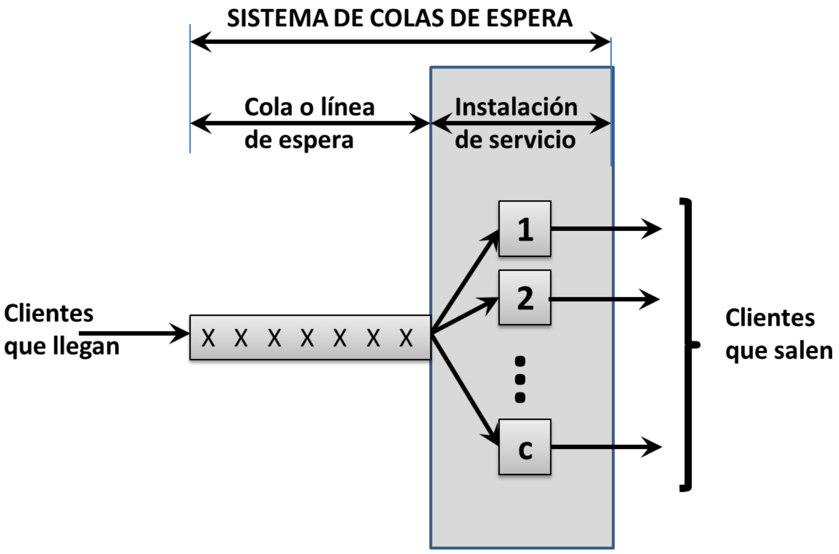
Teoría de Colas

Partes del Sistema de Colas



- Los clientes (part, entity) buscan
- Servicio de unas instalaciones de servicios o servidores (server, machine)
- Esperando en una cola (queue, buffer)

Ejemplos

- Banco
- Supermercado
- Semáforo
- Aviones dispuestos a despegar
- Ascensor
- Peluquería
- Mantenimiento de equipos
- Alumnos y ordenares e impresoras libres
- Peaje en una autopista

Los Procesos de Llegada y Salida

Las distribuciones de los procesos de llegada y de salida son muy características en un sistema de colas.

Derivación de las distribuciones

Si las condiciones en un intervalo de tiempo h son:

1) La probabilidad de que ocurra un evento entre t y t+h depende sólo de h y no de lo ocurrido antes de t o del valor de t.
2) La probabilidad de que un evento ocurra en h es positiva y menor de 1.
3) Sólo puede ocurrir un evento en h.

Si:

$P_n(t)$ = probabilidad de que ocurran n eventos en t

entonces:

1) $p_0(t+h) = p_0(t)\, p_0(h)$
2) $0 < p_0(h) < 1$

$$p_0(h) = e^{-\alpha h} = 1 - \frac{(\alpha h)^1}{1!} + \frac{(\alpha h)^2}{2!} + \frac{(\alpha h)^3}{3!} + \cdots \cong 1 - \alpha h,$$

3) $p_1(h) = 1 - p_0(h) \cong \alpha h$

El anterior resultado indica que la probabilidad de ocurrencia depende proporcionalmente de h. Intentemos hallar la:

- f(t) = función de densidad de probabilidad del intervalo de tiempo t entre la incidencia de eventos sucesivos, $t \geq 0$
- F{t} = función densidad acumulada de t = $\int_0^t f(x)dx$

Si T es el tiempo desde que sucedió el último evento:

$$P\begin{Bmatrix} el\ tiempo\ entre\ eventos \\ no\ es\ menor\ que\ T \end{Bmatrix} = P\begin{Bmatrix} no\ ocurren\ eventos \\ durante\ T \end{Bmatrix}$$

$$P\{t \geq T1\} = p_0(T)$$

$$\int_T^\infty f(t)dt = e^{-\alpha T}$$

$$1 - F(T) = e^{-\alpha T}, \qquad T > 0$$

La distribución obtenida es la DISTRIBUCIÓN EXPONENCIAL!!

$$f(T) = \alpha e^{-\alpha T}, \qquad T > 0$$

El resultado genera tres conclusiones:

1) Para el proceso que describen las probabilidades de $p_n(t)$, el tiempo para la incidencia de eventos sucesivos debe seguir una distribución exponencial

2) El valor esperado de la distribución exponencial es:

E{T} = $\frac{1}{\alpha}$ unidades de tiempo

y

$\frac{1}{E(T)} = \alpha$ eventos/unidad de tiempo h

es la tasa de generación de eventos.

3) La distribución exponencial padece **olvido o falta de memoria**, propiedad fundamental para demostrar en un sistema de colas las características de *aleatoriedad completa*.

