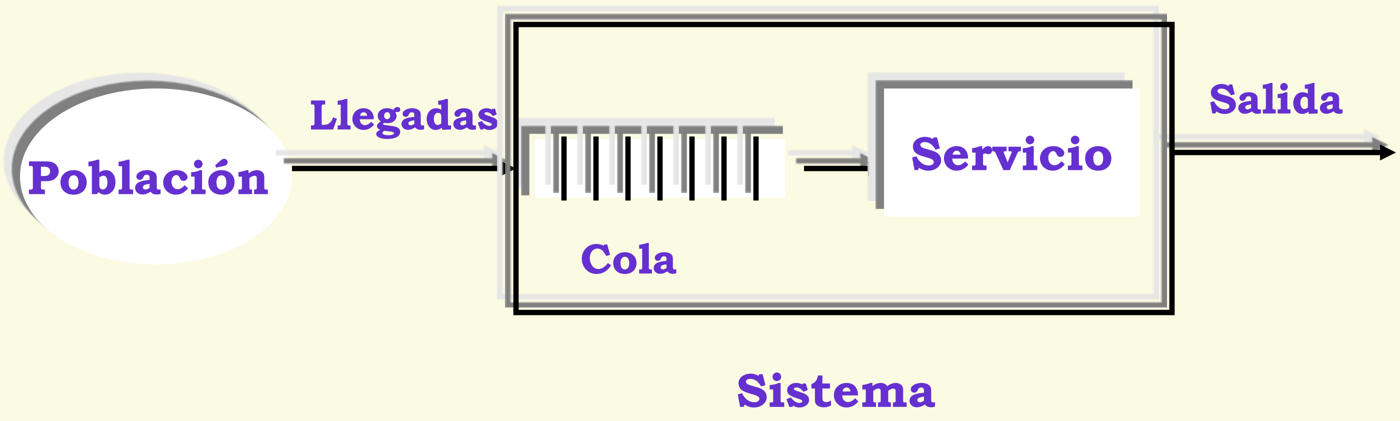
COMPONENTES DE UN SISTEMA DE COLAS

Los elementos de un sistema de colas son:

1.- Población de clientes
2.- Llegadas de clientes
3.- Servicio de los clientes
4.- Características de la cola
5.- Parámetros de rendimiento del sistema



1.- POBLACIÓN DE CLIENTES

Población: conjunto de usuarios potenciales del sistema.

Se definen:

λ_k = Tasa de llegadas al sistema cuando en el sistema ya hay k clientes/unidad tiempo= nº medio de llegadas cuando en el sistema ya hay k clientes/ unidad tiempo

λ = Tasa global (efectiva) de llegadas al sistema= nº medio efectivo de llegadas/unidad tiempo

1.- POBLACIÓN DE CLIENTES

Tipos de población:

- Finita: λ_k disminuye al aumentar k
- Infinita: λ_k permanece constante al aumentar k

Ejemplos de tipos de población:

- Finita: Sistema con 10 componentes y un servicio de reparación. Cada vez que un componente falla y va al servicio de reparación, la tasa λ_k disminuye en un 10%
- Infinita: Acceso a un servidor de Internet. La población serían los potenciales millones de personas que pueden conectarse al servidor. Aunque alguno de ellos se conecte, la tasa λ_k permanece constante.

2.- LLEGADAS DE CLIENTES

- Tamaño:
  - * Individuales
  - * En grupo: distribución del tamaño del grupo

- Distribución del tiempo entre llegadas: es el que transcurre entre dos llegadas consecutivas de clientes al sistema. Se considera que son v.a. independientes con la misma distribución.
- Modelos más usuales:
  - * Exponencial (notación M)
  - * Erlang
  - * Hiperexponencial
  - * General (notación G)

2.- LLEGADAS DE CLIENTES

- Exponencial(λ):
  $f(x) = \lambda e^{-\lambda x}, \; x > 0$
  $E(X) = 1/\lambda, \; V(X) = 1/\lambda^2, \; CV(X) = 1$
  Propiedad de falta de memoria: la distribución exponencial es la única distribución continua que verifica:
  $$P(X>t+h / X>t) = P(X>h)$$

- Erlang(n, λ) = Gamma(n, λ) = $\sum Ex(\lambda)$
  $E(X) = n/\lambda, \; V(X) = n/\lambda^2, \; CV(X) = 1/\sqrt{n} < 1$

2.- LLEGADAS DE CLIENTES

- Hiperexponencial: en este modelo con probabilidad p_i se produce una llegada de una fuente con distribución exponencial (λ_i), $\sum p_i=1$.

$$f(x) = p_1 \lambda_1 e^{-\lambda_1 x} + p_2 \lambda_2 e^{-\lambda_2 x} + \ldots + p_n \lambda_n e^{-\lambda_n x}$$

$$E(X) = \sum p_i/\lambda_i \,,\ E(X^2) = \sum p_i\, 2/\lambda_i^2 \,,\ CV(X) > 1$$

Sirve para modelizar situaciones en las que hay distintos tipos de clientes con distintas tasas de llegada y cada llegada es de tipo exponencial

3.- SERVICIO (I)

- Servidores: Puede haber uno o varios servidores. Llamamos S al nº de servidores.

