mbps), y varían conforme surgen nuevas tecnologías y se abandonan otras.

**$\gamma_{jk}$**  Valor medio de la tasa de llegada de mensajes *externos* con origen en el nodo  $j$  y destino el nodo  $k$ . Se mide en mensajes por segundo (mensajes/s). También le se llama tráfico externo, tráfico a cursar o tráfico ofrecido con origen en  $j$  y destino  $k$ . Se modelan estas llegadas como procesos de Poisson de parámetros  $\gamma_{jk}$ . Se llaman *mensajes externos* porque los nodos también reciben mensajes provenientes de otros nodos y no de sistemas finales; en este caso, el nodo sirve de paso intermedio para conseguir el encaminamiento de los mensajes desde el origen al destino. Si se particularizan la tasa de llegadas para cada par  $(j, k)$ , se pueden recoger estos valores en una matriz cuadrada de orden  $N$ , cada uno en la fila  $j$ -ésima y la columna  $k$ -ésima correspondientes.

**$\gamma$**  Tasa media total de llegada de mensajes desde el *exterior* de la red, medida en mensajes/s. También se llama tráfico externo total, tráfico total a cursar o tráfico total ofrecido a la red. Será igual a la suma de las tasas para cada par de nodos origen-destino:

$$\gamma = \sum_{j=1}^N \sum_{k=1}^N \gamma_{jk} \quad (4.1)$$

$$\frac{1}{\mu}$$

Valor medio de la longitud de los mensajes, que se expresa en bits. Será la media de una variable aleatoria exponencial.

**$\lambda_i$**  Valor medio de la tasa de mensajes que *atraviesan* el canal  $i$ , expresado en mensajes/s. También se conoce como flujo de mensajes o tráfico interno en el canal  $i$ . Este flujo incluye mensajes generados por los sistemas finales conectados a los nodos de entrada al canal  $i$ , así como otros almacenados y reenviados por éstos y provenientes de otros nodos. En ocasiones se hablará de flujo de bits en el canal  $i$ , en cuyo caso se

refiere a la cantidad  $\lambda_i / \mu$ , expresada en bps.

**$\lambda$**  Valor medio de la tasa total de mensajes que *atraviesan* todos los canales, expresado en mensajes/s. También se conoce como flujo total de mensajes o tráfico interno total. Será la suma de sus componentes para cada canal:

$$\lambda = \sum_{i=1}^M \lambda_i \quad (4.2)$$

**$T_i$**  Retardo medio asociado a la estancia en el canal  $i$ , definido como el tiempo medio que permanece un mensaje en el canal  $i$  (incluyendo la cola de entrada asociada al canal). También se llama retardo medio en el canal  $i$ . Hablando en términos del modelo de sistema M/M/1,  $T_i$  es el tiempo medio de estancia del cliente (mensaje) en el sistema (cola y servidor —canal—).

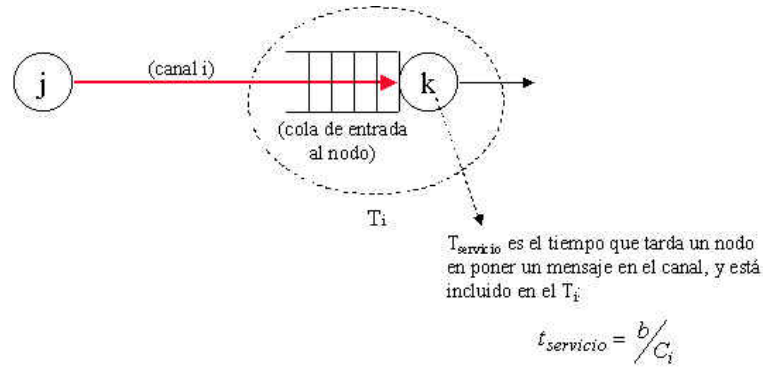


Figura 4.3 Modelo de cola M/M/1.

**$Z_{jk}$**  Retardo medio de los mensajes con origen en el nodo  $j$  y destino el nodo  $k$ , definido como el tiempo medio que tarda un bit en viajar desde el nodo  $j$  de origen, hasta el nodo  $k$  de destino.

**$T$**  Retardo medio de tránsito, definido como el valor medio del tiempo que transcurre desde que un bit de un mensaje entra en un nodo origen hasta que llega al nodo destino. No incluye el tiempo que permanece ese bit en la cola de salida del nodo destino, antes de emprender su camino hacia el sistema final, ni el retardo de los enlaces que unen los sistemas finales y la red, es decir, difiere ligeramente del retardo de tránsito. De las definiciones realizadas anteriormente se deriva la siguiente expresión:

$$T = \sum_{j=1}^N \sum_{k=1}^N \frac{\gamma_{jk}}{\gamma} Z_{jk} \quad (4.3)$$

ya que se puede ver el retardo como una variable aleatoria discreta, cuyos valores posibles son  $Z_{jk}$ , y sus probabilidades correspondientes:

$$P(Z_{jk}) = \frac{\gamma_{jk}}{\gamma} \quad (4.4)$$

Esta última expresión no es más que un cociente de la forma:

$$P(Z_{jk}) = \frac{\text{nº de casos favorables a } Z_{jk}}{\text{nº de casos totales}} \quad (4.5)$$

**4.3 Problemas de optimización.** La optimización del comportamiento de la red que se está considerando puede abordarse de distintas formas. En general, se supone que el objetivo de la optimización es reducir el retardo medio de tránsito ( $T$ ). Para ello se deben ajustar los valores de algún o algunos parámetros de la red. Los otros, en general, vendrán fijados por la especificación del problema, y se deben adaptar a ellos sin modificarlos.

Según esto, se pueden clasificar los problemas de optimización atendiendo a los siguientes elementos:

- **Parámetros fijos** Como ya se ha comentado, son aquellos que no se pueden modificar sino que vienen dados de antemano.
- **Parámetro a optimizar** En general, será el retardo medio de tránsito, en el sentido de minimizarlo.
- **Parámetros de diseño** Aquellos que se modificarán o ajustarán para conseguir la optimización.
- **Condición de contorno** Condición adicional, sobre una función de los parámetros, que delimitará los valores posibles de los mismos.

Se estudiarán dos problemas, que se conocerán como asignación de capacidades y asignación de flujos.

#### 4.4 Análisis del retardo

El objetivo de esta sección es profundizar en las relaciones matemáticas que ligan el retardo medio de tránsito ( $T$ ) con otros parámetros de la red. Esto es necesario porque el modelo que se ha introducido es demasiado general y no permite extraer directrices de diseño.

