Tema 1

Teletráfico

Sistemas de Conmutación

Grado en Ingeniería de Tecnologías de Telecomunicación

3er Curso

Profesores : ***Pablo Ameigeiras Gutiérrez, Pablo Muñoz Luengo***

Departamento de Teoría de la Señal, Telemática y Comunicaciones

E.T.S. Ingenierías Informática y Telecomunicación – Universidad de Granada

C/ Periodista Daniel Saucedo Aranda, s/n - 18071 – Granada (Spain)

Teléfono: +34-958 242306 - Fax: +34-958 243032 - Email: pameigeiras@ugr.es

ción - Grado en Ing. de Tecn. de Telecomunicación - UGR

utiérrez

Motivación y Objetivos

- Motivación
  - Las redes de telecomunicación se caracterizan por la transmisión de datos a través de enlaces
  - Los diferentes mensajes o solicitudes de transmisión compiten por los mismos recursos
  - Ejemplos:
    - Llamadas telefónicas llegan a un conmutador y deben ser encaminadas a través de un conjunto limitado de enlaces de salida
    - Diferentes paquetes de datos han de ser transmitidos a través de un único enlace

Motivación y Objetivos

- El objetivo de la teoría de colas es entender el fenómeno del almacenamiento de los mensajes o peticiones de servicio en cola y ser capaz de predecir indicadores de rendimiento (e.g. el retardo) de manera teórica
- La teoría de colas ejerce un papel clave en ingeniería de redes:
  - Análisis del retardo requerido por un paquete para atravesar un enlace o una red
  - Probabilidad de bloqueo de una llamada telefónica
  - Dimensionado de los elementos de red necesarios para cursar el tráfico ofrecido con una determinada calidad de servicio
  - Evaluación del rendimiento de algoritmos dependientes del retardo

Motivación y Objetivos

- Objetivos
  - Los objetivos del presente tema son los siguientes:
    1. Entender las colas como procesos estocásticos
    2. Aprender la fórmula de Little y su aplicabilidad
    3. Predecir el retardo medio en los sistemas de colas básicos (M/M/1, M/M/m, y M/G/1)
    4. Predecir el retardo medio en redes de colas
    5. Aprender a diferenciar entre sistemas de espera y de pérdidas, así como analizar y predecir el bloqueo en

Sistemas de Conmutación - Grado en Ing. de Tecn. de Telecomunicación - UGR
© Pablo Ameigeiras Gutiérrez

Índice de Contenidos

Sistemas de Conmutación - Grado en Ing. de Tecn. de Telecomunicación - UGR
© Pablo Ameigeiras Gutiérrez

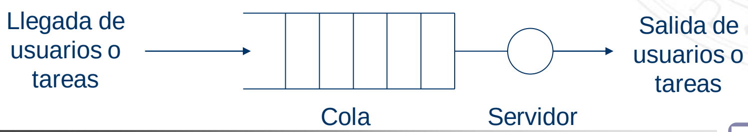
Bibliografía

- D. Bertsekas y R. Gallager: *Data Networks, Prentice Hall, 1992*
- L. Kleinrock : *Queueing Systems, Theory, Volume 1, Wiley, 1975*
- G. Giambene: *Queueing Theory and Telecommunications, Springer, 2005*
- M. Schwartz: *Redes de Telecomunicaciones. Protocolos, Modelado y Análisis, Addison-Wesley, 1994*
- J. Martínez y V. Casares: *Conmutadores de Paquetes, Arquitectura y Prestaciones, Editorial Universidad Politécnica de Valencia, 2001*

1 Colas y Procesos Estocásticos

Colas y Procesos Estocásticos

- Las colas son casos especiales de procesos estocásticos que están representadas por el estado $N(t)$ que indica el número de entidades en el sistema
- La cola está caracterizada por los siguientes factores:
  - El proceso estocástico de llegada de las peticiones de servicio (pej. Poisson)
  - La lista de peticiones de servicio a procesar
  - La disciplina de servicio (pej. FIFO)
  - El proceso estocástico del tiempo de servicio (pej. Constante)
  - El número de servidores
  - El tamaño máximo de la lista de peticiones de servicio



1 Colas y Procesos Estocásticos

- Un proceso estocástico $N(t)$ es una función del tiempo cuyos valores son variables aleatorias $N(t,n)$
- Ejemplos de procesos estocásticos:
  - La temperatura en un determinado punto en función del tiempo
  - La potencia instantánea de ruido en un receptor de comunicaciones
- Un proceso estocástico se caracteriza por:
  - El *espacio de estados* o conjunto de todos los posibles valores que puede tomar $N(t)$
    - espacio continuo
    - espacio discreto → el proceso se denomina *cadena*

1 *Colas y Procesos Estocásticos*

- El comportamiento conjunto del proceso estocástico en los instantes de tiempo t_1, t_2, … t_i está caracterizado mediante la función de distribución de probabilidad conjunta de las variables aleatorias en dichos instantes $\boldsymbol{N}=\{N(t_1), N(t_2),\ldots N(t_i)\}$

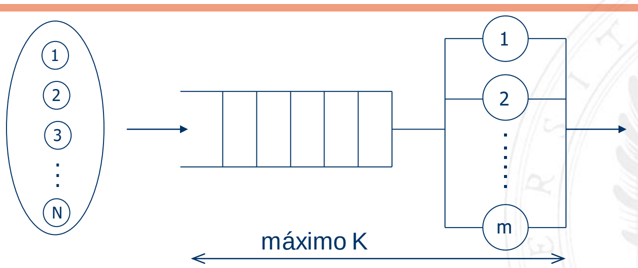
$$F_{\mathbf{N}}(\mathbf{n},\mathbf{t}) = P\{N(t_1) \le n_1, N(t_2) \le n_2,\ldots N(t_i) \le n_i\}$$

