



## APÉNDICE

- 3.A1. Introducción
- 3.A2. Caracterización de un sistema de colas
- 3.A3. Ejemplos de sistemas de colas
- 3.A4. Aplicación al análisis de prestaciones en TDM

## TEORÍA DE COLAS

## 3.A1. Introducción

El retardo medio en la transmisión es un factor importante a tener en cuenta a la hora de determinar el rendimiento en una red de comunicaciones. En su evaluación pueden intervenir varios aspectos, como el tiempo de transmisión de la trama, el tiempo de propagación por el canal, el tiempo de acceso al canal o el tiempo de espera en las colas de los sucesivos dispositivos de encaminamiento. La mayoría de estos factores pueden ser determinados a priori de forma relativamente simple a partir de parámetros de los sistemas involucrados, siendo independientes del nivel de ocupación de la red o de la actividad de otros nodos. Sin embargo, dado que se utilizan memorias temporales para almacenar los datos en varios contextos, como es el caso del acceso al canal, aparecerán colas en las que los bloques de datos deben esperar su turno para ser atendidos. El tiempo de permanencia en estas colas dependerá de su ocupación y, por tanto, del volumen de tráfico asociado. Para evaluar estos tiempos es necesario utilizar la teoría de colas, que proporcionará valores para algunos parámetros relevantes.

Antes de describir los fundamentos de teoría de colas, es importante incidir en la naturaleza estadística de los valores que proporciona, que serán valores esperados o medios, no valores instantáneos. En este sentido, la teoría de colas no permite evaluar el retardo que sufrirá una trama concreta, sino el valor promedio para todas las tramas.

Un sistema de colas puede ser descrito (Figura 3.A1) como aquel compuesto por una *cola de servicio*, en la que se almacenan trabajos en espera de ser atendidos, secuencialmente, por uno o varios *módulos de servicio*. Su funcionamiento queda definido por tres elementos fundamentales:

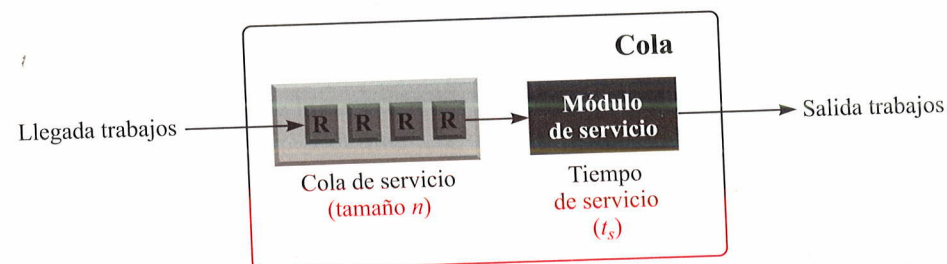


Figura 3.A1. Esquema general de un sistema de colas.

1. forma de llegada de los trabajos a la cola,
2. forma en que estos son atendidos (servidos) y
3. el número de servidores.

Así pues, para llevar a cabo el estudio acerca del comportamiento de un sistema de colas, es necesario definir de algún modo los elementos anteriormente comentados. La especificación general nemónica de este tipo de sistemas es de la forma  $X/Y/n$ , donde  $X$  indica la distribución de llegada de trabajos a la cola,  $Y$  la distribución de servicio ofrecida y  $n$  el número de servidores que atienden o sirven a la cola.

Existen distintos tipos de distribuciones, tanto en lo que se refiere a la recepción de trabajos, como en lo que respecta al servicio de los mismos. Estas distribuciones hacen referencia a los tiempos entre los eventos considerados, que pueden ser:

- *Poisson*. Identificada mediante la letra  $M$ , la expresión matemática de este tipo de distribución es la indicada en (3.7). En este caso, los tiempos entre eventos tienen una distribución exponencial.
- *Determinista*. En este caso, la expresión de la distribución es determinista, no estadística. La letra con la que se referencia esta distribución es  $D$ .
- *Genérica*. Como indica su nombre, una distribución genérica ( $G$ ) es aquella cuya expresión puede establecerse de forma general por el usuario.

De acuerdo con esto, un sistema  $M/D/3$  hace referencia a un sistema en el que la forma de llegada de los trabajos sigue una distribución de Poisson y estos son servidos por 3 servidores de acuerdo con una distribución determinista.