Antes de nada, conviene introducir unos elementos de notación. Se llama  $\pi_{jk}$  al camino, ruta o secuencia de canales que atraviesa un mensaje desde el nodo  $j$  hasta el nodo  $k$ , y se dice que el canal  $i$  está incluido en  $\pi_{jk}$  cuando los mensajes que siguen dicho camino atraviesan el canal  $i$ . Se simboliza esto último de la siguiente forma:

$$C_i \in \pi_{jk} \quad (4.6)$$

Asimismo, se escribe

$$i: C_i \in \pi_{jk} \quad (4.7)$$

para referirnos a los valores de  $i$  tales que el canal  $i$  está incluido en  $\pi_{jk}$  y

$$j, k: C_i \in \pi_{jk} \quad (4.8)$$

para representar los valores de  $j$  y  $k$  tales que el canal  $i$  está incluido en  $\pi_{jk}$ .

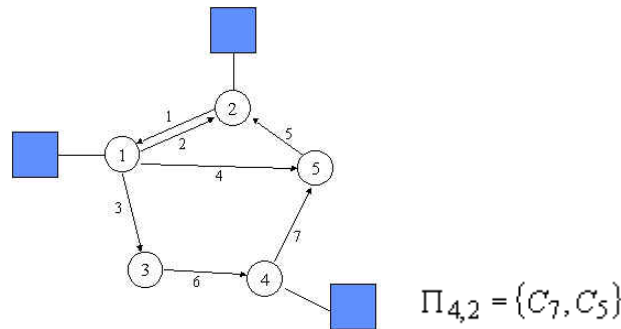


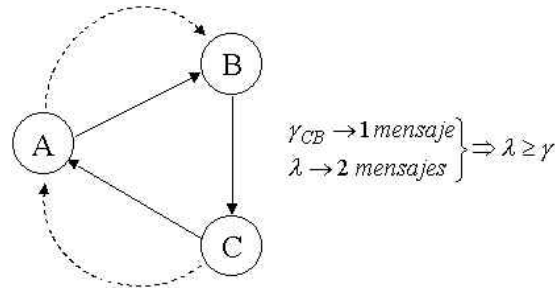
Figura 4.4 Camino o ruta de la red.

**4.4.1 Cálculo de  $T$**  En primer lugar, se supone que la red se encuentra vacía, es decir, no tiene mensajes que cursar. Si en este momento comienzan a llegar mensajes externos, la red comenzará a cursarlos siguiendo un camino fijo para cada par origen-destino. Los mensajes serán almacenados y reenviados en nodos intermedios, pero el flujo de tráfico externo no cesará. Cuando la red se encuentra en pleno funcionamiento, cada par de nodos  $(j, k)$  contribuye en media con  $\gamma_{jk}$  mensajes/s al flujo de mensajes en cada canal incluido en  $\pi_{jk}$ . Es decir, según la notación introducida en la expresión (4.8), se cumple:

$$\lambda_i = \sum_j \sum_{k: C_i \in \pi_{jk}} \gamma_{jk} \quad (4.9)$$

En el siguiente ejemplo, se pueden observar las diferencias entre  $\gamma$  y  $\lambda$ :



Figura 4.5 Diferencia entre  $\gamma$  y  $\lambda$ .

Por ejemplo, si se supone que la red se encuentra en el período permanente, y que, en estas condiciones, los mensajes que atraviesan un canal genérico  $i$  tienen origen en los nodos  $j$  y  $u$ , y por destino los nodos  $k$  y  $v$ , respectivamente, entonces por dicho canal estarán pasando, en un momento dado,  $\gamma_{jk}$  mensajes/s con destino al nodo  $k$ , y  $\gamma_{uv}$  mensajes/s con destino al nodo  $v$ . Por tanto, en total, pasarán  $\gamma_{jk} + \gamma_{uv}$  mensajes/s por el canal  $i$ .

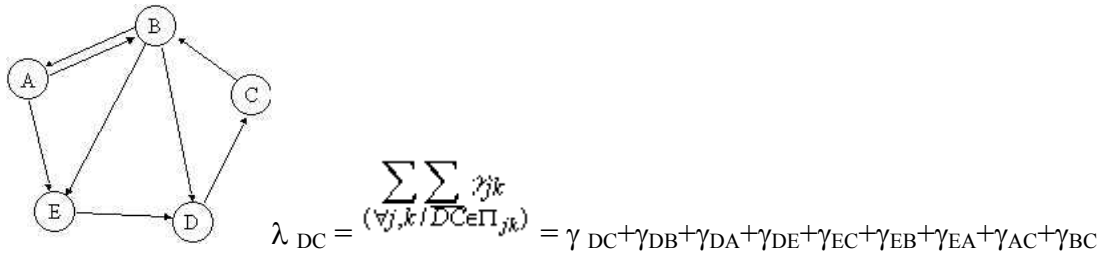


Figura 4.6 Tráfico total por un canal

En segundo lugar, el retardo medio de los mensajes con origen en el nodo  $j$  y destino el nodo  $k$  ( $Z_{jk}$ ) será la suma de los retardos medios en los canales ( $T_i$ ) por los que pasan estos mensajes, es decir, según la notación introducida en la expresión (4.7),

$$Z_{jk} = \sum_{i \in \Pi_{jk}} T_i \quad (4.10)$$

Llevando estos resultados a la expresión del retardo medio de tránsito ( $T$ ) ecuación (4.3), se obtiene:

$$T = \sum_{j=1}^N \sum_{k=1}^N \frac{\gamma_{jk}}{\gamma} \sum_{i \in \Pi_{jk}} T_i \quad (4.11)$$

Intercambiando el orden de las sumatorias, y observando que la condición sobre  $i$  se convierte en una condición sobre  $j, k$ , se obtiene:

$$T = \sum_{i=1}^M \frac{T_i}{\gamma} \sum_j \sum_{k: j,k \in C_i} \gamma_{jk} \quad (4.12)$$

Utilizando la expresión del flujo de mensajes calculada anteriormente, ecuación (4.9), se obtiene finalmente:

$$T = \sum_{i=1}^M \frac{\lambda_i}{\gamma} T_i \quad (4.13)$$

Con estas manipulaciones se ha conseguido descomponer  $T$  en sus componentes para cada canal. Sólo resta calcular  $T_i$  ( $i = 1, \dots, M$ ) para la red que se está considerando.

