Sistemas de Línea de Espera

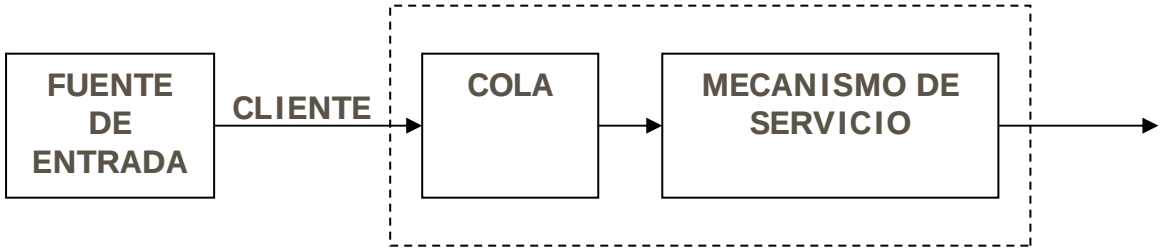
Sabaroni, Andrea

Garello Torres, Melina Valeria

Firmapaz, Maximiliano

Caif, Pablo

Los **clientes** que requieren un servicio se generan a través del tiempo en una fuente de entrada. Estos clientes entran al sistema de cola y se unen a una cola. En determinado momento se selecciona un miembro de la cola mediante una regla desconocida llamada **disciplina de cola**. Posteriormente en un **mecanismo de servicio** se lleva a cabo el servicio requerido por el cliente, después de lo cual éste sale del sistema de cola.



SISTEMA DE COLAS

Componentes

Fuente de entrada

- La caracteriza su tamaño (cantidad de clientes que pueden requrerir un servicio en determinado momento) lo que se denomina **población de entrada**

Cola

- Se caracteriza por el número máximo permisible de clientes que puede admitir.
- Pueden ser finitas o infinitas (estándar)

Disciplina de la Cola

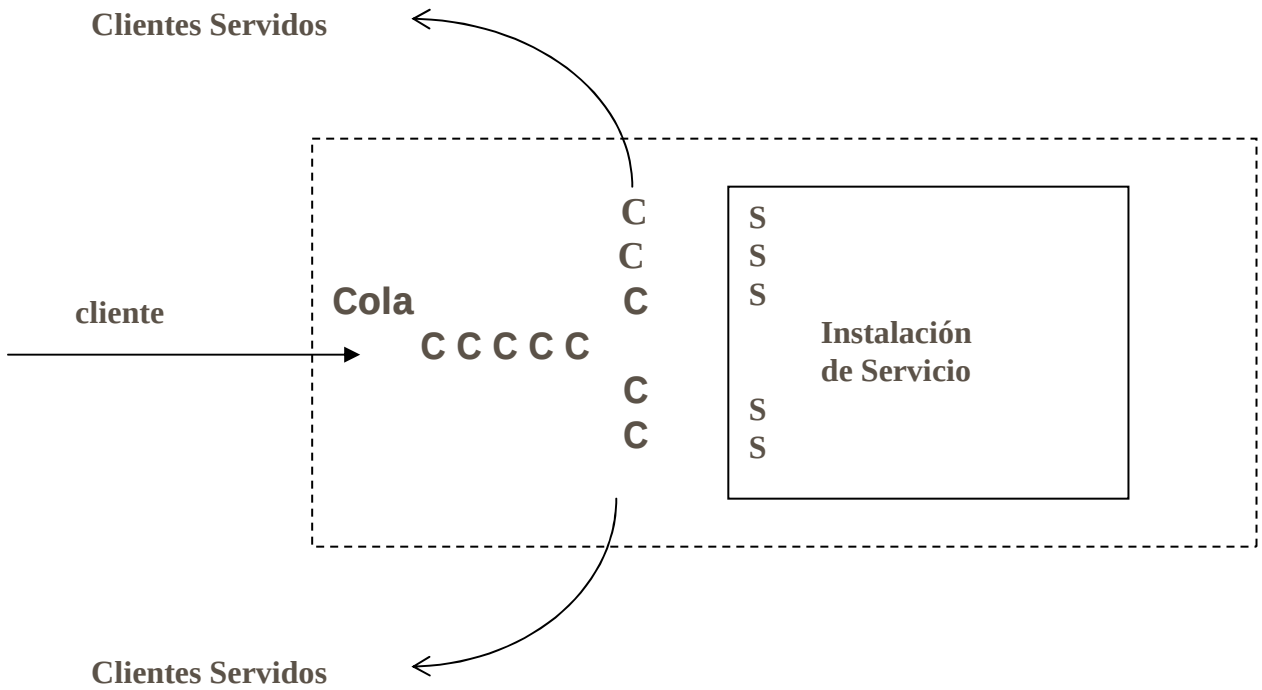
- Se refiera al orden al que seleccionan sus miembros para recibir el servicio.