En la descripción previa se ha considerado que el número de posiciones en la cola de servicio es infinito, esto es, existen suficientes posiciones para almacenar todos los trabajos que se reciban. Aunque esta suposición es habitual en el modelado de la mayoría de los sistemas de colas, en los casos en los que sea conveniente considerar un tamaño finito para la cola se añade un cuarto parámetro, pasándose a la notación  $X/Y/n/N$ , siendo  $N$  el número máximo de trabajos en la cola.

## 3.A2. Caracterización de un sistema de colas

El estudio de un sistema de colas consiste básicamente en la determinación de cuatro parámetros característicos (Figura 3.A2):

- a)  $q$ : número total de trabajos en la cola, es decir, los que están siendo servidos más los que se encuentran en espera en la cola de servicio.
- b)  $t_q$ : tiempo total de los trabajos en la cola o, lo que es lo mismo, tiempo desde que los trabajos entran en la cola hasta que salen del módulo de servicio.
- c)  $t_w$ : tiempo de espera, es decir, tiempo que están los trabajos en la cola de servicio.
- d)  $w$ : número de trabajos en espera, es decir, en la cola de servicio.

En la obtención de los parámetros anteriores aparecen involucrados otros:

- a)  $t_s$ : tiempo de servicio de los trabajos en el sistema de colas. Es el tiempo empleado en atender (servir) un trabajo por el módulo de servicio.
- b)  $E[t_s]$ : valor esperado del tiempo de servicio. A partir de este se define también la tasa de servicio,  $\mu$ , como  $\mu = 1/E[t_s]$ .
- c)  $\phi_s$ : varianza del tiempo de servicio.



en consecuencia, el tiempo de espera  $t_w$  es cero, el tiempo total en cola de los trabajos,  $t_q$ , se deberá únicamente al tiempo de servicio,  $t_s$ .

Para concluir la breve introducción a la teoría de colas aquí desarrollada, hemos de señalar la relación entre un sistema de colas genérico tal como se ha presentado y una red de computadores. Esta relación puede establecerse en múltiples escenarios. Por ejemplo, en el caso de un sistema TDM, como veremos a continuación, el servidor será el conmutador y los trabajos cada una de las ranuras que deben ser enviadas sobre la línea. Si consideramos una red de difusión, el servidor correspondería al medio o línea de transmisión y los trabajos a las tramas generadas por cada una de las estaciones que tienen acceso al mismo. Así, una red LAN suele plantearse como un sistema de tipo M/M/1, donde los datos, recibidos según una distribución de Poisson, se transmiten siguiendo este mismo tipo de esquema sobre el único canal que existe a tal fin.

### 3.A4. Aplicación al análisis de prestaciones en TDM

A modo de ejemplo de aplicación de la teoría de colas en sistemas de comunicaciones, consideraremos el análisis del retardo y rendimiento de un sistema TDM.

Supongamos un sistema TDM real en el que, por cuestiones de eficiencia, se ha dimensionado la capacidad de salida de forma que es menor que la suma de las velocidades nominales de las fuentes, aunque superior a la suma de sus velocidades efectivas. Esto es, en promedio, el sistema es capaz de transportar los datos de todas las fuentes, aunque no lo es en periodos de pico en la transmisión. Esta situación podría corresponder, por ejemplo, a una red telefónica, en la que, habitualmente, no todos los usuarios utilizan la línea simultáneamente. Por cuestiones de coste, la capacidad de las líneas troncales no es suficiente para alojar todas las posibles comunicaciones, sino que se dimensiona para alojar el volumen habitual de tráfico, con un pequeño margen para posibles picos.

En este escenario será necesario establecer memorias temporales para almacenar los datos durante los periodos de pico, dando lugar a un sistema de colas. En este, el módulo de servicio será el conmutador, que debe transmitir cada una de las ranuras sobre el canal de comunicaciones. Los trabajos serán, por tanto, las ranuras procedentes de las fuentes y el tiempo de servicio se corresponderá con el tiempo de transmisión de cada ranura. Consideremos  $I$  fuentes, todas ellas con las mismas características, esto es:

- $I$ , número de fuentes;
- $V$ , velocidad nominal de cada fuente, en bps;
- $\alpha$ , fracción de tiempo que transmite cada fuente. Evidentemente,  $0 \leq \alpha \leq 1$  y la velocidad efectiva,  $V_e$ , de cada fuente será  $V_e = \alpha V$ .