**4.4.2 Teorema de Jackson** En las redes abiertas de colas, que son aquéllas en las que los clientes o usuarios (en este caso los mensajes) entran por un extremo de la red (el nodo origen) y salen por otro (el nodo destino) tras recibir el servicio demandado, si la tasa de llegada a un nodo y el tiempo de servicio son variables aleatorias independientes, entonces los servidores pueden modelarse independientemente unos de otros, esto es, sin que las características estadísticas de unos influyan en las de los otros. En realidad tiene una formulación matemática más rigurosa, en términos de probabilidades, pero se mantiene con esta idea, que será suficiente en este contexto.

Desgraciadamente, en nuestra red estas variables no son independientes. En efecto, el **tiempo de servicio** de un mensaje en un canal  $i$  es

$$t_i^{serv} = \frac{b}{C_i} \quad (4.14)$$

es decir, el tiempo necesario para verter al canal los  $b$  bits del mensaje a una tasa de  $C_i$  bps. El tiempo entre llegadas al siguiente canal (a la cual se llama  $i + 1$ ), definido como el tiempo que transcurre entre la llegada de dos mensajes consecutivos, no puede ser menor que  $b/C_i$  (el tiempo necesario para verter al canal  $i$  el primero de ellos). Como los nodos conservan la longitud de los mensajes, el tiempo de servicio en el canal  $i+1$  y el patrón de llegada de mensajes a dicho servidor no son independientes, incumpliendo las condiciones del teorema de Jackson.

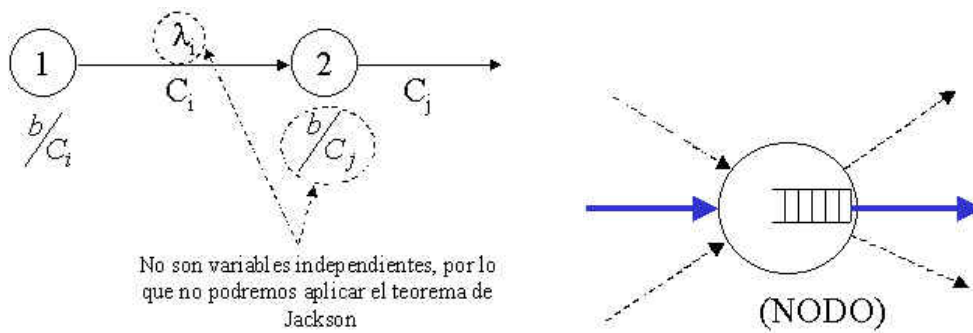


Figura 4.7 Teorema de Jackson

**4.4.3 Hipótesis de Kleinrock.-** Kleinrock estudió este problema y llegó a la conclusión de que si los mensajes llegan a un nodo provenientes de varias fuentes (varios canales de entrada y sistemas finales) y salen por varios canales de salida, la dependencia antes descrita se suaviza, hasta el punto de que puede eliminarse sin cometer mucho error. ¿Por qué? Si se mira la cola por donde va a salir un mensaje, en realidad un mensaje cualquiera no me da por sí solo información sobre el canal por el que podría haber llegado. Entonces puesto que no se conoce el camino por el que ha llegado el mensaje, daría igual suponer que los mensajes que llegan desaparecen, y se genera un mensaje nuevo en el propio nodo por cada mensaje que desaparece. Según Kleinrock, en estos sistemas, que se denominan redes de conectividad moderada, caracterizados por que cada nodo tiene varios canales de entrada y de salida, se puede suponer que, cada vez que entra un mensaje a un nodo, se genera un nuevo mensaje cuya longitud tiene una distribución exponencial independiente de la que tenía antes; este mensaje es el que sale por el canal de salida, por lo que su tiempo de servicio estará relacionado con su nueva longitud, y, por tanto, es una variable aleatoria independiente. A esta conclusión se llega mediante la observación de la cola asociada al nodo. En ella los mensajes que provienen de canales diferentes pasan a compartir un mismo recurso, perdiendo por tanto su naturaleza distinta, y suavizando la dependencia entre variables que imposibilitaba la aplicación del teorema de Jackson.

Se utiliza esta aproximación, conocida como hipótesis de Kleinrock, de modo que se podrá aplicar el teorema de Jackson y modelar los canales de la red de forma independiente. Según esto, cada canal corresponderá a un servidor aislado, con un patrón de llegadas poissoniano de media  $\lambda_i$ , y un tiempo de servicio, distribuido exponencialmente, de media  $1/\mu C_i$  (sistema M/M/1).

El retardo medio en el canal  $i$  es, para un sistema de estas características,

$$T_i = \frac{1}{\mu C_i - \lambda_i} \quad (4.15)$$

Llevando este valor a la ecuación (4.13), se obtiene definitivamente:

$$T = \sum_{i=1}^M \frac{\lambda_i}{\gamma} \frac{1}{\mu C_i - \lambda_i} \quad (4.16)$$

**4.4.4 Modelo umbral** Si se representa el retardo medio de tránsito ( $T$ ) en función del tráfico total ofrecido a la red ( $\gamma$ ), según la ecuación (4.16), se obtiene una curva con el siguiente aspecto:

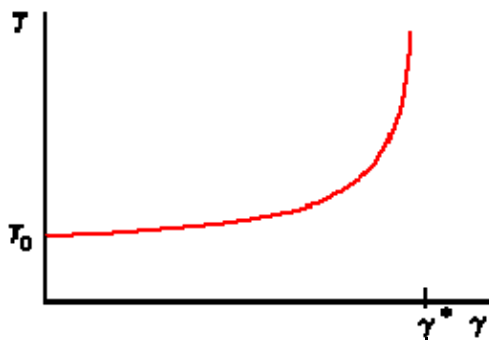


Figura 4.8 Retardo medio de tránsito en función del tráfico externo total

Como puede verse, en las proximidades de un valor umbral  $\gamma^*$ , que se llama **punto de saturación de la red** o **carga máxima**, la gráfica de  $T$  aumenta rápidamente. Ello es debido a que algún canal estará próximo a su saturación, es decir, tendrá dificultades para dar servicio a los paquetes que esperan en la cola asociada al mismo.

Por el contrario, para valores bajos de  $\gamma$ , el retardo se mantiene cercano a un valor  $T_0$ , al que se llama **retardo medio en vacío**, y que, como su propio nombre indica, es el retardo medio que sufren los mensajes cuando la red trabaja sin carga apreciable, es decir, llevando este requisito al extremo, cuando no hay otros mensajes cursándose al mismo tiempo.