- Un proceso estocástico es estacionario en sentido estricto si su función de distribución es invariante a desplazamientos temporales

$$F_{\mathbf{N}}(\mathbf{n},\mathbf{t}) = F_{\mathbf{N}}(\mathbf{n},\mathbf{t}+\tau)$$

- Una Cadena de Markov es un proceso estocástico en el que $N(t)$ sólo puede tomar valores discretos y su estado en el instante t_{i+1}, $N(t_{i+1})$, depende exclusivamente del estado en el instante previo t_i, $N(t_i)$

1 *Notación Kendall*



- A / B / m / K / N / Z
  - A : distribución del proceso de llegadas de peticiones de servicio

1 Colas y Procesos Estocásticos

- m: número de recursos del sistema (idéntica capacidad)
- K: capacidad del sistema en número de usuarios
  - si no se especifica, $K = \infty$
- N: número de fuentes de peticiones
  - si no se especifica, $N = \infty$
- Z: disciplina de gestión de la cola
  - FIFO (First In First Out)
  - LIFO (Last In First Out)
  - RR (Round Robin)
  - Random
  - si no se especifica, Z = FIFO

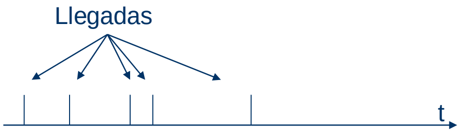
2 Proceso de Poisson

- Introducción
  - El proceso de llegada de Poisson tiene una gran relevancia en el campo de la telecomunicación ya que puede modelar la llegada de distintos tipos de eventos:
    - Llegadas de llamadas a un nodo de conmutación de circuitos de una red telefónica
    - Llegada de sesiones Web a un proveedor de servicio de Internet
    - Llegada de e-mails en una red de paquete de datos
    - Llegada de paquetes en protocolos de acceso aleatorio y en protocolos de acceso no aleatorio
  - Un proceso de Poisson es generalmente considerado como un

2 Proceso de Poisson

Definición

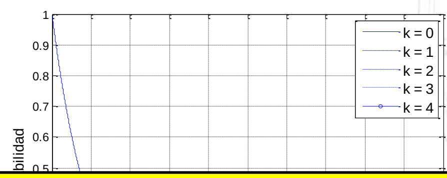
- Un proceso estocástico $\{A(t) | \ t \geq 0\}$ tomando valores enteros no negativos se dice que es un *Proceso de Poisson* con tasa λ si al considerar un pequeño intervalo de tiempo Δt ($\Delta t \rightarrow 0$):
  - La probabilidad de una llegada en el intervalo Δt se define como $\lambda \cdot \Delta t + o(\Delta t)$, siendo $\lambda \cdot \Delta t << 1$
  - La probabilidad de cero llegadas en Δt es $1 - \lambda \cdot \Delta t - o(\Delta t)$
  - Las llegadas son procesos sin memoria: cada llegada en un intervalo de tiempo es independiente de eventos en intervalos previos o futuros
- El proceso de Poisson es un caso particular de una cadena de Markov (proceso de renovación)



2 Proceso de Poisson

- La probabilidad de k llegadas en un intervalo de tiempo finito T sigue una distribución de Poisson:

$$P\{A(t+T) - A(t) = k\} = (\lambda \cdot T)^k e^{-\lambda \cdot T} / k! \qquad k = 0,1,2\ldots$$



2 Proceso de Poisson

- El número medio de llegadas en un intervalo T es:

$$E(k) = \Sigma_k k \cdot P\{ A(t+T) - A(t) = k\} = \lambda \cdot T$$

esto conduce a la interpretación de λ como una tasa de llegada (número medio de llegadas por unidad de tiempo)

- La varianza resulta ser igual a la media:

$$\sigma^2 = E(k) = \lambda \cdot T$$

2 Proceso de Poisson

- Los tiempos entre llegadas son independientes y exponencialmente distribuidos con parámetro λ, es decir, si t_i denota el tiempo de la i-esima llegada, el intervalo $T_i = t_{i+1} - t_i$ sigue una distribución:

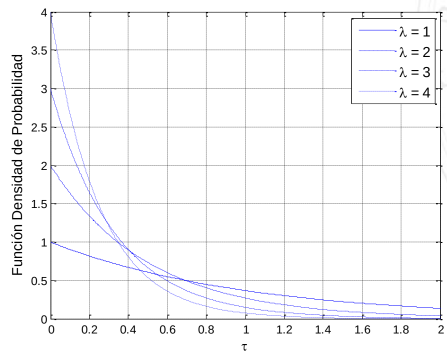
$$P\{T_i \leq \tau\} = 1 - e^{-\lambda \cdot \tau} \qquad \tau \geq 0$$

$$f_{Ti}(\tau) = \lambda\, e^{-\lambda \cdot \tau} \qquad \tau \geq 0$$

- El valor medio de los tiempos entre llegadas cumple:

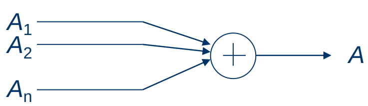
2 Proceso de Poisson

- Función densidad de probabilidad de la distribución exponencial



2 Proceso de Poisson

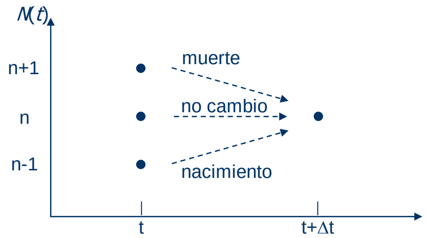
- Si dos o más procesos de Poisson independientes se unen en un único proceso $A = A_1+A_2+A_3+...+A_n$, este proceso es de Poisson y con una tasa de llegadas que iguala la suma de las tasas de cada uno de sus componentes

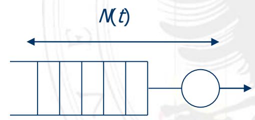


- Si un proceso de Poisson se divide en otros dos procesos asignando de manera independiente cada llegada al primero de dichos procesos con probabilidad p los dos procesos