Si  $C$  es la capacidad de la línea de salida, en bps, para que el sistema funcione adecuadamente deberá cumplirse

$$C \geq \alpha IV \quad (3.A5)$$

El módulo de servicio deberá enviar las ranuras, para lo que empleará un tiempo de servicio de valor

$$t_s = \frac{L}{C} \quad (3.A6)$$

siendo  $L$  la longitud, en bits, de cada ranura.

La tasa de llegada de ranuras será

$$\lambda = I \cdot \frac{L}{V_e} = \alpha I \cdot \frac{L}{V} \quad (3.A7)$$

Por tanto, de acuerdo a teoría de colas, la utilización de la línea será

$$u = \lambda \cdot E[t_s] = \frac{\alpha IV}{C} \quad (3.A8)$$

A partir de estos valores podremos evaluar el retardo sufrido por las ranuras sin más que suponer una distribución para las llegadas y aplicar teoría de colas. En primera aproximación, supondremos que los tiempos de servicio son deterministas, ya que corresponden a la duración de las ranuras en el canal de salida, y que los tiempos entre llegadas siguen una distribución de Poisson. La cola será, por tanto, del tipo M/D/1, por lo que

$$E[t_q] = \frac{2-u}{2 \frac{C}{L}(1-u)}; \quad E[t_w] = \frac{u}{2 \frac{C}{L}(1-u)}; \quad q = \frac{u(2-u)}{2(1-u)} \quad (3.A9)$$

Si representamos el retardo,  $E[t_q]$ , en función del valor de la utilización de la línea obtendremos la gráfica mostrada en la Figura 3.A4. En ella podemos observar que el retardo sufre un brusco incremento cuando la utilización se aproxima a 1. De hecho, el valor del retardo crece asintóticamente a medida que  $u$  tiende a 1. Esto es lógico si consideramos el comportamiento del sistema. Si la línea de salida está dimensionada de forma que su capacidad es muy superior a la necesaria para las velocidades efectivas de las fuentes, el retardo será muy reducido, pero la utilización será baja. Si intentamos que la capacidad de la línea sea aprovechada en su totalidad, el número de ranuras en la cola de servicio crecerá, lo que podemos deducir también a partir de (3.A9), y, consecuentemente, el retardo también lo hará. A partir de este análisis podemos concluir que un buen diseño

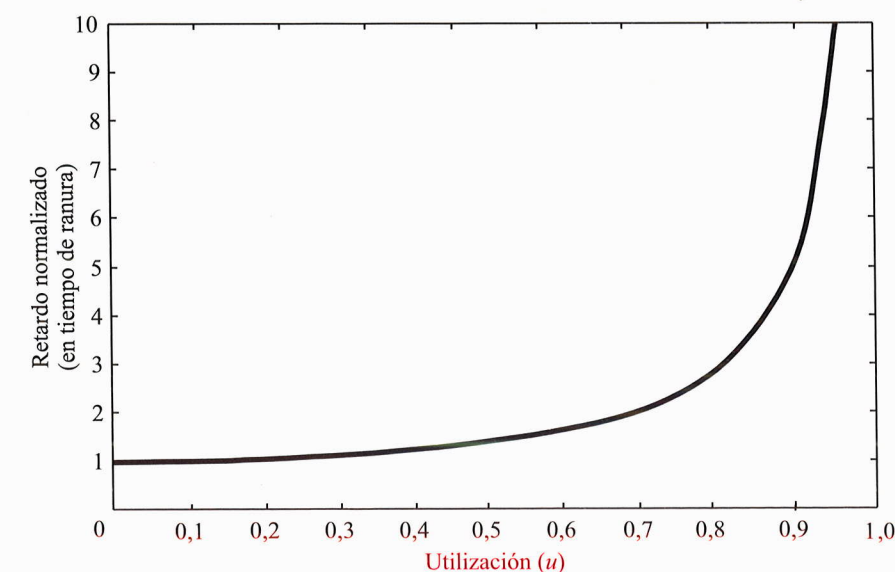
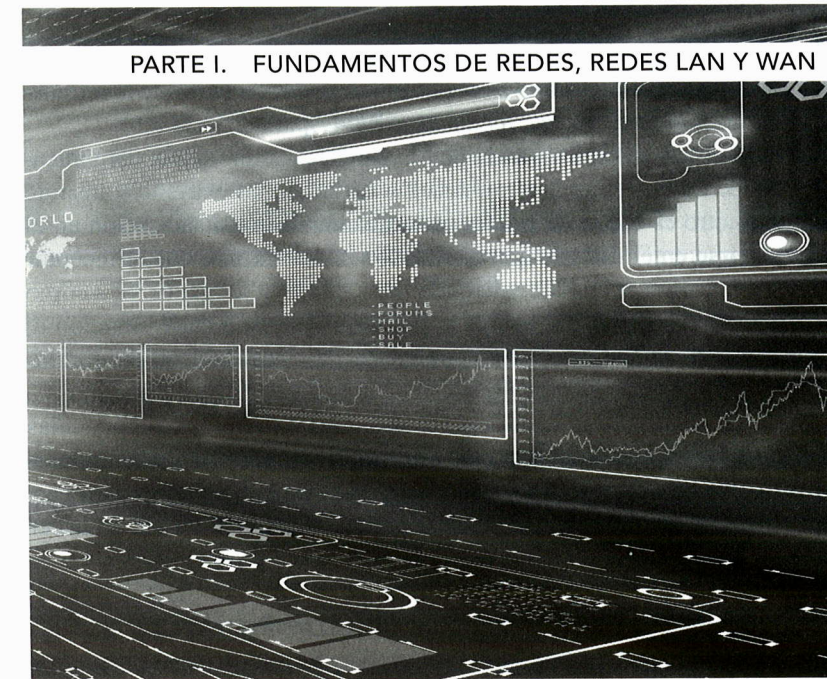