Mecanismo de Servicio

- Consiste en una o más instalaciones de servicio, cada una de ellas con uno o más canales paralelos de servicios, llamados **servidores**. El cliente entra por una instalación y el servidor le presta un servicio.
- El tiempo que transcurre desde el inicio de servicio para un cliente hasta su terminación en una instalación se llama **tiempo de servicio**.

PROCESO DE COLA ELEMENTAL

El tipo que más prevalece es el siguiente:
Una sola línea de espera (que puede estar vacía en cierto tiempo) se forma frente a una instalación de servicio, dentro de la cual se encuentra uno o más servidores. Cada cliente generado por una fuente de entrada recibe el servicio de el/los servidores, quizás después de esperar un poco en la cola (línea de espera). No es necesario que se forme físicamente la línea de espera. El único requisito es que los cambios en el número de clientes que esperan un servicio, ocurran como si prevalecieran la situación física que se describe en la figura.



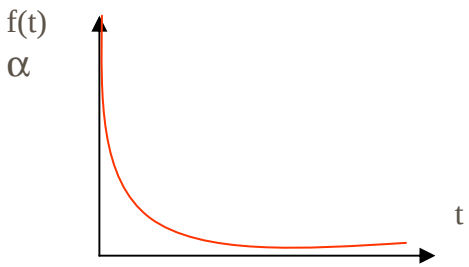
DISTRIBUCIÓN EXPONENCIAL

T= tiempo entre llegadas [o] tipos de servicios

eventos = eventos que marcan el final de estos tiempos

T tiene una distribución exponencial de parámetros α si su función de probabilidad es:

$$f_T(t) = \begin{cases} \alpha e^{-\alpha t} & t >= 0 \\ 0 & t < 0 \end{cases}$$



JEMPLOS DE SISTEMAS DE LÍNEA DE ESPERA

- **Sistema Comercial:**

liente: reciben servicio de 1 organismo comercial.

ervicio: persona -> persona.

eluquería (fija)

eparación de aparatos domésticos (servidor va al cliente).

áquina de monedas (servidor = máquina)

asolinería (cliente = automóviles)

- **Sistema de servicio de Transporte:**

os clientes son automóviles (o aviones)

axis (servidores = automóviles)

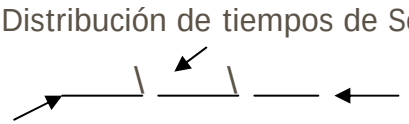
anejo de materiales (servidores = camiones)

istema de mantenimiento (servidores = reparan máquinas de clientes)

ervicios secretariales (servidores = máquinas, clientes = tareas)

- **Sistema de servicios internos en la industria y los negocios:**

PRINCIPALES CARACTERÍSTICAS OPERATIVAS DEL SISTEMA DE COLAS



Tipos de Distribución

- **M**: Distribución Exponencial
- **D**: Distribución Degenerada
- **E_k**: Distribución de Erlang
- **G**: Distribución General

Por ejemplo:

- M/M/s

Tiene tiempo de llegada con distribución exponencial, Tiempo de servicio con distribución exponencial y s Servidores.

- M/G/1

Tiene tiempo de llegada con distribución exponencial, tiempo de servicio con distribución degenerarada (sin restricciones) y 1 Servidor.

TERMINOLOGÍA Y NOTACIÓN

•***Estado del sistema:*** número de clientes en el sistema.

•***Longitud de cola:*** número de clientes que esperan en el servicio [o] estado - número de clientes que están siendo servidos.

•***M(t):*** número de clientes en el tiempo t (t>=0).

•$P_n(t)$**:** probabilidad de que n clientes estén en el tiempo t, dado el número en t = 0.

•***s***: número de servidores.