3 Procesos de Nacimiento y Muerte

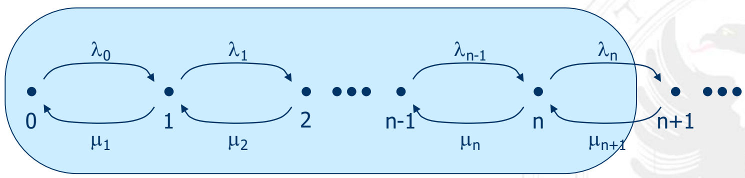
- Un Proceso de Nacimiento y Muerte es una Cadena de Markov en la que las transiciones de estados sólo pueden tener lugar entre estados adyacentes
- Consideremos una Cadena de Markov de tiempo continuo:





- Los procesos de nacimiento y muerte se caracterizan porque el tiempo entre nacimientos y el tiempo entre muertes son exponencialmente distr. ⇒ permiten modelar sistemas M/M/…

3 Procesos de Nacimiento y Muerte

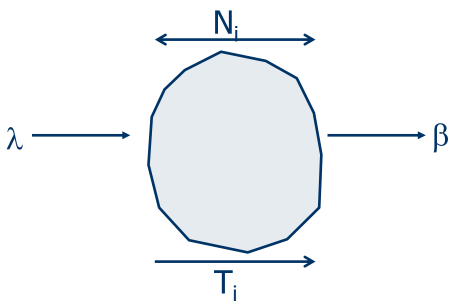


- La cadena cumple la ecuación de balance:

$$\lambda_n \cdot p_n = \mu_{n+1} \cdot p_{n+1}$$

4 El Teorema de Little

- El Teorema de Little expresa la idea natural de que sistemas con gran ocupación están asociados con largas esperas y viceversa



$$T = E[T_i]$$

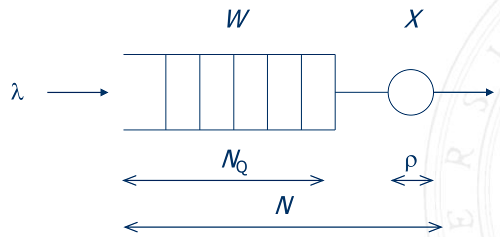
$$N = E[N_i]$$

4 El Teorema de Little

$$\boxed{N = \lambda \cdot T}$$

- Por tanto, el número promedio de usuarios (o paquetes) en un sistema de colas es igual a la tasa de llegada de usuarios (o paquetes) a dicho sistema por el tiempo medio esperado en dicho sistema
- El teorema de Little es de gran relevancia debido a su gran aplicabilidad, ya que es aplicable a casi cualquier sistema de colas que alcance el estado estacionario

4 El Teorema de Little



- Si denotamos N_Q al número medio de usuarios en la cola, ρ el número medio de usuarios en servicio, W el tiempo medio de espera en cola, y X el tiempo medio de servicio, entonces

$$T = X + W \Rightarrow \lambda T = \lambda(X + W) = \rho + N_Q = N$$

5 Análisis del Sistema M/M/1

- Sistema de colas M/M/1
  - Consideremos una cola con un proceso de llegadas de Poisson (tasa media λ), con un tiempo de servicio exponencialmente distribuido (con valor medio $1/\mu$), con un único servidor, con infinitas posiciones en la cola, y con una población de usuarios igual a ∞
  - El sistema M/M/1 se puede modelar como un caso especial de Proceso de nacimiento y muerte en el que $\lambda_i = \lambda$ y $\mu_i = \mu$

5 Análisis del Sistema M/M/1

- De la ecuación de balance:

$$p_n = p_0 \cdot \prod_{i=0}^{n-1} \lambda/\mu = p_0 \left(\lambda/\mu\right)^n = p_0 \cdot \rho^n$$

- La condición de equilibrio exige que $\lambda/\mu < 1$, es decir que la tasa de servicio exceda la tasa de llegadas

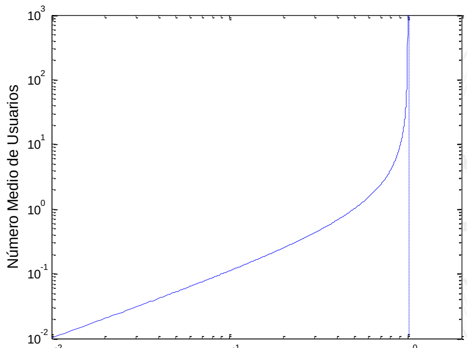
$$\sum_{n=0}^{\infty} p_n = 1 \Rightarrow \sum_{n=0}^{\infty} p_0 \cdot \rho^n = 1 \Rightarrow \frac{p_0}{1-\rho} = 1 \Rightarrow p_0 = 1-\rho$$

donde $\rho = \lambda/\mu$ se define como el Factor de Utilización o probabilidad de que el servidor esté ocupado

- El número medio de usuarios en el sistema se calcula:

$$N = \sum_{n=0}^{\infty} n \cdot p_n = \sum_{n=0}^{\infty} n \cdot \rho^n \cdot (1-\rho) \Rightarrow \boxed{N = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}}$$

5 Análisis del Sistema M/M/1



5 Análisis del Sistema M/M/1

- El tiempo medio de espera por usuario (tiempo de espera en cola más tiempo de servicio) se puede calcular mediante el teorema de Little

$$\boxed{T = N / \lambda = 1 / (\mu - \lambda)}$$

- El tiempo medio de espera en cola W es:

$$W = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\rho}{\mu - \lambda}$$

- Y por el teorema de Little, el número medio de usuario en cola es:

$$N_Q = \lambda W = \frac{\rho^2}{1 - \rho}$$

5 Ejemplo: Retardo en un Nodo de una Red de Paquetes

- Consideremos un nodo de una red que recoge tráfico de paquetes
- El nodo está formado por un buffer y una línea de transmisión
- El tiempo entre llegadas de paquetes sigue una distribución exponencial, y la tasa media de llegada de paquetes al nodo es 100 paq/seg

