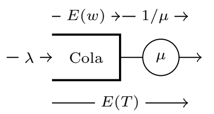
PRM: "Modelado de Tráfico"

3º Ingeniería de Telecomunicaciones — UPV/EHU

"Under-promise and over-deliver."

Javier de Martín – 2016/17

Fundamentos de la Teoría de Colas



λ y μ para una cola de un servidor: $\lambda \to \mu$ la cola se hace inestable, para que sea **estable**: $\lambda < \mu$. Si $\rho \to 1$ los tiempos de espera aumentan y la cola se llena rápidamente.

El **estado de la cola** es el número de usuarios/paquetes en la cola incluido el que pueda estar en el servidor.

Un **proceso de Poisson** es un caso particular del proceso de Markov en el cual la probabilidad de ocurrencia de un evento en $t + \Delta t$ es independiente de sucesos anteriores.

En una **distribución de Poisson** la probabilidad de k llegadas es:

$$p(k) = \frac{(\lambda \cdot T)^k}{k!} \cdot e^{-\lambda T} \quad \rightarrow \quad \sum_{k=0}^{\infty} p(k) = 1$$

La **esperanza** es $E(k) = \lambda \cdot T$ y **varianza** $\sigma^2(k) = \lambda \cdot T$.

Tiempo entre eventos sucesivos, función de densidad de probabilidad:

$$f_\tau(\tau) = \lambda e^{-\lambda\tau} = \frac{\partial F_\tau(x)}{\partial x},\ E(\tau) = \frac{1}{\lambda} \text{ y } \sigma^2(\tau) = \frac{1}{\lambda^2}$$

Relación de los procesos exponenciales con los poissonianos, probabilidad de una llegada en un intervalo:

$$Pr\{\tau > x\} = p_x(0) = e^{-\lambda x}$$

Función de distribución de probabilidad acumulativa, mide la probabilidad de que τ sea menor o igual a x:

$$F_\tau(x) = 1 - e^{\lambda x}$$

Distribución del tiempo de servicio:

$$f_r(r) = \mu \cdot e^{-\mu r}$$

Propiedad: Suponer m procesos de Poisson independientes de tasas λ_1, λ_2, ..., λ_m que se combinan, el proceso resultante será también de Poisson con una tasa $\lambda = \sum_{i=1}^{m} \lambda_i$

Notación Kendall

A/B/m/K/c/z

A: Distribución del tiempo entre llegadas
B: Distribución del tiempo de servicio
m: Número de servidores en paralelo en el sistema
K: Capacidad de la cola (incluyendo servidores)
c: Tamaño de la fuente
z: Disciplina de la cola

Las distribuciones de llegadas y servicio pueden ser:

- **M**: Exponencial
- E_k: Erlang de k etapas
- H_k: Hiperexponencial de k etapas
- **G**: General
- **D**: Determinística (constante)

Por defecto para colas A/B/N la cola es infinita y la disciplina FIFO.

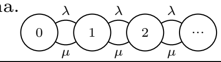
Parámetros medidos en las colas:

Medidas de rendimiento en las colas: $E(n)$, $E(T)$, γ y P_N.

Probabilidad de estado $n(p_n)$: Probabilidad de que haya n usuarios en la cola, incluyendo los que hay en los servidores. Cuando el sistema alcanza su estado más estable se supone que la probabilidad de usuarios en el mismo no varía con el tiempo. Con el tiempo los valores de p_n se irán aproximando a los valores más estables.

M/M/1

Las **ecuaciones de balance** permiten calcular la probabilidad de estado p_n. En el diagrama de estados, el # es el número de clientes en el sistema.



$$\boxed{(\lambda + \mu) \cdot p_n(t) = \lambda \cdot p_{n-1}(t) + \mu \cdot p_{n+1}(t)}$$

$$\begin{cases} (n-1) & \to \text{Ha llegado un usuario} \\ n & \to \text{No ha llegado ni salido nadie} \\ (n+1) & \to \text{Ha salido un usuario} \end{cases}$$

Principio de balance estacionario:
$(\lambda + \mu)p_n$: Tasa de salida de n.
$\lambda p_{n-1} + \mu p_{n+1}$: Tasa de llegada al estado n.

$$\lambda \cdot p_n = \mu \cdot p_{n+1} \quad n = 0, 1, 2... \quad \sum_{n=0}^{\infty} p_n = 1$$

$$p_{n+1} = \rho \cdot p_n$$

$$p_0 = (1 - \rho)$$
$$p_n = (1 - \rho)\rho^n$$

Para una cola M/M/1/N:

$$p_0 = \frac{1 - \rho}{1 - \rho^{N+1}}$$
$$p_n = \frac{1 - \rho}{1 - \rho^{N+1}} \cdot \rho^n$$

La **probabilidad de bloqueo** es $P_N = p_n$, es decir, cuando la última cola se llena.