Nótese que para dibujar esta gráfica es necesario conocer la topología de la red y alguna información sobre las componentes para cada par de nodos del tráfico total ofrecido a la red ( $\gamma^*$ ). Por eso, en cada caso concreto será diferente, pero siempre exhibirá el mismo aspecto.

La figura 4.8 da origen a un modelo simplificado del comportamiento de las redes de conmutación de paquetes, conocido como modelo umbral, en el que dicha figura se transforma en la figura 4.9:



Figura 4.9 Modelo umbral

En esta figura se aprecian dos tramos, a saber:

- Tramo con retardo constante, correspondiente a la zona de baja carga, cuando no existe ningún canal saturado.
- Tramo con pendiente infinita, correspondiente al estado de saturación, en el que el retardo medio se hace infinito.

Nótese, sin embargo, que la existencia de un canal saturado no implica la saturación de los demás, de modo que sólo los mensajes que atraviesan el canal saturado experimentarán retardos enormes.

Se puede ir más allá y encontrar expresiones que relacionan el retardo en vacío y el punto de saturación con otros parámetros de la red. Es lo que se hará en las siguientes secciones.

**4.4.5 Cálculo del retardo medio en vacío** En primer lugar, de la definición de  $T_0$ , se obtiene:

$$T_0 = \lim_{\lambda, \gamma \rightarrow 0} \sum_{i=1}^M \frac{\lambda_i}{\gamma} \frac{1}{\mu C_i - \lambda_i} \quad (4.17)$$

Para calcular este límite se necesita saber lo siguiente:

$$\lim_{\lambda, \gamma \rightarrow 0} \frac{\lambda_i}{\gamma} \quad (4.18)$$

que en principio presenta una indeterminación del tipo 0/0. Para resolverlo, se supone que el tráfico externo que cursa la red aumenta proporcionalmente entre cada par de

nodos. Es decir, si, por ejemplo,  $\gamma_{12}$  se duplica, entonces el resto de los  $\gamma_{jk}$  también se duplican, de modo que el tráfico total  $\gamma$  también aumenta al doble, según la siguiente ecuación:

$$\begin{aligned}\gamma(t_2) &= \sum_{j=1}^N \sum_{k=1}^N \gamma_{jk}(t_2) = \sum_{j=1}^N \sum_{k=1}^N 2 \cdot \gamma_{jk}(t_1) \\ &= 2 \cdot \sum_{j=1}^N \sum_{k=1}^N \gamma_{jk}(t_1) = 2 \cdot \gamma(t_1)\end{aligned}\tag{4.19}$$

donde  $\gamma(t_1), \gamma_{jk}(t_1)$  son los tráficos en el instante  $t_1$  y  $\gamma(t_2), \gamma_{jk}(t_2)$  son los tráficos en el instante  $t_2$ .

Dicho con otras palabras, se supone que para los aumentos de tráfico no priman ningún par de nodos, sino que todos los pares origen-destino ven aumentado su tráfico proporcionalmente, de ahí que

$$\frac{\gamma_{jk}}{\gamma} = P(j, k)\tag{4.20}$$

donde  $P(j, k)$  es constante para cada par  $(j, k)$ .

Dado un camino  $\pi_{jk}$ , se llama longitud del mismo al número de canales que atraviesan los mensajes que lo siguen. También se le denomina número de saltos. Se lo simboliza como  $n_{jk}$ .

Interesa conocer la longitud media, o número medio de saltos, definido como el valor medio de los  $n_{jk}$ , según la ecuación:

$$\bar{n} = \sum_{j=1}^N \sum_{k=1}^N \frac{\gamma_{jk}}{\gamma} n_{jk}\tag{4.21}$$

ya que se puede ver el número de saltos como una variable aleatoria discreta, cuyos valores posibles son  $n_{jk}$ , y sus probabilidades correspondientes:

$$P(n_{jk}) = \frac{\gamma_{jk}}{\gamma}\tag{4.22}$$

y, por tanto, según (4.20), constantes para cada par  $(j, k)$ .

Esta última expresión no es más que un cociente de la forma:

$$P(n_{jk}) = \frac{\text{nº de casos favorables a } n_{jk}}{\text{nº de casos totales}} \quad (4.23)$$

En la expresión (4.21) todo se mantiene constante, por tanto  $\bar{n}$  también es constante.

Por último, como se deduce de lo que se explicó al comienzo, cuando la red se encuentra en el período permanente, cada  $\gamma_k$  contribuye al tráfico interno total ( $\lambda$ ) con  $\gamma_k n_k$  mensajes/s (es como una estructura *pipeline* que se va llenando de objetos), de modo que se puede escribir:

$$\square \text{ INCLUDEPICTURE "C:\\Tesis\\Teoría de Colas\\TEMA 2 Modelado y"} \quad (4.24)$$

Teniendo en cuenta (4.21) y (4.24), se obtiene:

$$\bar{n} = \frac{\lambda}{\gamma} \quad (4.25)$$

y, como el número medio de saltos es constante, la expresión (4.25) también, y por tanto el límite de la expresión (4.18) lo mismo.

Llevando este resultado a la expresión de  $T_0$ , ecuación (4.17), se obtiene finalmente:

$$l_0 = \bar{n} \sum_{i=1}^M \frac{\lambda_i / \lambda}{\mu C_i} \quad (4.26)$$

donde la razón  $\lambda_i / \lambda$  permanece constante y sólo depende del algoritmo de encaminamiento, que es el que distribuye el tráfico entre los distintos caminos posibles (se ha considerado encaminamiento fijo, pero esto no deja de ser una simplificación de la realidad).

**4.4.6 Cálculo del punto de saturación** Observando la ecuación (4.16), se llega a la conclusión de que el retardo medio de tránsito aumenta cuando el cociente  $\gamma(\mu C_i - \lambda_i)$  aumenta en alguno de los sumandos de dicha expresión. En el límite,

cuando el denominador de dicho cociente se anula, el canal correspondiente se encontrará totalmente saturado, el retardo correspondiente al mismo será infinito, el retardo de los caminos que incluyan dicho enlace también, y finalmente, el retardo medio de tránsito será infinito.

En principio, no se sabe qué canal, se lo llama  $i_0$ , se va a saturar primero. Por ello, se procede de la siguiente manera:

Para una situación cualquiera de la red, se mide el cociente  $\rho_i = \lambda_i / (\mu C_i)$  para todo canal  $i$ . En el caso de encaminamiento fijo, por ejemplo, primero se calculan los flujos de mensajes a partir de los tráficos externos, según la ecuación (4.9), y después dicho cociente.