5 Ejemplo: Retardo en un Nodo de una Red de Paquetes

- Sean las longitudes de los paquetes independientes, y exponencialmente distribuidas
- Se pide:
  - La capacidad de la línea para que el retardo medio de un paquete al atravesar el nodo no exceda de 50 ms
  - El factor de utilización de la línea
  - El número medio de paquetes en el nodo, y el número medio de paquetes en el buffer
  - La probabilidad de que el número de paquetes en el nodo sea mayor o igual que 10
- El retardo medio de un paquete es $T = 1/(\mu - \lambda) \Rightarrow \mu = \lambda + 1/T = 100 + 1/50\text{ms} = 120$ paq/seg
- El factor de utilización de la línea es $\rho = \lambda/\mu = 0.83$
- El número medio de paquetes en el nodo es $N = \lambda\ T = 5$ paq
- El número medio de paquetes en el buffer es $N_Q = N - \rho = 4.17$ paq

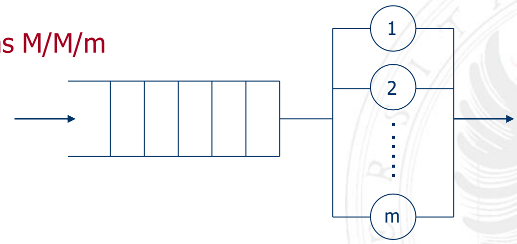
5 Ejemplo: Retardo en un Nodo de una Red de Paquetes

- La probabilidad de que el número de paquetes en el nodo sea mayor o igual que 10 es:

$$\sum_{n=10}^{\infty} p_n = 1 - \sum_{n=0}^{9} p_n = 1 - \sum_{n=0}^{9} p_0 \cdot \rho^n = 1 - p_0 \frac{1-\rho^{10}}{1-\rho} = \rho^{10} = 0.16$$

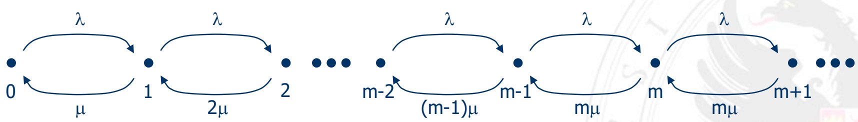
5 Análisis del Sistema M/M/m

- Sistema de colas M/M/m



- El sistema M/M/m es idéntico al M/M/1 pero con un máximo de m servidores (pej. canales en una linea de transmisión) de manera que cada usuario al principio de la cola se enruta al servidor disponible
- El sistema M/M/m también se puede modelar como un proceso de nacimiento y muerte en el que $\lambda_i = \lambda$ y μ_i = min[i μ,mμ]

5 Análisis del Sistema M/M/m



- De las ecuaciones de balance se obtiene:

$\lambda\, p_{n-1} = n \cdot \mu \cdot p_n \qquad n \leq m$

$\lambda\, p_{n-1} = m \cdot \mu \cdot p_n \qquad n \geq m$

5 Análisis del Sistema M/M/m

- Cálculo de p_n para $n \leq m$
  - $\lambda p_0 = \mu p_1 \Rightarrow p_1 = p_0 \cdot \frac{\lambda}{\mu}$
  - $\lambda p_1 = 2\mu p_2 \Rightarrow p_2 = p_1 \cdot \frac{\lambda}{2\mu} \Rightarrow p_2 = p_0 \cdot \left(\frac{\lambda}{\mu}\right)^2 \cdot \frac{1}{2}$
  - $\lambda p_2 = 3\mu p_3 \Rightarrow p_3 = p_2 \cdot \frac{\lambda}{3\mu} \Rightarrow p_3 = p_0 \cdot \left(\frac{\lambda}{\mu}\right)^3 \cdot \frac{1}{2 \cdot 3}$
  - $\lambda p_{n-1} = n\mu p_n \Rightarrow p_n = p_0 \cdot \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1}{2 \cdot 3 \cdot \dots \cdot n} = p_0 \cdot \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1}{n!}$
  - $p_n = p_0 \cdot \left(\frac{\lambda}{m \cdot \mu}\right)^n \cdot \frac{m^n}{n!} = p_0 \cdot \rho^n \cdot \frac{m^n}{n!}$

5 Análisis del Sistema M/M/m

- Cálculo de p_n para $n \geq m$
  - $\lambda p_m = m\mu p_{m+1} \Rightarrow p_{m+1} = p_m \cdot \frac{\lambda}{m \cdot \mu} = p_0 \cdot \rho^m \cdot \frac{m^m}{m!} \cdot \rho \Rightarrow$
    $p_{m+1} = p_0 \cdot \rho^{m+1} \cdot \frac{m^m}{m!}$
  - $\lambda p_{m+1} = m\mu p_{m+2} \Rightarrow p_{m+2} = p_{m+1} \cdot \frac{\lambda}{m \cdot \mu} = p_0 \cdot \rho^{m+1} \cdot \frac{m^m}{m!} \cdot \rho \Rightarrow$
    $p_{m+2} = p_0 \cdot \rho^{m+2} \cdot \frac{m^m}{m!}$

5 Análisis del Sistema M/M/m

- Cálculo de p_0
  - $$\sum_{n=0}^{\infty} p_n = 1 = \sum_{n=0}^{m-1} p_0 \cdot \rho^n \cdot \frac{m^n}{n!} + \sum_{n=m}^{\infty} p_0 \cdot \rho^n \cdot \frac{m^m}{m!}$$
    $$= p_0 \left[\sum_{n=0}^{m-1} \rho^n \cdot \frac{m^n}{n!} + \frac{m^m}{m!} \cdot \sum_{n=m}^{\infty} \rho^n \right] =$$
    $$= p_0 \left[\sum_{n=0}^{m-1} \rho^n \cdot \frac{m^n}{n!} + \frac{m^m}{m!} \cdot \rho^m \sum_{n=m}^{\infty} \rho^{n-m} \right] \Rightarrow$$
  - $$\boxed{p_0 = \left[\left(\sum_{n=0}^{m-1} \rho^n \cdot \frac{m^n}{n!} \right) + \frac{m^m}{m!} \cdot \frac{\rho^m}{1-\rho} \right]^{-1}}$$