•λ_n**:** tasa media de llegada de nuevos clientes cuando hay n clientes**.**

•μ_n**:** tasa media de servicio cuando hay n clientes. Representa tasa combinada a la que todos los servicios ocupados logran terminar sus servicios.

$$\begin{cases} Si\ \lambda_n = cte. \ \rightarrow\ \lambda_n = \lambda \\ Si\ \mu_n = cte. \ \rightarrow\ \mu_n = \mu \end{cases}$$

$$\begin{cases} 1/\lambda = tiempo\ de\ llegada \\ 1/\mu = tiempo\ de\ servicio \end{cases}$$

$\rho = \lambda/s\mu$ **=** fracción de tiempo que los servidores individuales están ocupados*.*
$s\mu$**=** capacidad de servicio del sistema**.**

ESTADOS

ESTADO

Cuando comienza el estado del sistema se ve afectado por el estado inicial y el tiempo de servicio.

Condición Transitoria => Estado Transitorio.
Después es independiente del Estado inicial y tiempo transcurrido.

Condición de estabilidad => conserva la distribución de probabilidad de estado del sistema.

ESTADO ESTABLE (por defecto)

NOTACIÓN EN CONDICIÓN ESTABLE

- **P_n:** probabilidad de que n clientes estén en el sistema.
- **L:** número esperado de clientes.
- **L_q:** longitud esperada de cola.
- ***W*** : tiempo de espera en el sistema para cada cliente.
- **W :** E(*W*).
- ***W*** **:** tiempo de espera en la cola para cada cliente.
- ***W*** **:** E(W).

RELACIÓN ENTRE L, W, L_q Y W_q

λ = nº promedio de llegadas por unidad de tiempo

μ = nº promedio de clientes atendidos por unidad de tiempo por servidor.

$W = W_q + 1/\mu$

$L = \lambda * W$

$L_q = \lambda * W_q$

SISTEMA DE COLAS

Para formular un modelo de teoría, especificar forma supuesta para cada distribución. Para que sea útil debe ser suficientemente realista para hacer predicciones razonables y sencillas para que sean matemáticamente manejable.

CARACTERÍSTICAS OPERATIVAS

- llegadas: Distribución de probabilidad de los tiempos entre
- servidores: Distribución de probabilidad de los tiempos de

PROCESO DE NACIMIENTO Y MUERTE

Nacimiento: llegada de un nuevo cliente al sistema de colas.

Muerte: salida del cliente servido.

Estado del sistema: Nt / t >= 0 : número de clientes que hay en el momento t.

Este proceso describe en términos probabilísticos Nt.

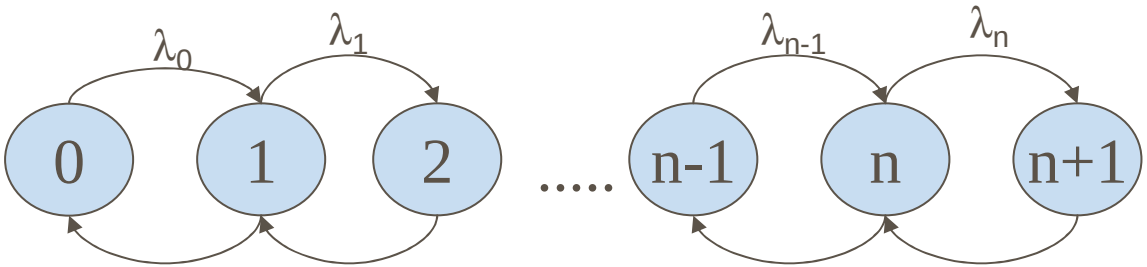
Suposiciones:

1º- $N_T = N$ Distribución probabilidad de tiempo que falta para el próximo nacimiento es exponencial con parámetro α .

2º- $N_T = N$ Distribución probabilidad de tiempo que falta para el próximo nacimiento es exponencial con parámetro μ.

3º- Solo un nacimiento o muerte a la vez.

Estado:



Flecha = transición

Datos = Tasa media de transferencia.

CARACTERÍSTICAS DE UN SISTEMA M/M/1

1. Una población de clientes finita
2. Un proceso de llegada en el que los clientes se presentan de acuerdo con una distribución de Poisson con una tasa promedio de λ clientes por unidad de tiempo.
3. Un proceso de colas que consiste en una sola línea de espera de capacidad infinita, con una disciplina de colas de primero en entrar, primero en salir.
4. Un proceso de servicio que consiste en un solo servidor que atiende a los clientes de acuerdo a una distribución exponencial con un promedio de μ clientes por unidad de tiempo.