Se selecciona el canal en el que el cociente sea mayor; ése será el  $i_0$ . Se incrementan los tráficos externos uniformemente (esto es, proporcionalmente), de modo que los flujos en los canales también aumentarán de esa manera. Se lo hace hasta que el canal  $i_0$  se sature ( $\rho_i = 1$ ). El valor de  $\gamma$  correspondiente a esta nueva situación será el punto de saturación ( $\gamma^*$ ).

#### 4.5 Asignación de capacidades

**4.5.1 Introducción.-** Como ya se dijo, se puede plantear la optimización del comportamiento de la red desde distintos puntos de vista. En este apartado se estudia el problema de asignación de capacidades que consiste, básicamente, en hallar los valores de las capacidades de los canales ( $C_i$ ) que minimizan el retardo medio de tránsito ( $T$ )

**4.5.2 Hipótesis de diseño** Para el cálculo de las  $C_i$ , se parten de las siguientes premisas:

- Se conoce la topología de la red.
- Se conocen los tráficos a cursar ( $\gamma_{jk}$ ).
- Se conoce la longitud media de los mensajes ( $\bar{V}$ ).
- El encaminamiento es fijo y conocido. Es decir, los paquetes con el mismo origen y destino siguen una misma ruta, que se conoce. Por tanto, se conocen los flujos de mensajes en los canales ( $\lambda_i$ ), que se calculan a partir del encaminamiento y los tráficos a cursar, según la ecuación (4.9).

**4.5.3 Topologías de red que favorecen  $T$  bajos** Se deduce que, para tener unos valores de  $T$  bajos, se ha de intentar tener:



- Número medio de saltos bajo. Es lógico: cuantos menos saltos tengan que dar los paquetes, menor será el retardo medio de tránsito ( $T$ ).
- Número de canales bajo y con gran capacidad.

Se ven topologías que minimizan uno de estos parámetros (malla y anillo) y topologías que mantienen bajos ambos (estrella).

#### 4.5.4 Número medio de saltos bajo

$$\bar{n} \text{ bajo} \Rightarrow T \text{ bajo} \quad (4.27)$$

Las topologías que favorecen el mínimo número medio de saltos son las de alta conectividad, es decir, aquellas que tienden a conectar los nodos con todos los demás. El ejemplo extremo es la estructura en malla, donde el número medio de saltos es 1.

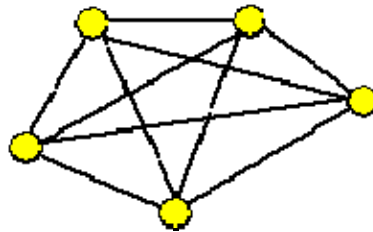


Figura 4.10 Estructura de red en malla.

El inconveniente de estas topologías es que requieren una inversión elevada, aunque sólo sea por la necesidad de cablear todos los nodos entre sí, con los consiguientes costos de obra civil, materiales, etc...

#### 4.5.5 Número de canales bajo

$$\sum_{i=1}^M \sqrt{\lambda_i / \lambda} \text{ bajo} \Rightarrow T \text{ bajo} \quad (4.28)$$

El caso extremo sería que todo el flujo viajara por un canal y los demás no llevaran información; en este caso la sumatoria sería 1. Sin embargo, esta red serviría de poco, de ahí que una solución más razonable consiste en concentrar el tráfico en un número de canales reducido, y de gran capacidad.

Las estructuras de red que favorecen el mínimo número de canales son las topologías en anillo, donde se cumple que el número de canales (unidireccionales) es igual que el número de nodos. Este tipo de estructura aparece en la siguiente figura:

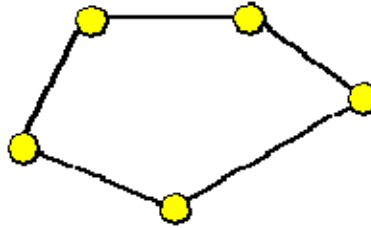


Figura 4.11 Estructura de red en anillo.

El inconveniente de esta topología es que el número medio de saltos tiende a  $N/2$ , contradiciendo el caso anterior (que requería número medio de saltos bajo).

**4.5.6 Topología con número de canales y de saltos bajos.** La topología de red que favorece un mínimo número de canales y de saltos es la topología de red en estrella (figura 4.12), en la que se cumplen:

$$M=2(N-1)$$

El número medio de saltos tiende a 2, o incluso a 1 si el nodo central es un *centro de proceso* que recibe mucho tráfico de los otros nodos (cosa por otro lado bastante normal, por ejemplo, en una entidad bancaria).

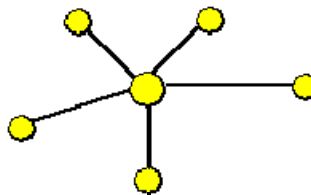


Figura 4.12 Estructura de red en estrella.

**4.5.7 Consideraciones sobre la realidad** En la práctica se llega a un compromiso: se empieza con una topología de red suficiente para las necesidades de ese momento, y se va interconectando la red en función del factor de utilización de los canales, y de las necesidades cambiantes de los usuarios.

En redes profesionales conviene habilitar rutas alternativas que protejan frente a:

- Caída de rutas principales.
- Saturación de rutas principales.

Aunque esto represente un consiguiente aumento del retardo.

Para hacer el estudio anterior, se ha supuesto conocidas las  $\gamma_k$ , pero en realidad éste es un parámetro no sólo desconocido, sino también variable, aunque sí estimable. Por eso se toman precauciones y se sobredimensionan los canales y además se instalan más canales de los estrictamente necesarios, con el objeto de hacer frente a posibles picos de tráfico. Además, desde la solicitud de un determinado servicio a un proveedor (por ejemplo, el aumento de capacidad de un enlace) hasta la disponibilidad del mismo transcurre un tiempo en el que la red ha de seguir en funcionamiento. Si se

han dispuesto canales alternativos que hagan frente a estas contingencias, la red puede seguir funcionando hasta proveer una mejor solución.

En la práctica se tienen en cuenta otros parámetros, además del retardo medio de tránsito ( $T$ ). En ocasiones se pone una cota al retardo de cada canal ( $T_i$ ), o al retardo de ciertos caminos ( $Z_{jk}$ ). Esto se hace para garantizar que en ningún caso se producirá un retardo muy grande (en cuyo caso, al usuario afectado le importará bien poco que la red tenga un retardo medio pequeño).