5 Análisis del Sistema M/M/m

- La probabilidad de que una llegada encuentre todos los servidores ocupados y tenga que esperar es:

$$P_Q = P\{\text{Queuing}\} = \sum_{n=m}^{\infty} p_n = \sum_{n=m}^{\infty} \frac{p_0 m^m \rho^n}{m!} = \frac{p_0 m^m \rho^m}{m!} \sum_{n=m}^{\infty} \rho^{n-m}$$

y finalmente

$$\boxed{P_Q = \frac{p_0 m^m \rho^m}{m!(1-\rho)}}$$

5 Análisis del Sistema M/M/m

- El número medio de usuarios esperando en cola (pero no en servicio) es:

$$N_Q = \sum_{n=0}^{\infty} np_{m+n} = \sum_{n=0}^{\infty} np_0 \frac{m^m \rho^{m+n}}{m!} = \frac{p_0 m^m \rho^m}{m!} \sum_{n=0}^{\infty} n\rho^n$$

y usando la fórmula de Erlang C para expresar p_0 en función de P_Q

$$N_Q = P_Q \frac{\rho}{(1-\rho)}$$

- El número medio de usuarios esperando en cola coincide con el de una cola M/M/1 multiplicado por la probabilidad de que dicho usuario esté forzado a esperar en cola

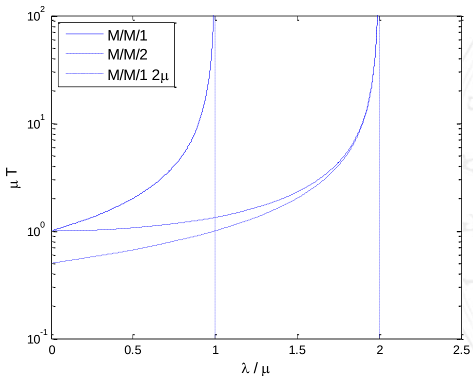
5 Análisis del Sistema M/M/m

- Usando el teorema de Little podemos obtener el tiempo medio que un usuario espera en cola

$$W = \frac{N_Q}{\lambda} = \frac{\rho P_Q}{\lambda(1-\rho)}$$

- Y por tanto el tiempo medio que un usuario espera en el sistema es

5 Análisis del Sistema M/M/m

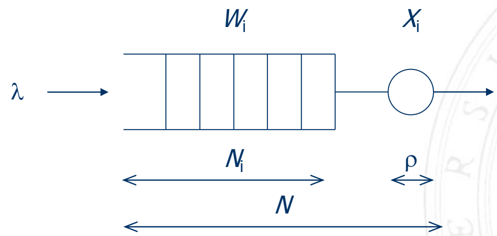


- En condiciones de baja carga, el retardo del sistema M/M/1 con tiempo de servicio $1/m\mu$ se reduce porque el tiempo medio de transmisión de un paquete se reduce por m

6 Análisis del Sistema M/G/1

- Sistema de colas M/G/1
  - Consideremos un sistema de un solo servidor donde los usuarios (o paquetes) llegan según una distribución de Poisson con una tasa λ, pero con un tiempo de servicio que sigue una distribución general (pero conocida)
  - Si el tiempo de servicio no sigue una distribución exponencial, el proceso de servicio tiene memoria
    - Si hay un usuario en servicio en un instante dado, el tiempo residual de servicio tiene una distribución que depende del tiempo transcurrido desde que dicho usuario empezó su servicio

6 Análisis del Sistema M/G/1



- Sea W_i el tiempo de espera en cola del i-esimo usuario (o paquete)
- Sea R_i el tiempo residual de servicio percibido por el i-esimo usuario (o paquete)
- Sea X_i el tiempo de servicio del i-esimo usuario (o paquete)
- Sea N_i el número de usuarios esperando en cola cuando el i-esimo usuario (o paquete) llega al sistema

6 Análisis del Sistema M/G/1

- Entonces:

$$W_i = R_i + \sum_{j=i-N_i}^{i-1} X_j \Rightarrow W = E\{W_i\} = E\{R_i\} + E\left\{\sum_{j=i-N_i}^{i-1} X_j\right\}$$

$$E\{W_i\} = E\{R_i\} + E\left\{\sum_{j=i-N_i}^{i-1} E\{X_j\}\right\} = E\{R_i\} + E\left\{\sum_{j=i-N_i}^{i-1} \frac{1}{\mu}\right\}$$

$$E\{W_i\} = E\{R_i\} + \frac{1}{\mu} \cdot E\left\{\sum_{j=i-N_i}^{i-1} 1\right\} = E\{R_i\} + \frac{N_Q}{\mu} = E\{R_i\} + \frac{\lambda \cdot W}{\mu} \Rightarrow$$

6 Análisis del Sistema M/G/1

- Se puede demostrar (fórmula de Pollaczek-Khinchin):

$$W = \frac{\lambda \overline{X^2}}{2(1-\rho)} \qquad T = \frac{1}{\mu} + \frac{\lambda \overline{X^2}}{2(1-\rho)}$$

o equivalentemente:

$$T = \frac{1/\mu}{1-\rho}\left[1 - \frac{\rho}{2}\left(1 - \sigma_x^2 \mu^2\right)\right]$$

