

TEORÍA DE COLAS Y STOCKS

Índice

Elementos de teoría de colas

Descripción de un sistema de colas 3

El proceso de Poisson y la distribución exponencial 4

Variables asociadas a un sistema de colas 9

Notación de Kendall 11

Cola determinística 11

Colas tipo M/M/1

Comprobación de la distribución de las llegadas 14

Estudio de la cola 16

Medidas de eficacia en estado estacionario 19

Fórmulas de Little 21

Ejemplos 22

Colas tipo M/M/C

Estudio de la cola 24

Medidas de eficacia en estado estacionario 28

Ejemplos 29

Cola tipo M/M/C/K 34

Cuadro resumen 37

Colas tipo M/G/1

Estudio de la cola 38

Fórmula de Pollaczek–Khintchine 40

Distribución estacionaria de los puntos de salida 42

Ejemplos 43

Colas tipo M/Ek/1 45

Colas tipo M/G/1/N 45

Colas con parámetros variables

Caso general 47

Cola con desaliento 48

Cola binomial 48

Cola con desaliento dependiente del tiempo de servicio 49

Cola con tasa de servicio dependiente del estado 50

Cola con servidor adicional cuando la cola es grande 50

Colas con pérdidas 51

Problema de las máquinas 52

Redes de colas

Redes de colas en serie 55

Redes de colas abiertas 56

Redes de colas cerradas 57

Redes de colas cíclicas 57

ELEMENTOS DE TEORÍA DE COLAS

Descripción de un sistema de colas

La teoría de colas aparece a principios del presente siglo para estudiar los problemas de congestión de tráfico que se presentaban en las recientemente inventadas comunicaciones telefónicas. Entre 1903 y 1905, Erlang es el primero en tratar el tráfico telefónico de forma científica, y establece la unidad de tráfico telefónico, que recibe su nombre. Posteriormente esta teoría se ha aplicado a multitud de problemas de la vida real, como el tráfico de automóviles, la regulación de semáforos en una ciudad, la determinación del número de cajeros en los hipermercados, o el control de los tiempos de espera de los procesos que acceden al procesador de un ordenador que trabaja en tiempo compartido.

Los elementos más importantes de un sistema de colas son: las llegadas, la cola, el servicio y la salida.

En general, un sistema de colas consiste en uno o varios servidores que prestan un servicio a uno o varios usuarios que acceden al sistema. El **proceso de llegadas** lo regula una fuente generadora de usuarios y, en general, estas llegadas serán de forma aleatoria. Esta fuente generadora de usuarios puede ser finita o infinita.

Interesa saber cuál es el intervalo de tiempo entre las llegadas de dos usuarios consecutivos. Además, según cómo sea el proceso de llegadas, los usuarios pueden llegar individualmente o en grupos

Si cuando un usuario llega al sistema el servidor está libre, se le da servicio. Si el tiempo de servicio es mayor que el intervalo entre llegadas, el siguiente usuario, cuando accede al sistema, encuentra que el servidor está ocupado, por lo que debe quedar en espera, formando la cola.

Otra cuestión importante es saber cuánto tiempo debe esperar un usuario que llega al sistema hasta que recibe el servicio, lo cual entra dentro del concepto QOS (Quality of Service, calidad de servicio). Cuando en la cola hay más de un usuario, al quedar el servidor libre hay que determinar cuál de los usuarios en espera será el que pase a recibir servicio. Es decir, es necesario un proceso para decidir qué usuario va a ser llamado de la cola; esto es lo que se llama **disciplina de la cola**. Los modelos más importantes son los siguientes:

- FIFO (First-In-First-Out): se le da servicio al primero que ha llegado, de forma que la cola está ordenada según el orden de llegada de los usuarios.
- LIFO (Last-In-First-Out): se le da servicio al último que ha llegado, de forma que la cola está ordenada en orden inverso al de llegada de los usuarios.
- SIRO (Service-In-Random-Order): Se sortea aleatoriamente cuál de los usuarios en espera accederá al servicio.

No obstante, otro procedimiento para establecer la disciplina de la cola puede ser el de establecer determinadas prioridades a los diferentes usuarios según algunas de sus características.

En sistemas finitos, en los que el número de usuarios en espera es limitado, es necesario establecer además qué sucede con aquellos usuarios que acceden al sistema cuando la cola de espera está completa. Por último, en los sistemas en que los usuarios son humanos, hay que tener en cuenta otros factores propios del comportamiento humano como el hecho de que hay individuos que no respetan el orden establecido en la cola o bien que hay usuarios que, a la vista de la cola, renuncian a acceder al sistema.

Otra característica importante de un sistema de colas es el diseño de la ejecución del servicio. El servicio puede estar ejecutado por uno o varios servidores. Si el tiempo que tardan los usuarios en salir del sistema es mayor que el intervalo entre llegadas, la cola aumentará indefinidamente y el sistema puede llegar a colapsarse. Por tanto es necesario diseñar el sistema de forma que el tiempo de servicio sea igual o menor que el intervalo entre llegadas. En esta situación es importante saber cuánto tiempo va a estar un servidor inactivo, tiempo que ha de ser mínimo para optimizar el rendimiento del sistema. No obstante, en la mayoría de los sistemas la duración del servicio es también una magnitud aleatoria.

Por último, los usuarios que salen del sistema pueden hacerlo al exterior o pueden integrarse en otro sistema similar, en cuyo caso se habla de colas enlazadas o **redes de colas**.

El proceso de Poisson y la distribución exponencial

En la mayoría de los sistemas de colas, el proceso de llegadas sigue una distribución de Poisson. Se demuestra que si se da esta circunstancia, la duración de los intervalos entre llegadas tiene una distribución exponencial o una combinación continua de exponenciales, es decir, una distribución gamma, que recibe el nombre de **distribución erlangiana**, o distribución K.

En efecto, si llamamos