Tiempo medio por cliente en el sistema:

$$\frac{1}{\mu(1 - \rho)}$$

Fórmula de Little: $\gamma \cdot E(T) = E(n)$.

Número de usuarios en la cola: $E(q) = \lambda E(w)$.

Tiempo medio de espera en cola:

$$\frac{\rho}{\mu(1 - \rho)}$$

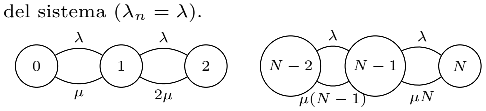
Número medio de usuarios en la cola M/M/1:

$$E(n) = \frac{\rho}{1 - \rho}$$

Cola M/M/N

Permite procesamiento en paralelo al tener múltiples servidores, el factor de utilización es $\rho = \lambda / m\mu$.

La tasa media de llegadas λ es constante e independiente del estado del sistema ($\lambda_n = \lambda$).



$$p_n = \frac{n+1}{N\rho} \cdot p_{n+1} \quad n = 0, ..., N-1$$
$$p_{n+1} = \rho \cdot p_n \quad n \geq N$$

Throughput (γ)

$$\gamma = \lambda \cdot (1 - P_B)$$
$$= \mu \cdot (1 - p_0)$$

Para una **cola infinita**: $\gamma = \mu \cdot \rho = \lambda$.

Para una **cola finita**:

$$P_B = \frac{1 - \rho}{1 - \rho^{N+1}} \cdot \rho^N = P_N$$

Cola M/M/1/N en función de ρ: Si el sistema está lleno, N, no se permite la entrada a nuevos clientes al sistema. La tasa de llegadas no es constante y varía con el tiempo en función de si el sistema está lleno o no:

$$\lambda_{ef} = \lambda(1 - p_k)$$

$$p_n = \rho^n \cdot p_0$$

Para estos sistemas no existe el estado $k + 1$.

$$p_n = \frac{1 - \rho}{1 - \rho^{N+1}} \cdot \rho^n$$

Región de congestión

Throughput Normalizado (γ/μ): Mide la relación entre los usuarios que son atendidos en el sistema por unidad de tiempo y los que potencialmente podrían haber sido atendidos.

$$\frac{\gamma}{\mu} = \frac{1 - \rho^N}{1 - \rho^{N+1}} \cdot \rho = \frac{Usuariosatendidos/seg}{Usuariospotenciales/seg}$$

Insertar Gráfica

Tiempo Medio de Estancia en la Cola E(T)

$$E(T) = \frac{E(n)}{\gamma} \quad \rightarrow \quad E(T)_{M/M/1} = \frac{1}{\mu - \lambda}$$

Colas dependientes del estado: Proceso de nacimiento y muerte

En las **colas dependientes del estado** las tasas de llegada λ y las de servicio μ dependen del estado del sistema, presentes en colas de servidores múltiples: $M/M/m$ y $M/M/1/N$.

$$\lambda = \lambda(n) = \lambda_n \to \text{ Procesos de nacimiento}$$
$$\mu = \mu(n) = \mu_n \to \text{ Procesos de muerte}$$

Un **proceso de nacimiento** tasas de llegada son función del estado. Un **proceso de muerte** salidas son función del estado de la cola.

Ecuaciones de Balance

$$p_n = \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^{n} \mu_i} \cdot p_0 \quad p_0 = \frac{1}{\sum_{n=1}^{N} \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} + 1}$$

Cola M/G/1

Fórmulas de Pollaczek Khinchine para distribuciones M/D/1:

Número medio de usuarios E(n):

$$E(n) = \frac{\rho}{1 - \rho} \cdot \left(1 - \frac{\rho}{2}\right)$$

Tiempo medio de estancia en la cola E(T):

$$E(T) = \frac{E(n)}{\gamma} = \frac{1}{\mu(1 - \rho)} \cdot \left(1 - \frac{\rho}{2}\right)$$

Modelado de Protocolos de N.Enlace

Stop & Wait

Suposiciones para el estudio analítico:

Sólo se permite transmisión de datos `A` → `B`, `B` sólo envía `ACK` o `NACK`.

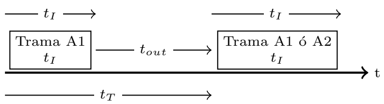
`A` transmite continuamente → saturación.

