Simulación a Eventos Discretos

Teoría de colas y SED

Teoría de colas y SED

SED en muchos casos se utiliza para analizar sistemas de colas.

Elementos: cliente, cola, servidor.

El tomador de decisiones tiene control sobre algunos parámetros (por ejemplo, cantidad de servidores, tiempo de atención), por lo tanto puede influir sobre el desempeño del sistema (por ejemplo, tiempo de espera, largo de cola).

En sistemas sencillos es posible calcular las medidas de desempeño de forma analítica (no es necesario utilizar SED); cuando no son sencillos cobra relevancia SED, pero aún son útiles los modelos analíticos (por ejemplo, para evaluación preliminar).

Elementos y terminología de un sistema de colas

Clientes: Arriban al sistema y requieren un servicio.

Servidores: Entidades que proveen el servicio.

Población: Clientes potenciales, pueden ser infinitos (la tasa de arribos no depende de quienes ya dejaron la población y entraron al sistema) o finitos (por ejemplo, la cantidad fija de máquinas de un taller, que pueden estar disponibles o no).

Arribos: En general dados por un tiempo entre arribos aleatorio; pueden estar agendados en tiempos fijos.

Capacidad: De las colas, afecta la tasa de arribos (real vs. efectiva).

Servicio: Tiempo constante o aleatorio; servidores paralelos.

Colas: comportamiento y disciplina

Comportamiento:

- Balking: El cliente se retira antes de ingresar al sistema.
- Reneging: El cliente se retira luego de esperar un cierto tiempo, antes de ser servido.
- Swopping: El cliente se cambia a una cola (más corta o más rápida).

Disciplinas:

- FIFO: Primero en llegar, primero en salir.
- LIFO: Último en llegar, primero en salir.
- SIRO: Select In Random Order.

Colas: caracterización

Notación de Kendall: $A/B/c/N/K$, donde:

- A es la distribución del tiempo entre arribos.
- B es la distribución del tiempo de servicio.
- c es la cantidad de servidores paralelos.
- N es la capacidad del sistema.
- K es el tamaño de la población.

Por ejemplo, $M/M/1/\infty/\infty$ es un sistema de un solo servidor con capacidad ilimitada y potenciales arribos infinitos. Los tiempos entre arribos y los de servicio tienen una distribución de probabilidad exponencial negativa (con diferentes tasas).

Es equivalente decir $M/M/1$.

Medidas sobre sistemas de servidores paralelos

Donde λ es la tasa de arribos de clientes al sistema y μ es la tasa de servicio.

Medidas para colas $G/G/c/N/K$.

Dos tipos de medidas: promedio de la muestra y promedio ponderado en el tiempo de la muestra.

El sistema se dice estable si la relación entre los arribos y la capacidad de atención no favorece la formación de colas de largo infinito.

Número medio de clientes en el sistema

$$L = \frac{1}{T}\int_0^T L(t)dt = \frac{1}{T}\sum_{i=0}^{M} iT_i$$

Donde

- T_i es la cantidad de tiempo (eventualmente en intervalos disjuntos) en el cual en el sistema hay i clientes, cumple que $\sum_{i=0}^{M} T_i = T$
- $L(t)$ es la cantidad de clientes en el sistema en el instante t.
- M es la máxima cantidad de clientes observada.

Asintótico cuando t (M) tiende a infinito en sistemas estables.

Tiempo medio de permanencia de clientes en el sistema

$$w = \frac{1}{N} \sum_{i=1}^{N} W_i$$

Donde

- N es la cantidad de clientes observados.
- W_i es el tiempo de permanencia del cliente i.

Asintótico cuando N tiende a infinito en sistemas estables.

Ecuación de conservación de Little: $L = \lambda w$

Utilización de servidores

$$\rho = \frac{\lambda}{c\mu}$$

Tiene sentido para sistemas estables, es decir, que cumplen $\lambda < c\mu$.

Para sistemas inestables, tampoco tienen sentido las medidas asintóticas de L y w.

Medidas en estado estacionario para colas $M/M/c$

Arribos y servicios según distribución exponencial negativa.

P_0 es la probabilidad en estado estacionario de tener 0 clientes en el sistema.

$$P_0 = 1/[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)}]$$

$$P(L(\infty) \geq c) = \frac{(c\rho)^c P_0}{c!(1-\rho)}$$

$$L = c\rho + \frac{\rho P(L(\infty) \geq c)}{1-\rho} = c\rho + L_q \qquad (L_q\text{: clientes en la cola } q)$$

$$W_q = L_q/\lambda \qquad (W_q\text{: tiempo medio de espera en la cola } q)$$

Otras consideraciones

Sistemas con capacidad.

Población finita.

Redes de colas.