Probabilidad de que no haya clientes en el sistema (P_0):	$P_0 = 1-\rho$
Número promedio en la fila (L_q):	$L_q=\rho^2/1-\rho$
Tiempo promedio de espera en la cola (W_q):	$W_q= L_q/\lambda$
Tiempo promedio de espera en el sistema(W):	$W=W_q + 1/\mu$
Número promedio en el sistema(L):	$L=\lambda * W$
Probabilidad de que un cliente que llega tenga que esperar (p_w):	$p_w = 1- P_0= \rho$
Probabilidad de que haya n clientes en el sistema(P_n):	$P_n = \rho^n * P_0$
Utilización(U):	$U = \rho$

MODELO M/G/1

Suposiciones

- iene un servidor y un proceso de entradas Poisson con una tasa media de llegada fija λ.
- os clientes tienen tiempos de servicios independientes con la misma función de probabilidad, pero no se imponen restricciones sobre cual debe ser esta distribución de tiempos de servicios. Solo es necesario conocer la media $1/\mu$ y la varianza σ^2 distribución.
- uede alcanzar una condición estable si $\rho = \lambda / \mu < 1$

Resultados

$$P_0 = 1 - \rho$$

$$W_q = L_q / \lambda$$

$$L_q = (\lambda^2\sigma^2 + \rho^2) / [2 (1 - \rho)]$$

$$W = W_q + 1 / \mu$$

$$L = \rho + L_q$$

$1 / \mu$, L_q, L, W_q y W se incrementan cuando σ^2 aumenta (ya que indica su velocidad promedio y que la consistencia del servidor tiene mucha trascendencia en el desempeño de la instalación del servicio.

Si la distribución de tiempos es exponencial, $\sigma^2 = 1 / \mu^2$.

MODELO M/D/S

Cuando el servicio consiste básicamente en la misma tarea rutinaria que el servicio realiza para los clientes, tiende a haber poca variación el tiempo de servicio requerido.

•Supone que todos los tiempos de servicio en realidad son iguales a una constante fija (distribución de tiempos de servicios degenerada)

• También se supone que tiene un proceso de entrada Poisson con tasa media de llegada fija λ.

•Cuando se tiene un solo servidor, el modelo M/D/1 es un caso especial del modelo M/G/1 donde $\sigma^2 = 0$, por lo tanto:

$L_q = \rho^2 / [2 (1 - \rho)]$

Para mas de un servidor se dispone de un método complicado y se dispone de gráficas.

MODELO $M/E_K/S$

Supone una **variación cero en los tiempos de servicios** ($\sigma = 0$). La distribución exponencial de tiempos de servicios supone una variación muy grande ($\sigma = 1 / \mu$). En estos casos extremos existe un intervalo ($0 < \sigma > 1 / \mu$).La función densidad de probabilidad para la distribución Earlang es:

$f(t) = [(\mu k)^2 / (k - 1)!]. t^{k-1} e^{-k\mu t}$ para $t >= 0$

μ y k: parámetros positivos, k es entero

media = $1 / \mu$

desviación estándar = $1 / k^{1/2} . 1 / \mu$

k:especifica el grado de variabilidad de los tiempos de servicio con relación a la media.

El tiempo requerido para realizar cierto tipo de tareas puede incluir una secuencia de k tareas. Será de distribución **Earlang si el servidor debiera realizar la misma tarea exponencial k veces para cada cliente**. Es útil debido a su gran familia de distribuciones que permiten solo valores no negativos.

La exponencial y la degenerada son casos especiales de Earlang con k = 1 y K = ∞ respectivamente.

Si aplicamos $\sigma 2 = 1 / k\mu 2$ (en el modelo M/G/1)

$W_q = [(1 + k) / 2k] . \{\lambda / [\mu (\mu - \lambda)]\}$

$W = W_q + 1 / \mu$

$L = \lambda W$

Para varios servidores no ha sido posible determinar una solución general de entrada estable. Estos resultados se han obtenido tabulando para casos numéricos.