- μ_K : Tasa de servicio cuando hay k clientes en el sistema = nº medio de clientes servidos en el sistema por unidad de tiempo.

- μ : tasa de servicio de cada servidor = nº medio de clientes servidos en cada servidor por unidad de tiempo.

3.- SERVICIO (II)

Los tiempos de servicio de los clientes, en cualquiera de los servidores, son v.a. independientes, con la misma distribución. Las distribuciones más habituales son:

* Exponencial (μ) (Notación: M)
* Erlang(n, λ)
* Constante (redes ATM)
* Pareto (c, α)
* General (Notación: G)

3.- TRÁFICO EN INTERNET (I): Distribución de Pareto

- Se utiliza como modelo del tamaño de los mensajes que circulan por Internet (para $1 < \alpha < 2$), y como consecuencia de ello también para el tiempo de procesamiento de dichos mensajes.

- Su función de densidad es:

$$f(t) = \alpha\, c^{\alpha}\, t^{-(\alpha+1)} \quad \text{si } t > c$$
$$= 0 \quad \text{resto}$$

La media es $c\,\alpha/(\alpha-1)$, y la varianza infinita.

- Tiene "cola pesada": $P\{T>t\} = c^{\alpha}\, t^{-\alpha}$ decrece lentamente cuando t crece mucho.

3.- TRÁFICO EN INTERNET (II): Distribución de Pareto

➢ Comparación de Exponencial y Pareto con la misma media (3):

x	Pareto (1, 1.5) $P(T>x) = x^{-1.5}$	Exponencial (1/3) $P(T>x) = e^{-x/3}$
10	0.032	0.036
30	0.0061	0.000045
100	0.0010	$3.3 \cdot 10^{-15}$
600	0.00007	$1.4 \cdot 10^{-87}$

4.- COLA (I)

- Capacidad :

  finita : se pierden los clientes que al llegar se encuentran la cola llena.

  Infinita: se admiten todos los clientes.

- Posibilidades:

  una cola para cada servidor

  una cola para varios servidores

  varias colas para un servidor

  En este caso hay que indicar como elige el servidor entre las distintas colas, por ejemplo de la que tenga más clientes, al azar, con probabilidades prefijadas, ...

4.- COLA (II)

Disciplina: Regla para elegir clientes de una cola.

- FIFO (o FCFS): se sirven por orden de llegada
- LIFO: se sirven por orden inverso de llegada
- SPT: se sirven primero los que menor tiempo de procesamiento necesiten.
- RANDOM: se elige al azar entre los de la cola
- ROUND ROBIN: se procesa una cantidad fija de tiempo para cada cliente, y si no ha terminado se devuelve a la cola.
- PRIORIDADES: se sirve por orden de prioridad, y FIFO entre los de la misma prioridad

5.- PARÁMETROS DE RENDIMIENTO DEL SISTEMA(I)

En un sistema de colas interesan estudiar, cuando menos, cuatro procesos estocásticos:

- $L(t)$= número de clientes en el sistema en el instante t
- $L_q(t)$= número de clientes en la cola en el instante t
- $W(n)$= tiempo de estancia del cliente n-ésimo en el sistema
- $W_q(n)$= tiempo de espera del cliente n-ésimo en la cola

5.- PARÁMETROS DE RENDIMIENTO DEL SISTEMA(II)

- El proceso L(t) cuando los tiempos entre llegadas de clientes al sistema y los tiempos de servicio son exponenciales es un PROCESO DE NACIMIENTO Y MUERTE DE PARAMETRO CONTINUO (marco teórico para el estudio de sistemas de colas).
- El cálculo de las probabilidades de estado para estos procesos en el periodo transitorio es complejo, por lo que nos centraremos en estudiar estos procesos cuando el sistema alcanza el ESTADO ESTACIONARIO.
- En estado estacionario nos interesa estudiar alguna de las características de estos procesos: medias, varianzas, percentiles,...

5.- PARÁMETROS DE RENDIMIENTO DEL SISTEMA(III)

- Nº clientes en el sistema. L= nºmedio clientes en el sistema en estado estacionario.
- Nº clientes en cola. L_q= nº medio clientes en cola en estado estacionario.
- Tiempo total en el sistema. W= tiempo medio en el sistema en estado estacionario.
- Tiempo de espera en cola. W_q= tiempo medio en cola en estado estacionario.
- Probabilidad de pérdida de un cliente: P. Sólo se pierden clientes si la cola es finita.

5.- PARÁMETROS DE RENDIMIENTO (IV)

Tasa de ocupación de cada servidor: $\rho = \phi$ = proporción de tiempo que cada servidor está ocupado.

Si hay un solo servidor: ρ = nº medio de llegadas por unidad de tiempo x tiempo medio de servicio = λ/μ.

Si hay S servidores: $\rho = \lambda/(S\mu)$.