El cálculo del throughput se realiza utilizando el límite máximo.

Envía cada trama y espera recibir un ACK o NACK:

Si llega un ACK se libera el buffer que almacenaba el paquete transmitido y se envía otra trama.

Si llega un NACK o expira el temporizador, se retransmite la trama.



nº	Tiempo	Probabilidad
1	$t_T = t_I + t_{out}$	$1-p$
2	$2 \cdot t_T$	$p(1-p)$
n	$n \cdot t_T$	$p^{n-1}(1-p)$

t_I: Tiempo requerido para transmitir una trama

t_{out}: Tiempo del temporizador de espera de ACK

$t_{out} \geq 2 \cdot t_p + t_{proc} + t_s$

t_p: Tiempo de retardo de propagación

t_{proc}: Tiempo de procesado en recepción

t_s: Tiempo de transmisión de respuesta

t_T: Tiempo mínimo entre tramas de datos sucesivas: $t_T = t_I + t_{out}$

Tiempo medio para una transmisión correcta:

$$t_v = t_T(1-p) + 2t_T(1-p)p + 3t_T(1-p)p^2 + \ldots =$$
$$= t_T(1-p)\sum_{i=1}^{\infty} i \cdot p^{i-1} = t_T(1-p)\frac{1}{(1-p)^2} =$$
$$= \frac{t_T}{1-p}$$

Throughput máximo ($TH_{real} < TH_{max}$ ← debido a retransmisiones provocadas por tramas erróneas)

$$TH_{max_{S\&W}} = \lambda_{max_{S\&W}} = \frac{1}{t_v} = \frac{1-p}{t_T}$$
$$= \frac{1-p}{a \cdot t_I}(paq/seg) \qquad a = \frac{t_T}{t_I} \geq 1$$

Throughput normalizado real ρ:

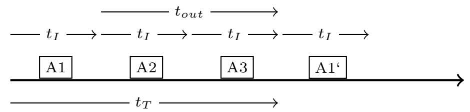
$$\rho \equiv \lambda \cdot t_I \leq \lambda_{max} \cdot t_I = \frac{1-p}{a} < 1$$

Go-Back-N

Envía continuamente las tramas, mejorando el throughput. Es útil para comunicaciones full-duplex.

Se realizan las siguientes **suposiciones**: Los números de secuencia se consideran indefinidos, la retransmisión por `NACK` o t_{out} se estudian igual, transmisión en régimen de saturación de `A` → `B`, longitud de trama t_I fija y temporizador t_{out} fijo y es múltiplo de t_I.

Mínimo tiempo entre transmisiones es el tiempo de transmisión de una trama t_I.



nº	Tiempo	Probabilidad
1	t_I	$1-p$
2	$t_I + t_T$	$p(1-p)$
n	$t_I + (n-1) \cdot t_T$	$p^{n-1}(1-p)$

Tiempo medio de transmisión de una trama:

$$t_V = (1-p)\cdot t_I + (1-p)\cdot p(t_T + t_I) + (1-p)\cdot p^2(2t_T + \cdot t_I) + \ldots =$$
$$= (1-p)\cdot t_I + (1-p)t_T\sum_{i=1}^{\infty} i \cdot p^i + (1-p)\cdot t_I \sum_{i=1}^{\infty} p^i =$$
$$= (1-p)\cdot t_I + a \cdot t_I \cdot (1-p)\frac{p}{(1-p)^2} + t_I \cdot (1-p)\cdot \frac{p}{1-p} =$$
$$= t_I(1-p+\frac{ap}{1-p}+p) = t_I\frac{1-p+ap}{1-p} =$$
$$= t_I\frac{1+(a-1)p}{1-p} \qquad a \equiv \frac{t_T}{t_I} = \frac{t_I + t_{out}}{t_I} = 1 + \frac{t_{out}}{t_I}$$

Throughput máximo:

$$\lambda_{max} = \frac{1}{t_v} = \frac{1-p}{t_I \cdot (1+(a-1)p)}$$

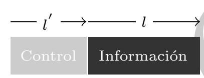
Throughput normalizado:

$$\rho = \lambda \cdot t_I \leq \frac{1-p}{1+(a-1)p} < 1$$

$$\lambda_{max_{GBN}} = a \cdot \lambda_{max_{S\&W}}$$

ACKs embebidos en tramas B-A

Influencia de la tasa de error



Erorres en enlaces de satélite, como **hipótesis** se toma que p_b es la probabilidad de error de bit independiente de la posición del bit debido al ruido aleatorio. Probabilidad de error de trama p:

$p = 1 - (1-p_b)^{l+l'} = 1 - q_b^{l+l'}$. Si $p_b << 1 \to p \approx (l+l')p_b \to p << 1$.