6 Análisis del Sistema M/G/1

- El tiempo medio de espera por usuario (o paquete) equivale al caso en el que el tiempo de servicio sigue una distribución exponencial multiplicado por un factor de corrección
  - El factor de corrección depende de la razón de cambio entre la varianza del tiempo de servicio σ^2 y el valor promedio al cuadrado $1/\mu^2$
  - En el caso de la distribución exponencial $\sigma^2 = 1/\mu^2$ con lo que la ecuación P-K nos da los resultados derivados para el sistema M/M/1
  - Si $\sigma^2 > 1/\mu^2$, el tiempo medio de espera y la ocupación aumentan con respecto al sistema M/M/1
  - Si $\sigma^2 < 1/\mu^2$, el tiempo medio de espera y la ocupación disminuyen con respecto al sistema M/M/1

6 Análisis del Sistema M/G/1

- El comportamiento dominante del tiempo medio de espera es siempre $1/(1-\rho)$
- Todas las colas con tamaño infinito, sin importar su distribución de servicio, tienden a exhibir un comportamiento de congestión cuando $\rho \rightarrow 1$
- Nótese además que la fórmula de Pollaczek-Khinchin es aplicable a cualquier disciplina de servicio siempre que el orden sea independiente del tiempo de servicio

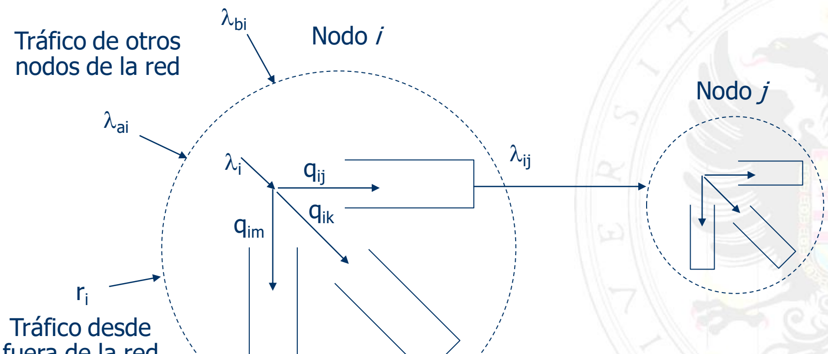
7 Redes de Colas

- Redes de Colas
  - La interconexión de una multiplicidad de nodos que intercambian tráfico de paquetes puede analizarse mediante modelos de Redes de Colas

7 Redes de Colas

- En la red, cada nodo recibe tráfico desde fuera de la red, y a su vez también recibe tráfico proveniente de otros nodos de la red que contribuyen al tráfico total de entrada a dicho nodo
- El tráfico de llegada se divide aleatoriamente entre los diferentes enlaces de salida del nodo
- Cada enlace lo vamos a modelar como una cola (infinita) y una línea de transmisión (servidor)

7 Redes de Colas



7 Redes de Colas

- Dado un conjunto de nodos {1,2,..K} de una red abierta, podemos escribir un sistema de K ecuaciones para encontrar la tasa de llegada a cada nodo:

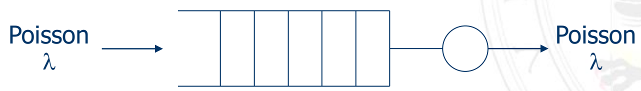
$$\lambda_j = r_j + \sum_{i=1}^{K} \lambda_{ij} = r_j + \sum_{i=1}^{K} q_{ij} \cdot \lambda_i$$

- Para cada nodo se debe cumplir que $\sum_{j=1}^{K} q_{ij} \leq 1$
- Al menos debe existir un nodo para el cual $r_j > 0$
- Igualmente, en al menos un nodo se debe cumplir que el tráfico abandone la red:

$$1 - \sum_{j=1}^{K} q_{ij} > 0$$

7 Redes de Colas

- Teorema de Burke
  - Consideremos un sistema M/M/1, M/M/m o M/M/∞ con una tasa de llegada λ:
    - El proceso de salida de dicho sistema es Poisson con tasa λ



7 Redes de Colas

- Supongamos dos líneas de transmisión de igual capacidad concatenadas, donde la primera se puede modelar como un sistema M/M/1

- Los instantes de llegadas a la segunda cola está fuertemente correlados con el tiempo de servicio en la primera
- Aquellos usuarios (o paquetes) con un tiempo de servicio largo en la primera cola tienden a esperar menos en la segunda cola que los usuarios con un tiempo de servicio corto
- ¡¡La segunda línea no se puede modelar como M/M/1!!
  - Difícil analizar redes de colas con dicha correlación

7 Redes de Colas

- Aproximación de Independencia de Kleinrock
  - La unión de varios flujos de tráfico sobre una misma línea de transmisión tiene un efecto similar a restaurar la independencia entre los instantes de llegada y los tiempos de servicio
  - Si en el ejemplo de las colas concatenadas, la segunda línea de transmisión recibiera una cantidad sustancial de tráfico adicional de Poisson, la correlación de los tiempos de llegada y de servicio en dicha cola se vería considerablemente debilitada

7 Redes de Colas

Teorema de Jackson

- Sea una red de K colas FIFO de un solo servidor cada una y con las siguientes suposiciones:
  - Los usuarios llegan desde fuera de la red a las colas mediante procesos de Poisson independientes y con tasas r_i (algunos r_i pueden ser igual a cero pero al menos uno debe ser mayor de cero)
  - Cuando un usuario es servido en la i-ésima cola, pasa a la j-ésima cola con probabilidad q_{ij} o deja la red con probabilidad $1-\Sigma_j\, q_{ij}$
  - Los tiempos de servicio en la i-ésima cola siguen una distribución exponencial y se asume que son mutuamente independientes e independientes del proceso de llegada a la cola
  - Sea $\rho_i = \lambda_i/\mu_i$ el factor de utilización para cada cola, de manera que $\rho_i < 1$ para todo i
  - Sea además N_i el número de usuarios en la cola i