Proceso de Llegada

El objetivo es determinar la distribución de los procesos de llegada durante un intervalo de tiempo.

$p_n(t + h)$ = P (n llegadas durante t y ninguna durante h

o bien

n-1 llegadas durante t y una durante h)

$$p_n(t+h) = p_n(t) * p_0(h) + p_{n-1}(t) * p_1(h), \qquad n = 1,2,3,\ldots$$

$$p_0(t+h) = p_0(t) * p_0(h), \qquad n = 0$$

Si llamamos λ a la tasa de llegada (número de llegadas por unidad de tiempo), y hacemos **h** lo suficientemente pequeña (h —> 0):

$$p_n'(t) = \frac{p_n(t+h) - p_n(t)}{h} \approx -\lambda p_n(t) + \lambda p_{n-1}(t), \qquad n = 1,2,3,\ldots$$

$$p_0'(t) = \frac{p_0(t+h) - p_0(t)}{h} \approx -\lambda p_0(t), \qquad n = 0$$

Por lo que:

$$p_n(t+h) \approx p_n(t)(1 - \lambda h) + p_{n-1}(t)(\lambda h), \qquad n = 1,2,3,\ldots$$

$$p_0(t+h) \approx p_0(t)(1 - \lambda h), \qquad n = 0$$

La distribución del número de llegadas es POISSON !!!, con Media = Varianza = λt

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Conclusión:

- Distribución del Tiempo entre Llegadas =EXP (λ)

  Esperanza = 1/λ

- Distribución del Número de Llegadas en t = POI (λ)

  Esperanza = λt

Procesos de salida

Las distribuciones de los procesos de salida se derivan suponiendo un número inicial N de clientes en el sistema que quieren salir una vez servidos a un ritmo de unidades por unidad de tiempo.

Si

$q_n(t)$ = probabilidad de que ocurran n salidas en t

entonces, resultarían del estudio las mismas ecuaciones, cambiando α por μ.

$$q_0(h) = e^{-\mu h} \cong 1 - \mu h,$$

$$q_1(h) = 1 - \ q_0(h) \cong \mu h,$$

o lo que es lo mismo:

$$q_N(t+h) \cong q_N(t) * 1 + q_{N-1}(t) * \mu h, \qquad n = N$$

$$q_n(t+h) \cong q_n(t) * (1 - \mu h) + q_{n-1}(t) * \mu h, \qquad 1 \leq n < N$$

$$q_0(t+h) \cong q_0(t) * (1 - \mu h), \qquad n = 0$$

y resolviendo las ecuaciones:

$$q'_N(t) = \mu q_{N-1}(t), \qquad n = N$$

$$q'_n(t) = -\mu q_n(t) + \mu q_{n-1}(t), \qquad 1 \leq n < N$$

$$q'_0(t) = -\mu q_0(t), \qquad n = 0$$

Entonces:

$$q_n(t) = \frac{(\mu t)^n}{n!} e^{-\mu t}$$ n = 0, 1, 2 …, N-1

$$q_N(t) = 1 - \sum_{n=0}^{N-1} q_n(t)$$ n = N-

Si lo que queremos hallar es:

$p_n(t)$ = probabilidad de que queden n clientes después de t

$p_n(t) = q_{N-n}(t)$

o sea,

$$p_n(t) = \frac{(\mu t)^{N-n}}{(N-n)!} e^{-\mu t}$$ n = 0, 1, 2 …, N-1

$$p_0(t) = 1 - \sum_{n=1}^{N} p_n(t)$$ n = N-

POISSON !!!, con Media = Varianza = μ t

Características de un sistema

Las condiciones previas sólo sirven para introducir los sistemas de colas. En la realidad, dichas condiciones pueden cambiar.

Las 6 características esenciales de un sistema de colas son:

a) Distribución de llegada de clientes

b) Distribución de los tiempos de servicio

c) Número de servidores en paralelo

d) La disciplina del servicio

e) Número máximo en el sistema (cola + servicio)

f) Tamaño de la fuente de clientes

La representación del sistema será:

(a/b/c):(d/e/f)

Descripción de los parámetros

1) (**a/b**/c):(d/e/f)

Las distribuciones de probabilidad de los procesos de entrada y salida pueden variar y tomar los siguientes valores:

- M = distribución de llegadas o salidas de Poisson (o markoviana); o, lo que es lo mismo, distribución exponencial entre llegadas o de tiempo de servicio
- D = tiempo entre llegadas o de servicio constante o determinista
- E_k = distribución de Erlang o gamma de la distribución de tiempo entre llegadas o de servicio con el parámetro k
- GI = distribución independiente general de llegadas (o tiempo entre llegadas)
- G = distribución general de salidas (o tiempo de servicio)