Otra consideración a hacer es que en un caso real los  $C_i$  no son parámetros continuos, sino discretos. Sólo se pueden escoger unos valores  $C_i$  dentro de un conjunto de valores determinados por las jerarquías sincronas de transmisión actualmente utilizadas (9600 bps, 19200 bps, 64 Kbps, 2 Mbps, 34 Mbps, 155 Mbps)

Por esta razón, no se pueden escoger los valores exactos que se deducen de las fórmulas usadas anteriormente, por lo que el método utilizado es el siguiente:

El diseñador empieza eligiendo arbitrariamente, es decir, como fruto de su experiencia, un conjunto o vector de valores para las capacidades de los canales (que se representan como  $\{C_i\}$ ), valores que se traducen en un  $D_0$  (costo inicial estimado).

Con este  $D_0$ , mediante el anterior análisis, se busca un conjunto o vector  $\{C_i\}$  que minimice  $T$ , o cumpla otros requisitos de calidad (ya se dijo con anterioridad que minimizar el retardo de tránsito es, en la práctica, sólo uno de los posibles. aunque importante, objetivos de optimización).

Como los  $C_i$  hallados no son, en general, valores discretos, hay que elegir unos que sí lo sean, es decir, hay que aproximar la solución por un vector  $\{C_i\}$  con valores comerciales.

Con los nuevos  $C_i$  se halla el nuevo valor del costo total, al cual se le llama  $D_1$ , que se utilizará para calcular un nuevo conjunto  $\{C_i\}$ , y así sucesivamente.

Como puede verse, esto se convierte en un círculo vicioso. En este momento entra la experiencia del diseñador, que detiene el proceso iterativo y escoge la solución definitiva.

## 4.6 Asignación de flujos

**4.6.1 Introducción** En este apartado se estudiará de un problema de optimización opuesto a este último, que se llama **asignación de flujos**, y que consiste en hallar un conjunto o vector de valores para los flujos de bits que atraviesan los canales que minimice ( $T$ ), manteniendo las capacidades y los otros parámetros constantes (hipótesis de diseño).

Fijémonos por un momento en la expresión del retardo medio de tránsito ( $T$ ) ecuación (4.16), que se repite aquí por conveniencia:

$$T = \sum_{i=1}^M \frac{\lambda_i}{\gamma} \frac{1}{\mu C_i - \lambda_i} \quad (4.29)$$

Esta expresión no presenta ningún mínimo relativo respecto de los  $\lambda_i$  y su mínimo absoluto se alcanza para  $\lambda_i = 0$ ,  $i = 1, \dots, M$ . Pero ésta no es una solución a nuestro problema, porque con estos flujos nulos la red no es capaz de cursar el tráfico ofrecido ( $\gamma$ ). De hecho, no es capaz de cursar ningún mensaje.

Por tanto, se debe emplear un método que minimice  $T$  al tiempo que asegure que todo el tráfico ofrecido es cursado. En concreto, los algoritmos que se estudiarán parten de un vector de flujos inicial (que es capaz de cursar todo el tráfico) y realizan pequeños ajustes sobre el mismo, de forma controlada e iterativa. Son, por tanto, algoritmos numéricos cuya convergencia hacia la solución exacta no se demostrarán, y que, por lo tanto, darán una solución aproximada, tanto más precisa cuantas más iteraciones se realicen.

En particular, se verá el algoritmo de Gerla o de desviación de flujos que sirve para mejorar el desempeño de la red.

**4.6.2 Hipótesis de diseño** Para el cálculo de los  $\lambda_i$ , se parten de las siguientes premisas:

- Se conoce la topología de la red.
- Se conocen las capacidades de los distintos canales de la red ( $C_i$ ).
- Se conocen los tráficos ofrecidos a la red ( $\gamma_k$ ).
- Se conoce el tamaño medio de los mensajes ( $1/\mu$ ).

- Se supone que existe solución al problema, es decir, que con las capacidades que se dan es posible, de alguna forma, cursar el tráfico ofrecido a la red ( $\mathcal{V}$ ) con retardo medio de tránsito finito.

De los algoritmos de asignación de flujos que existen, uno de ellos utiliza una única ruta para encaminar todos los mensajes con origen y destino dados (algoritmo de rutas fijas); este camino será el óptimo desde el punto de vista de la sensibilidad, concepto que se define en el siguiente apartado. El otro algoritmo que se verá utiliza, en general, y suponiendo que existan, varios caminos entre un nodo origen y otro destino dados (algoritmo de Gerla); esta solución es, en general, mejor que la anterior, y conduce a un balanceo de carga, esto es, un reparto del tráfico entre todos los caminos posibles, lo que permite, dada su mayor flexibilidad, alcanzar una solución óptima al problema.

**4.6.3 Concepto de longitud de un canal.** Los algoritmos de asignación de flujos que se estudiarán, y, en general, otros algoritmos para encontrar las mejores rutas entre dos nodos, hacen uso del concepto de longitud de un canal, que no es sino una métrica asociada a los canales que mide de algún modo el rendimiento de los mismos. También se le denomina distancia o costo. Estos términos no deben engañarnos, puesto que, en principio, cualquier función que mida comparativamente las prestaciones de los canales puede tomarse como longitud de un canal; se utilizarán una u otra dependiendo del objetivo perseguido. Incluso pueden manejarse varias métricas al mismo tiempo, para satisfacer las necesidades de un amplio abanico de usuarios.

La mejor ruta (o camino) entre dos nodos respecto de una métrica será aquella para la que la suma de las longitudes de los canales que la forman sea mínima. También se la conoce como ruta más corta. Por ejemplo, es muy común en algoritmos de encaminamiento la métrica asociada al número de saltos que dan los mensajes desde el origen hasta el destino. Atendiendo a este criterio, cada canal tendrá longitud unidad y la ruta más corta será la que atraviere menor número de canales.

El objetivo es minimizar el retardo medio de tránsito, parece lógico utilizar una métrica que prime aquellos enlaces con retardo poco sensible al aumento de flujo a su través.