Figura 3.A4. Retardo sufrido por las ranuras en función de la utilización en un sistema TDM.



del sistema debería proporcionar un valor de  $u$  alrededor de  $u = 0,8$ , a pesar de que ello implique la infrautilización de la línea de salida.

Evidentemente, en un sistema real no es posible que el retardo sea infinito, puesto que tampoco será infinito el tamaño de la cola de servicio. En este caso, habría que considerar un sistema de colas con tamaño de memoria finito,  $M/D/1/N$ , y aparecerá una probabilidad de pérdida de ranuras relacionada con la posibilidad de que se llene la memoria. Esta probabilidad también puede ser evaluada con teoría de colas, aunque no lo abordaremos en este libro.



## CONTROL DEL ENLACE DE DATOS

CAPÍTULO

4

- 4.1. Introducción
- 4.2. Delimitación de tramas
- 4.3. Códigos para la gestión de errores
- 4.4. Control de flujo
- 4.5. Técnicas ARQ
- 4.6. Ejemplo de protocolo de control de enlace: LLC

### 4.1. Introducción

A lo largo de este capítulo se aborda el estudio de las funciones de la capa de enlace ya apuntadas en el Apartado 1.3 del tema introductorio. El objetivo de esta capa es proporcionar las funcionalidades necesarias para posibilitar un servicio de transferencia de datos fiable a través del enlace físico. Esto es, para garantizar que todos y cada uno de los datos emitidos se reciben correctamente en el destino cuando estos pueden interactuar directamente entre sí. Para ello se requiere el análisis de los problemas que pueden aparecer y la propuesta de mecanismos o funcionalidades que los resuelvan.

Una de ellas, tal vez la más básica y evidente, es la del control de errores. Dado que todo canal de comunicación es real, es posible la ocurrencia de errores de transmisión que provoquen que los datos recibidos en el receptor difieran de los emitidos por el origen. Es por ello necesario arbitrar algún esquema o método en base al cual el receptor de una transmisión sea capaz de, en primer lugar, determinar la potencial ocurrencia de errores y, a continuación, actuar en consecuencia a fin de solucionarlos. Las técnicas de control de errores pasan por la inclusión de redundancia en la transmisión, de forma que se dota al receptor de capacidad para detectar la aparición de errores en los datos y, en algunos casos, incluso para su corrección por el mismo receptor. En el caso de que únicamente se pueda realizar la detección de los errores, la recuperación de los datos se realiza a partir de la solicitud al emisor de su retransmisión. Esta forma de operación, denominada solicitud de retransmisión automática (ARQ, «Automatic Repeat reQuest») suele ser la habitual en las redes de computadores y se suele combinar, como veremos a lo largo de este capítulo, con otras funcionalidades. Por ello, a la hora de