2) (a/b/**c**):(d/**e**/**f**)

El número de puestos de servicio **c** es finito pero puede variar

$C \in \{1, \ldots, \infty\}$

Lo mismo sucede con el espacio disponible (**e**) así como la población objetivo (**f**)

3) (a/b/c):(**d**/e/f)

Las prioridades en el servicio pueden ser del tipo:

- GD = General Discipline
- FCFS = First Come First Served (FIFO)
- LCLS = Last Come First Served (LIFO)
- SIRO = Service in Random Order (Aleatorio)

Medidas u objetivos-

El comportamiento del sistema en su régimen permanente se mide a través de las siguientes variables:

- ρ = relación entre λ y $\mu = \lambda/\mu$
- p_n = Probabilidad de n clientes en el sistema
- L_s = Número medio de clientes en el sistema
- L_q = Número medio de clientes en la cola
- W_s = Tiempo medio de espera en el sistema
- W_q = Tiempo medio de espera en la cola

Sabiendo la distribución de los procesos de llegada y de salida, donde

λ es el parámetro de la distribución de llegada

μ es el parámetro de la distribución de salida

Entonces:

$\rho=\lambda/\mu$

Podemos definir los valores de L_s y L_q como:

$$L_s = \sum_{n=0}^{\infty} np_n$$

$$L_q = \sum_{n=c}^{\infty} (n-c)p_n$$

y, por tanto, de W_s y W_q como:

$$L_s = \lambda\, W_s$$

$$L_q = \lambda\, W_q$$

El problema estriba en: ¿Qué ocurre si el sistema tiene capacidad finita? La tasa de llegada se convertirá entonces en:

Tasa efectiva = λ_{eff}

=tasa efectiva de llegadas para los que se unan al sistema= $\beta\lambda$, $0 <\beta<1$

Si

tiempo de espera estimado en el sistema

= tiempo de espera estimado en la línea de espera

+ tiempo estimado de servicio

entonces:

$$W_s = W_Q + \frac{1}{\mu}$$

$$L_s = L_Q + \frac{\lambda}{\mu}$$

Para cada combinación de los seis parámetros, entonces se procede como sigue para cualquier sistema de colas:

$$p_n \rightarrow L_s = \sum_{n=0}^{\infty} np_n \rightarrow W_s = \frac{L_s}{\lambda} \rightarrow W_q = W_s - \frac{1}{\mu} \rightarrow L_q = \lambda\, W_q$$

(M/M/1):(GD/∝/∝)

a	Poisson
b	Poisson
c	1
d	disciplina general en asignación de clientes a servidores
e	sin límite de capacidad
f	infinito número de posibles clientes

Recordamos que, según la distribución exponencial que rige los procesos de llegadas y de salidas:

$$P(0\ llegadas\ en\ h) = e^{-\lambda h} \cong 1 - \lambda h$$

$$P(1\ llegadas\ en\ h) = 1 - e^{-\lambda h} \cong \lambda h$$

$$P(0\ salidas\ en\ h) = e^{-\mu h} \cong 1 - \mu h$$

$$P(1\ salidas\ en\ h) = 1 - e^{-\mu h} \cong \mu h$$

Si sólo puede ocurrir un evento en h, para tener n elementos en t+h:

$$p_n(t+h) \approx p_n(t)(1-\lambda h)(1-\mu h) + p_{n-1}(t)(\lambda h)(1-\mu h) + p_{n-1}(t)(1-\lambda h)(\mu h),$$

$$n = 1,2,3,\ldots$$

$$p_0(t+h) \approx p_0(t)(1-\lambda h)(1) + p_1(t)(1-\lambda h)(\mu h), \qquad n = 0$$

Si despejamos $p_n(t)$, dividimos por h y hacemos h — > 0:

$$p_n'(t) = \lambda p_{n-1}(t) + \mu p_{n+1}(t) - (\lambda + \mu)p_n(t), \qquad n = 1,2,3,\ldots$$