Redes de Colas

7 Redes de Colas

- Entonces el Teorema de Jackson enuncia que:

$$P\{N_1 = n_1, N_2 = n_2, \ldots N_K = n_K\} = \prod_{i=1}^{K} P\{N_i = n_i\}$$

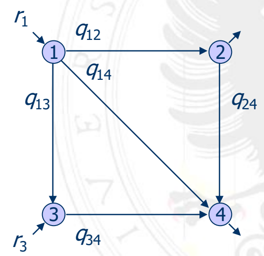
y además:

$$P\{N_i = n_i\} = \rho_i^{n_i}(1-\rho_i), \quad n_i \geq 0$$

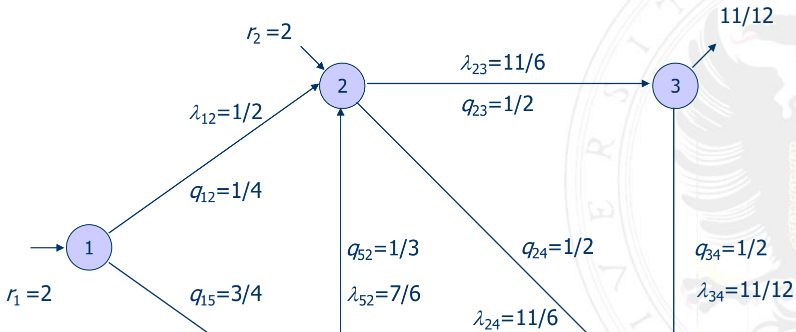
- Según el Teorema de Jackson, las colas de la red se comportan de manera independiente, y además el número de

7 Redes de Colas

- Por el Teorema de Jackson, el retardo medio en la cola (i , j) se puede calcular como $T_{(i,j)} = 1/(\mu_{ij}-\lambda_{ij})$
- El retardo medio experimentado por un flujo de tráfico al atravesar un camino p es $T_p = \sum_{\forall(i,j)\in p} T_{(i,j)}$
- Equivalentemente, el número medio de usuarios en la cola (i , j) es $N_{(i,j)} = \lambda_{ij}/(\mu_{ij}-\lambda_{ij})$
- El número medio de usuarios en la red es $N = \sum_{\forall(i,j)} N_{(i,j)}$



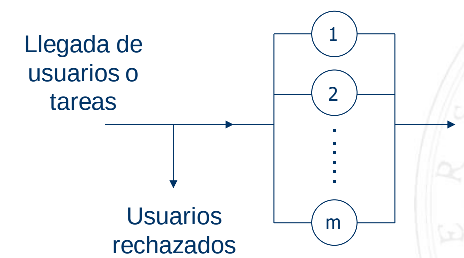
7 Ejemplo: Retardo en Red de 5 Nodos



8 Sistemas de Pérdidas y de Espera

- Sistemas de Pérdidas y de Espera
  - Los sistemas de telecomunicación no se dimensionan para que todos los usuarios se conecten simultáneamente
  - Por ejemplo en telefonía, sólo un pequeño porcentaje (p.ej. 5-8%) de entre todos los usuarios realizan llamadas de manera simultáneamente durante los periodos de mayor tráfico
  - Se explotan las ventajas de la multiplexación estadística para reducir el número de equipos en la red y por tanto su coste
  - Por ello, es posible que un usuario intente llevar a cabo una comunicación a través de la red pero no encuentre recursos disponibles y tenga que esperar o sea bloqueado
  - No obstante, cada usuario debería percibir un servicio como si tuviera un acceso ilimitado a los recursos del sistema de telecomunicación

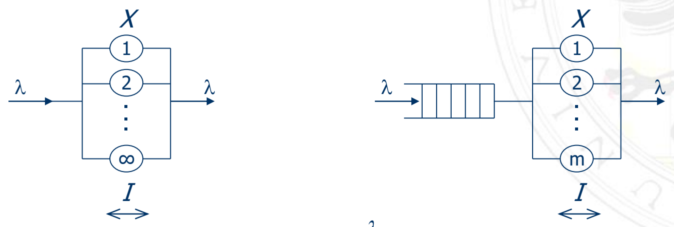
8 Sistemas de Pérdidas y de Espera



- **Sistemas de espera**: Si un usuario llega y todos los recursos están ocupados el usuario espera su turno en cola (capacidad ilimitada)

8 Sistemas de Pérdidas y de Espera

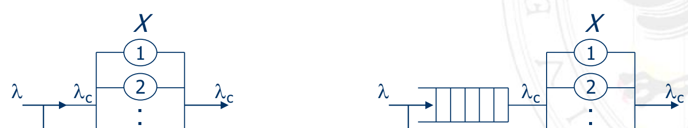
- Intensidad de Tráfico Ofrecido (I): Número medio de recursos ocupados si el sistema de colas es capaz de atender todas las peticiones de los usuarios
    - Se mide en unidades de tráfico o Erlangs
- En un sistema de espera con mμ > λ, o sistema de pérdidas con m = ∞:



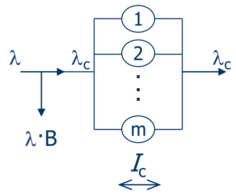
$$I = \lambda \cdot E[X] = \frac{\lambda}{\mu}$$

8 Sistemas de Pérdidas y de Espera

- Intensidad de Tráfico Cursado (I_c): Número medio de recursos ocupados
    - También se mide en unidades de tráfico o Erlangs
- La tasa media cursada λ_c sería el número de usuarios que salen del sistema por unidad de tiempo