Esto hacen los algoritmos que se verán: utilizan como longitud de un canal una función ( $l_i$ ) que mide la sensibilidad del retardo medio de tránsito al aumento en el flujo en dicho canal, esto es, el cociente entre el incremento en el retardo y el incremento en

el flujo, cuando estos incrementos son infinitesimales. Matemáticamente, se puede escribir:

$$l_i = \frac{\partial T}{\partial (\lambda_i / \mu)} \quad (4.30)$$

En este caso, la mejor ruta ( $\pi^*$ ) —o mejor camino— entre dos nodos será la que posea el menor

$$\sum_{i: C_i \in \pi^*} l_i \quad (4.31)$$

#### 4.6.4 Algoritmo de Gerla o de desviación de flujos

Este algoritmo consigue la minimización del retardo medio de tránsito desviando tráfico hacia las rutas más cortas respecto de la sensibilidad ( $l_i$ ).

Para aplicar el algoritmo de Gerla, se parte de un vector de flujos inicial (en bps y no mensajes/s) que cumpla los siguientes requisitos:

- Cursar todo el tráfico ofrecido a la red ( $\mathcal{V}$ ).
- No saturar ningún canal, es decir:

$$\frac{\lambda_i}{\mu} < C_i, \quad i = 1, \dots, M \quad (4.32)$$

Por conveniencia de notación, se simboliza a este vector inicial mediante:

$$\vec{f}^{(0)} = \left( \frac{\lambda_1^{(0)}}{\mu}, \dots, \frac{\lambda_M^{(0)}}{\mu} \right) \quad (4.33)$$

Análogamente, se representa el vector de flujos hallado en la iteración  $n$ -ésima por medio de

$$\vec{f}^{(n)} = \left( \frac{\lambda_1^{(n)}}{\mu}, \dots, \frac{\lambda_M^{(n)}}{\mu} \right) \quad (4.34)$$

Sin más preámbulos, el algoritmo consta de los siguientes pasos:

**Paso 1** Tomar  $n = 0$ .

**Paso 2** Para cada  $i = 1, \dots, M$  calcular:

$$l_i = \frac{C_i}{\gamma \left[ C_i - \left( \lambda_i^{(n)} / \mu \right) \right]^2} \quad (4.35)$$

**Paso 3**

Calcular la tasa incremental de costo para el vector de flujo  $\vec{f}^{(n)}$ , que se representa como  $\beta_n$ , definida mediante la siguiente expresión:

$$\beta_n = \sum_{i=1}^M l_i \frac{\lambda_i^{(n)}}{\mu} \quad (4.36)$$

**Paso 4** Calcular todas las rutas más cortas entre cada par de nodos, utilizando como métrica la sensibilidad de los canales (valores calculados en el paso 2). Calcular el vector de flujo resultante de enviar los mensajes por las rutas más cortas (nótese que sólo hay una mejor ruta por cada par de nodos, de modo que los mensajes entre cada par de nodos viajan por un único camino). Se simboliza este vector mediante:

$$\vec{\Phi} = (\phi_1, \dots, \phi_M) \quad (4.37)$$

**Paso 5**

Calcular la tasa incremental de coste para el flujo  $\vec{\Phi}$ , que se representa como  $b_n$ , y cuya expresión es

$$b_n = \sum_{i=1}^M l_i \phi_i \quad (4.38)$$

**Paso 6 (Condición de parada del algoritmo)**

Si  $\beta_n - b_n < \epsilon$ , donde  $\epsilon < 0$  es un parámetro de tolerancia escogido

adecuadamente, entonces PARAR; en este caso, el vector  $\vec{f}^{(n)}$  es la solución buscada.

Si no, ir al paso 7.

**Paso 7**

Encontrar el valor de  $\alpha$  en el rango  $0 \leq \alpha \leq 1$  tal que el vector de flujo

$(1 - \alpha) \vec{f}^{(n)} + \alpha \vec{\Phi}$  minimiza  $T$ . Se simboliza este valor por medio de  $a$ .

**Paso 8 (Desviación de flujo)**

Sea  $\vec{f}^{(n+1)} = (1 - a) \vec{f}^{(n)} + a \vec{\Phi}$ . Como resultado de esta asignación, parte del flujo en cada canal es desviado por las rutas más cortas con el mismo origen y destino, resultando un retardo medio de tránsito lógicamente menor.

**Paso 9** Tomar  $n=n+1$ .

Ir al paso 2.

El algoritmo de Gerla converge hacia la solución mejor, u óptima. Los algoritmos de encaminamiento que se utilizan en las redes reales tratan de aproximarse lo más posible a esta solución, pero nunca puede alcanzarse porque los parámetros de la red, y en concreto  $\mathcal{V}$ , varían continuamente de forma tan rápida que ningún algoritmo puede seguir estas variaciones y adaptarse a ellas en tiempo cero. Además, téngase en cuenta que ni siquiera el algoritmo de Gerla da una solución exacta, sino que interviene un parámetro de tolerancia que decide cuándo se para el algoritmo.

#### 4.7 Control de Flujo y Congestión.

El caudal depende del tipo de red y tiene un valor nominal máximo, que no se podrá superar en ningún caso. Pero además, la red no ofrece el mismo caudal real si se le ofrece poco tráfico o si se le ofrece mucho.

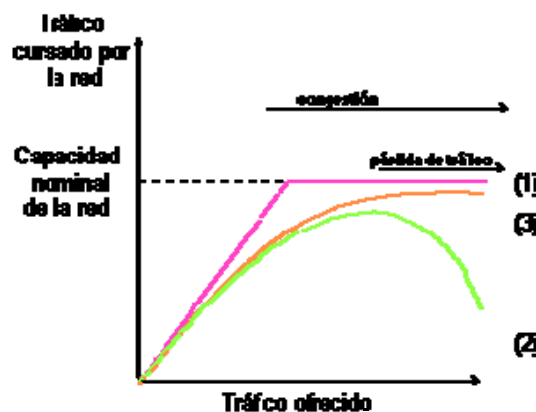


Figura 4.13 Caudal en función del tráfico ofrecido

La curva (1) representa el comportamiento **ideal** de la red: hay linealidad hasta llegar a la capacidad nominal de la red, momento en el que el tráfico cursado se satura. La curva (2) representa el comportamiento real típico de una red. Como puede observarse, al llegar a la zona de saturación, cuanto más tráfico se ofrece menos tráfico se cursa. Esto es debido, por ejemplo, a que los paquetes tardarán mucho tiempo en llegar a su destino, y mientras tanto serán retransmitidos por la fuente, pensando que se han perdido por el camino. Esto, a su vez, origina una explosión de tráfico, ya que cada paquete es retransmitido varias veces, hasta que consigue llegar a tiempo al destino.