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t), \qquad n = 0$$

Si:

$$\frac{\partial P_n(t)}{\partial t} \rightarrow 0$$

$$P_n(t) \rightarrow P_n$$

entonces:

$$0 = -\lambda P_0 + \mu P_1 \qquad 0 = \lambda P_0 - \mu P_1$$

$$0 = \lambda P_0 - (\mu + \lambda) P_1 + \mu P_2 \qquad 0 = \lambda P_0 - \mu P_1 - (\lambda P_1 - \mu P_2)$$

$$0 = \lambda P_1 - (\mu + \lambda) P_2 + \mu P_3 \qquad 0 = \lambda P_1 - \mu P_2 - (\lambda P_2 - \mu P_3)$$

$$. \; . \; . \; . \qquad . \; . \; . \; .$$

$$0 = \lambda P_{n-1} - (\mu + \lambda) P_n + \mu P_{n+1} \qquad 0 = \lambda P_{n-1} - \mu P_n - (\lambda P_n - \mu P_{n+1})$$

O simplificando:

$$\lambda P_0 = \mu P_1$$

$$\lambda P_1 = \mu P_2$$

$$\lambda P_2 = \mu P_3$$

$$. \; . \; . \; .$$

$$\lambda P_n = \mu P_{n+1}$$

Y resumiendo

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0$$

y dado que:

$$\sum_{n=0}^{\infty} p_n = 1 = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n p_0 = p_0 \sum_{n=0}^{\infty} \rho^n = p_0 \left(\frac{1}{1-\rho}\right) = 1$$

$$p_n = (1-\rho)\rho^n$$

$$p_0 = (1-\rho)$$

Para hallar, por ejemplo, el número medio de unidades en el sistema:

$$L_s = \sum_{n=0}^{\infty} n\, p_n = \sum_{n=0}^{\infty} n\,(1-\rho)\rho^n = (1-\rho)\rho \frac{\partial}{\partial \rho} \sum_{n=0}^{\infty} \rho^n = (1-\rho)\rho \frac{\partial}{\partial \rho}\left(\frac{1}{1-\rho}\right) = \frac{\rho}{1-\rho}$$

Resumiendo, y para régimen permanente estable (ρ < 1):

$$p_n = (1-\rho)p^n, \qquad n = 0,1,2,\ldots$$

$$L_s = \frac{\rho}{1-\rho}$$

$$L_q = L_s - \frac{\lambda}{\mu} = \frac{\rho^2}{1-\rho}$$

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu(1-\rho)}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\rho}{\mu(1-\rho)}$$

(M/M/1):(GD/N/∝)

a —> Poisson
b—> Poisson
c —> 1
d—> disciplina general en asignación de clientes a servidores
e—> límite de capacidad en N unidades
f—> infinito número de posibles clientes

$$\lambda_{eff} = \mu(L_s - L_q)W_s = \lambda(1-p_N)$$

$$p_n = \begin{cases} \left(\dfrac{1-\rho}{1-\rho^{N+1}}\right)\rho^n, & \rho \neq 1 \\ \dfrac{1}{N+1}, & \rho = 1 \end{cases} \qquad n = 0,1,2,\ldots,N$$

$$L_s = \begin{cases} \dfrac{\rho\{1-(N+1)\rho^N + N\rho^{N+1}\}}{(1-\rho)(1-\rho^{N+1})}, & \rho \neq 1 \\ \dfrac{N}{2}, & \rho = 1 \end{cases}$$

$$L_q = L_s - \frac{\lambda_{eff}}{\mu} = L_s - \frac{\lambda(1-p_N)}{\mu}$$

$$W_s = \frac{L_s}{\lambda_{eff}} = \frac{L_s}{\lambda(1-p_N)}$$

$$W_q = \frac{L_q}{\lambda_{eff}} = \frac{L_q}{\lambda(1-p_N)}$$

(M/M/c):(GD/∝/∝)

$\frac{\rho}{c} < 1 \; o \; \frac{\lambda}{\mu c} < 1,$

a —> Poisson
b —> Poisson
c —> c servidores en paralelo
d —>disciplina general en asignación de clientes a servidores
e —> sin límite de capacidad
f —> infinito número de posibles clientes

$$p_0 = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!\,(1 - {}^{\rho}/_{c})} \right\}^{-1}$$

$$p_n = \begin{cases} \left(\dfrac{\rho^n}{n!}\right) p_0, & 0 \leq n < c \\ \left(\dfrac{\rho^n}{c^{n-c} c!}\right) p_0, & n > c \end{cases}$$

$$L_q = \frac{\rho^{c+1}}{(c-1)!\,(c-\rho)^2}\, p_0 = \left[\frac{c\rho}{(c-\rho)^2}\right] p_0$$

$$L_s = L_q + \rho,\, W_s = \frac{L_s}{\lambda},\, W_q = \frac{L_q}{\lambda}$$

El modelo se puede aproximar de la siguiente forma:

- Para ρ<<1
  - $p_0 \cong 1 - \rho$
  - $L_q \cong \frac{\rho^{c+1}}{c^2}$
- Para ρ/c=1
  - $p_0 \cong \frac{(c-\rho)(c-1)!}{c^c}$
  - $L_q \cong \frac{\rho}{c-\rho}$

(M/M/c):(GD/N/∝), c≤N

a —> Poisson
b —> Poisson
c —> c servidores en paralelo
d —> disciplina general en asignación de clientes a servidores
e —> con límite de capacidad en N unidades
f —> infinito número de posibles clientes

$$p_n = \begin{cases} \frac{\rho^n}{n!} p_0, & si\ n < c \\ \frac{\rho^n}{c!\, c^{n-c}} p_0, & si\ c \le n \le N \end{cases}$$

$$p_0 = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c \left(1 - \left(\frac{\rho}{c}\right)^{N-c+1}\right)}{c!\,(1 - \rho/c)} \right\}^{-1}$$

$$L_q = \begin{cases} p_0 \frac{\rho^{c+1}}{(c-1)!\,(c-\rho)^2} \left\{1 - \left(\frac{\rho}{c}\right)^{N-c} - (N-c)\left(\frac{\rho}{c}\right)^{N-c}\left(1 - \frac{\rho}{c}\right)\right\}, & \rho/c \ne 1 \\ p_0 \frac{\rho^c (N-c)(N-c+1)}{2c!}, & \rho/c = 1 \end{cases}$$

$$L_s = L_q + (c - \bar{c}) = L_q + \frac{\lambda_{eff}}{\mu}$$

(M/GD/1):(GD/∝/∝)

Fórmula de Pollaczek-Khintchine

a —> Poisson
b —> General —> Media = E{t}; Varianza = var {t}
c —> 1
d —> disciplina general en asignación de clientes a servidores
e —> sin límite de capacidad
f —> infinito número de posibles clientes

$$\rho = \lambda E\{t\} < 1$$

$$L_s = \lambda E\{t\} + \frac{\lambda^2 (E^2\{t\} + var\{t\})}{2(1 - \lambda E\{t\})}$$

$$L_q = L_s - \lambda E\{t\}$$

$$W_s = \frac{L_s}{\lambda}$$

$$W_q = \frac{L_q}{\lambda}$$

(M/M/∝):(GD/∝/∝)

Modelo Self-Service

a —> Poisson
b —> Poisson
c —> infinito número de posibles servidores (=clientes)
d —> disciplina general en asignación de clientes a servidores
e —> sin límite de capacidad
f —> infinito número de posibles clientes

$$p_0 = \frac{1}{1 + \rho + \frac{\rho^2}{2!} + \cdots} = \frac{1}{e^\rho} = e^{-\rho}$$

$$p_n = \frac{e^{-\rho}\rho^n}{n!} \qquad n = 0,1,2,\ldots \qquad Poisson(\rho)$$

$$L_s = E(n) = \rho$$

$$W_s = \frac{1}{\mu}$$

(M/M/R):(GD/K/K), R<K

Modelo de Reparación, con R reparadores y K máquinas

a —> Poisson
b —> Poisson
c —> R reparadores
d —> disciplina general en asignación de clientes a servidores
e —> K máquinas
f —> K máquinas

$$\bar{R} = valor\ esperado\ de\ reparadores\ ociosos = \sum_{n=0}^{R} (R-n)p_n$$

$$\lambda_{eff} = \mu(R-\bar{R}) = \lambda(K-L_s)$$

$$p_0 = \frac{1}{1+\rho+\frac{\rho^2}{2!}+\cdots} = \frac{1}{e^{\rho}} = e^{-\rho}$$

$$p_0 = \left\{\sum_{n=0}^{n} \binom{K}{n} \rho^n + \sum_{n=R+1}^{K} \binom{K}{n} \frac{n!\,\rho^n}{R!\,R^{n-R}}\right\}^{-1}$$

$$p_n = \begin{cases} \binom{K}{n} \rho^n p_0, & 0 \le n \le R \\ \binom{K}{n} \frac{n!\,\rho^n}{R!\,R^{n-R}} p_0, & R \le n \le K \end{cases}$$

$$L_q = \begin{cases} K - \left(1+\frac{1}{\rho}\right)(1-p_0), & R = 1 \\ \sum_{n=R+1}^{K} (n-R)\,p_n, & R > 1 \end{cases}$$

$$L_s = \begin{cases} K - \left(\frac{1-p_0}{\rho}\right), & R = 1 \\ L_q + (R-\bar{R}) = L_q + \frac{\lambda_{eff}}{\mu}, & R > 1 \end{cases}$$