MODELOS DE COLAS CON DISCIPLINAS DE PRIORIDAD

La disciplina de cola se basa en sistema prioritario. El orden en que se seleccionan los clientes para darle el servicio está basado en sus prioridades asignadas. Con frecuencia proporciona un refinamiento bien aceptado en comparación con otros métodos.

Casi todos los resultados corresponden al caso de un servidor.

· Supone que **existen N clases de prioridad** (la 1 la más alta y la n la más baja). Si un servicio está libre para comenzar un nuevo cliente, **selecciona el de prioridad más alta**, y existe una cola dentro de cada prioridad.

· Supone un **proceso de entrada Poisson** y **tiempos de servicio exponencial** para cada clase prioritaria.

· El **tiempo medio de servicio es el mismo** para clase prioritaria, pero permite que la tasa media de llegadas difiera entre ellas.

· Si se ignora la distribución de clientes es un modelo M/M/s.

· Entonces las fórmulas de L y L_q también sirven al igual que W y W_q para un cliente elegido aleatoriamente.

· **Lo que cambia es la distribución de tiempos**, ya que **tiene una varianza menor los de mayor prioridad** que lo de menor prioridad.

· Se desea **mejorar las medidas de desempeño para cada cliente de prioridad alta,** a costo del desempeño de las clases de prioridad baja. Para determinar la mejora, es necesario obtener estas medidas en término de tiempo de espera esperado en el sistema, y número esperado de clientes en el sistema para las clases de prioridades individuales.

Prioridad

- Sin Interrupciones
  - No se puede interrumpir el servicio de un cliente para mandarlo a la cola si llega al sistema un cliente de prioridad más alta.
- Con Interrupciones
  - Se interrumpe el servicio del cliente de prioridad mas bajo (se expulsa y regresa a la cola) cuando entra al sistema un cliente de prioridad más alta.

REDES DE COLAS

Son redes de instalación de servicios en las que los clientes solicitan el servicio de algunas o todas ellas.Tienen un proceso de entradas Poisson y servicio exponencial.

Propiedad de equivalencia

Se tiene N servidores, un punto de entrada Poisson con parámetro λ y la misma distribución de los tiempos de servicio para cada servidor con parámetro μ donde s$\mu > \lambda$. Entonces la salida en estado estable de cada instalación de servicio es también un proceso Poisson con parámetro λ.

Colas infinitas en serie

Todos los clientes deben recibir servicio en una serie m de instalaciones en una secuencia fija. Cada instalación tiene una cola infinita, por lo tanto en serie forman un sistema de colas infinitas en serie. Los clientes llegan a la primera instalación del sistema de acuerdo a un proceso Poisson con parámetro λ y cada instalación i tiene la misma distribución de servicios exponenciales con parámetro $\mu_i > \lambda$. Por la propiedad de equivalencia cada instalación tiene una entrada Poisson con parámetro λ. Se puede usar el modelo elemental M/M/s.

ANÁLISIS DE COSTOS DEL SISTEMA DE COLAS

- **Un costo basado en el tamaño del personal**

Costo por hora para c/reparador
* Nº de reparadores

Costo total de personal por hora

- **Un costo por hora basado en el nº de maquinas fuera de operación**

Costo por hora por maquina fuera de operación
* Nº promedio de maquinas fuera de operación

Costo total de por la espera

ANÁLISIS DE COSTOS DEL SISTEMA DE COLAS

M/M/7 : M / M / C	
QUEUE STATISTICS	
Number of identical servers	70,000
Mean arrival rate	250,000
Mean service per server	40,000
Mean server utilization (%)	892,857
Expected number of customers in queue	58,473
Expected number of customers in system	120,973
Probability that a customer must wait	0,7017
Expected time in queue	0,2339
Expected time in system	0,4839

	NÚMERO DE REPARADORES				
	7	8	9	10	11
Utilización	892,857	781,250	694,444	625,000	568,182
Número esperando en la cola	58,473	14,936	0,5363	0,2094	0,0830
Número esperando en el sistema	120,973	77,436	67,863	64,594	63,330
Probabilidad de que un cliente tenga					
que esperar	0,7017	0,4182	0,2360	0,1257	0,0632
Tiempo esperado en la cola	0,2339	0,0597	0,0215	0,0084	0,0033
Tiempo esperado en el sistema	0,4839	0,3097	0,2715	0,2584	0,2533

Costo total = costo del personal + costo de la espera

=(50*7) + (100 * 12.07973)

= $ 1559.73