Errores en enlaces terrestres, los errores se producen en ráfagas. p es proporcional a $l + l'$, si $p_b << 1 \to p \approx (l+l')k$.

Tasa de datos normalizada D/C, la estación emisora está en saturación (λ_{max}) y se utiliza el protocolo tipo Go-Back-N.

$$D = \lambda_{max} \cdot l = l \cdot \frac{1-p}{t_I \cdot (1+(a-1))p} \qquad t_I = \frac{l+l'}{C}$$

$C \equiv$ Capacidad del canal

$$\frac{D}{C} = \frac{l}{l+l'} \cdot \frac{1-p}{t_I \cdot (1+(a-1))p} \leq 1$$

$$\frac{D}{C} \approx \frac{l}{l+l'} \cdot (1-p)$$

Cálculo de la longitud óptima

Hipótesis: $a = 1$ (Stop & Wait) o $(a-1)\cdot p \ll 1$.

Caso general de **satélite**:

$$l_{opt} = \frac{l'}{2} \cdot \left(\sqrt{1 - \frac{4}{l' \cdot L_n q_b}} - 1\right)$$
$$\approx \sqrt{\frac{l'}{p_b}} \quad \text{suponiendo } p_b \ll 1 \to p_b \cdot l' \ll 1$$

Caso **terrestre/satélite** simplificado $p = p_b(l+l')$:

$$l_{opt} = \approx \sqrt{\frac{l'}{p_b}} - l'$$

Protocolo HDLC

Longitud de tramas: Información (t_I) y de supervisión (t_S):

$$t_I = \frac{l+l'}{C} \qquad t_S = \frac{l'}{C}$$

Retardo de propagación (incluyendo t_{proc}): t_p

Tiempo de confirmación: $t_{ack} = 2 \cdot t_p + t_s$

Temporizador de reenvío: $t_{out} = 2 \cdot t_p + 2 \cdot t_I > t_{ack}$

TERMINAR

Cálculo del Throughput la **hipótesis** es que cuando ocurre un error se puede recuperar mediante: `REJ` (primer caso) o `TEMPORIZADOR` (siguientes casos).

Throughput máximo:

$$\lambda_{max} = \frac{1}{t_v} = \frac{1-p}{t_I(1+(a-1)p)}$$

p: Probabilidad de recibir una trama con error

T_1: Tiempo aleatorio requerido para la primera retransmisión (puede tomar diferentes valores)

T_2: Tiempo medio requerido para retransmisiones posteriores (es siempre igual en cada transmisión)

t_v: Tiempo medio de transmisión virtual de una trama de longitud t_I.

Tiempo medio de transmisión correcta de una trama:

$$t_V = t_I + E[T_1]p + \frac{T_2 p^2}{1-p}$$

Modelado de N.Red y Control de Flujo

$$\text{Potencia} = \frac{\text{Caudal}^{\alpha}}{\text{Retardo}}$$

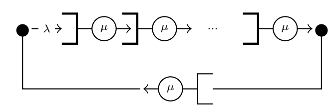
Retardos medios en redes abiertas sin mecanismos de control de flujo:

En CV → $E(T) = \sum_{i=1}^{N} \frac{1}{\mu_i - \lambda_i}$ (ROC)

En Red → $E(T) = 1/\gamma \sum_{i=1}^{N} \frac{\lambda_i}{\mu_i - \lambda_i}$ (NOC)

Mecanismos ROC

Ventana deslizante: CV. Homogéneo (M/M/1 idénticos en serie)



$$u(n, M) = \frac{n}{n+(M-1)} \cdot \mu$$

$\lambda \to \infty$ $\gamma = u(N) = \frac{N}{N+M-1}$ $E(N) = N$

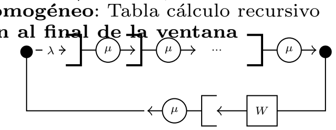
$E(T) = E(n)/\gamma = \frac{N+M-1}{\mu}$

$\lambda \approx \mu$ $\gamma = u(N) = \frac{N}{N+M-1}$ $E(N) = N$

$E(T) = E(n)/\gamma = \frac{N+M-1}{\mu} \cdot \frac{M}{M+1}$

CV No Homogéneo: Tabla cálculo recursivo

Confirmación al final de la ventana



$$n = W - (i+j)$$