Para evitar esa degradación, se introduce el control de congestión que trata de aproximar el comportamiento de la red al dado por la curva (3), evitando así entrar en una zona de degradación. El retardo de tránsito en la red sigue la siguiente curva:



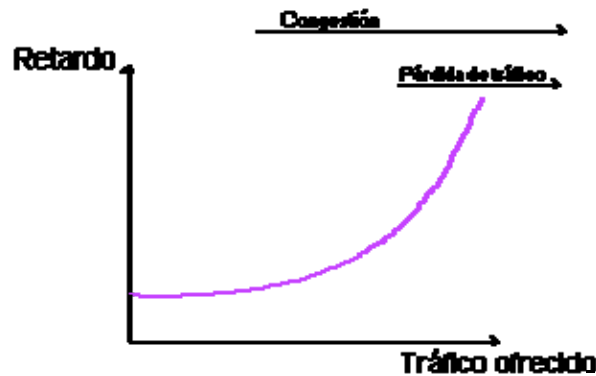


Figura 4.14 Retardo de tránsito en función del tráfico ofrecido

Se ve que el retardo no aumenta linealmente, sino que el aumento de éste es mayor que el aumento de tráfico ofrecido.

**4.7.1 Causas de la congestión.** Hay varias causas de congestión. Se enumeran aquí las más importantes:

**Memoria insuficiente de los conmutadores.** En la siguiente figura se observa:

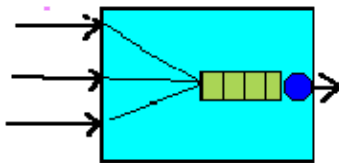


Figura 4.15 Saturación del buffer de un nodo

En ella se tiene un conmutador en el que tres líneas de entrada mandan paquetes a una de salida. Así puede llenarse el *buffer* (cola) de la línea de salida.

Además, si hay congestión en otros nodos, las colas no liberan la información de los paquetes transmitidos (que se guarda por si hay que retransmitir), con lo que la situación empeora aún más.

**Insuficiente CPU en los nodos.** Puede que el nodo sea incapaz de procesar toda la información que le llega, con lo que hará que se saturen las colas.

**Velocidad insuficiente de las líneas.** Se tiene el mismo problema que en el caso anterior.

#### 4.7.2 Diferencia entre control de flujo y control de congestión.

##### 4.7.2.1 Control de flujo.

Es una técnica que permite sincronizar el envío de información entre dos entidades que producen o procesan la misma a distintas velocidades. Por ejemplo, Se supone el caso representado en la siguiente figura:

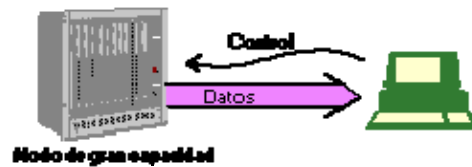


Figura 4.16 Conexión entre nodo de alta capacidad y PC

En este caso, dada la gran velocidad a la que produce y envía información, el nodo desborda al PC, por lo que éste debe enviar información de control (control de flujo) para que el nodo reduzca su tasa de envío de datos. De esta forma, parando a la fuente cada cierto tiempo, el PC puede procesar el tráfico que le envía el nodo.

#### 4.7.2.2 Control de congestión.

Es un concepto más amplio que el control de flujo. Comprende todo un conjunto de técnicas para detectar y corregir los problemas que surgen cuando no todo el tráfico ofrecido a una red puede ser cursado, con los requerimientos de retardo, u otros, necesarios desde el punto de vista de la calidad del servicio. Por tanto, es un concepto global, que involucra a toda la red, y no sólo a un remitente y un destinatario de información, como es el caso del control de flujo.

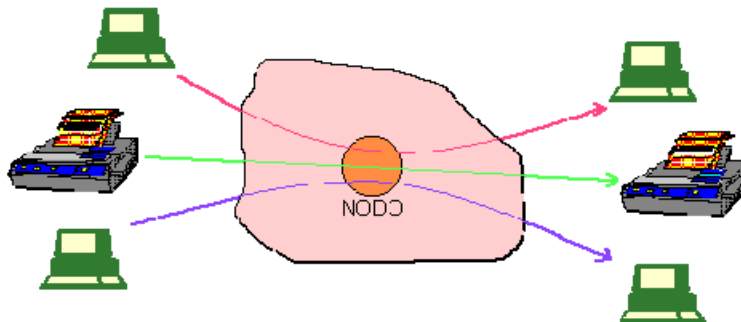


Figura 4.17 Congestión en un nodo

El control de flujo es una más de las técnicas para combatir la congestión. Se consigue con ella parar a aquellas fuentes que vierten a la red un tráfico excesivo.

#### 4.7.3 Soluciones. Mecanismos de control de congestión.

El problema del control de congestión puede enfocarse matemáticamente desde el punto de vista de la teoría de control de procesos, y según esto pueden proponerse soluciones en **bucle abierto** y en **bucle cerrado**.

**4.7.3.1 Soluciones en bucle abierto.** También llamadas **soluciones pasivas**. Combaten la congestión de las redes mediante un adecuado diseño de las mismas. Existen múltiples variables con las que el diseñador puede jugar a la hora de diseñar la red. Estas variables influirán en el comportamiento de la red frente a la congestión.

**4.7.3.2 Soluciones en bucle cerrado.** También llamadas **soluciones activas**. Actúan cuando se detectan problemas. Tienen tres fases:

- **Monitorización de parámetros.** Se vigilan los siguientes parámetros:
  - Ocupación de los enlaces y de los buffers (colas de espera en los nodos)
  - Porcentaje de descartes
  - Número de retransmisiones
  - Retardos y "jitters". Jitter: oscilaciones de la separación temporal entre paquetes. En aplicaciones que requieren sincronización (videoconferencia, sincronizar audio con video), es muy importante que esas oscilaciones sean pequeñas.
- **Reacción: envío de información a los puntos necesarios.** La comunicación se realiza gracias a:
  - Paquetes especiales. **NO** están sometidos a control de congestión y se saltan las colas de espera en los nodos. Los envía el nodo que, gracias a la monitorización, ha detectado la congestión.
  - Bits de cabecera. En los paquetes enviados, indico en la cabecera que empieza a haber congestión. (Ejemplo Frame Relay).
  - Información específica. Si se recibe una alerta de congestión (mediante bits de cabecera de paquetes que circulan por la red), se solicita más información.
- **Ajuste del sistema.** Hay varias medidas:
  - Reducir la velocidad de envío
  - Control de acceso. No se permiten más conexiones.
  - Tirar paquetes. Controlar ráfagas de paquetes que llegan.