8 Sistemas de Pérdidas y de Espera



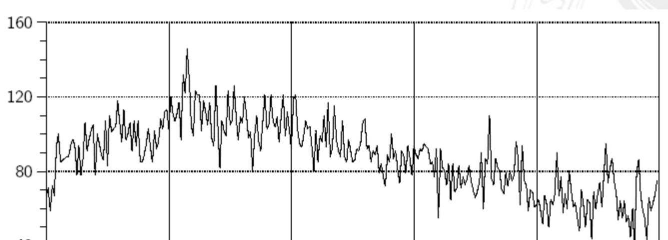
- En un sistema de pérdidas (o mixto), se define la probabilidad de bloqueo B como la probabilidad de que un usuario que llega al sistema sea rechazado por encontrar todos los recursos ocupados

$$\lambda_c = \lambda - \lambda \cdot B = (1 - B) \cdot \lambda$$

$$\boxed{I_c = \frac{\lambda_c}{\mu} = (1 - B) \cdot \frac{\lambda}{\mu} = (1 - B) \cdot I \leq I}$$

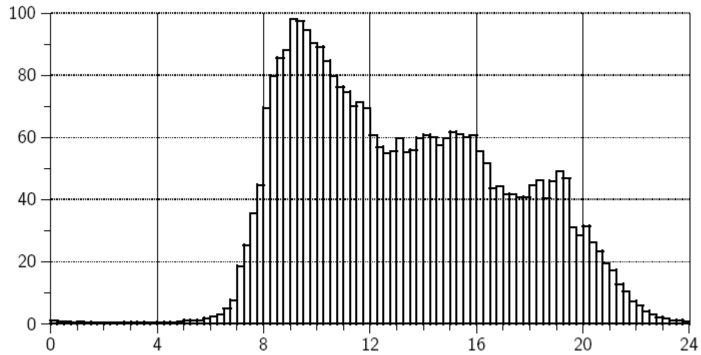
8 Sistemas de Pérdidas y de Espera

- Las variaciones del tráfico tienen una naturaleza parcialmente estocástica y parcialmente determinista



8 Sistemas de Pérdidas y de Espera

- Mediante la comparación de varios días, podemos reconocer la curva determinista



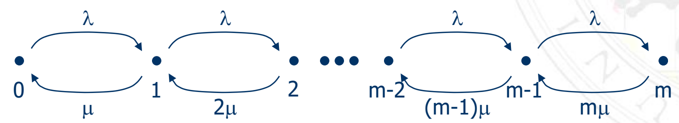
Número de llamadas por minuto promediado a lo largo de 10 días laborables

8 Sistemas de Pérdidas y de Espera

- En telefonía, se suele usar el concepto de *hora cargada*
- La **hora cargada** son aquellos 60 minutos los cuales, al promediar a lo largo de varios días, tienen la intensidad de tráfico más alta del todo el día
- En determinados días, puede suceder que la hora más cargada tenga una intensidad de tráfico mayor que la de la *hora cargada*

8 Sistemas de Pérdidas y de Espera

- Probabilidad de Bloqueo en Sistema de Pérdidas: ErlangB
  - Supongamos un sistema de pérdidas en el que las llegadas se pueden modelar mediante un proceso de Poisson (tasa media λ), y con un tiempo de servicio exponencialmente distribuido $\Rightarrow$ podemos modelar el sistema como un proceso de nacimiento y muerte M/M/m/m



8 Sistemas de Pérdidas y de Espera

- De las ecuaciones de balance podemos obtener:

$$\lambda p_{n-1} = n\ \mu\ p_n \qquad n=1,2,\ldots m$$

- Por tanto:

$$p_n = (\lambda/\mu)^n\ p_0 / n! \qquad n=1,2,\ldots m$$

- Y resolviendo p_0 a partir de $\Sigma_n\ p_n = 1$:

8 Sistemas de Pérdidas y de Espera

- La probabilidad de que la llegada de un usuario encuentre todos los m servidores ocupados (y por tanto se pierda) es:

$$p_m = \frac{(\lambda/\mu)^m / m!}{\sum_{n=0}^{m} (\lambda/\mu)^n / n!}$$

- Esta es la conocida fórmula de Erlang B, o fórmula de Erlang de primer tipo
- Este modelo se ha usado ampliamente como modelo básico de diseño para centrales telefónicas
- El modelo asume que las llamadas bloqueadas se pierden (no se reintentan)
  - En la práctica esto implica que estas llamadas se reintentan después de un cierto tiempo, de manera que pueden ser consideradas como incorreladas a las anteriores (e incluidas en la tasa media λ)
- ¡La fórmula de Erlang B es aplicable a cualquier modelo M/G/m/m!

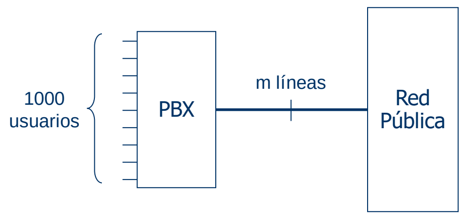
8 Sistemas de Pérdidas y de Espera

Probabilidad de Bloqueo

m	0.007	0.008	0.009	0.01	0.02	0.03	0.05	0.1	0.2	0.4	
1	.00705	.00806	.00908	.01010	.02041	.03093	.05263	.11111	.25000	.66667	1
2	.12600	.13532	.14416	.15259	.22347	.28155	.38132	.59543	1.0000	2.0000	2
3	.39664	.41757	.43711	.45549	.60221	.71513	.89940	1.2708	1.9299	3.4798	3
4	.77729	.81029	.84085	.86942	1.0923	1.2589	1.5246	2.0454	2.9452	5.0210	4
5	1.2362	1.2810	1.3223	1.3608	1.6571	1.8752	2.2185	2.8811	4.0104	6.5955	5
6	1.7531	1.8093	1.8610	1.9090	2.2759	2.5431	2.9603	3.7584	5.1086	8.1907	6
7	2.3149	2.3820	2.4437	2.5009	2.9354	3.2497	3.7378	4.6662	6.2302	9.7998	7
8	2.9125	2.9902	3.0615	3.1276	3.6271	3.9865	4.5430	5.5971	7.3692	11.419	8
9	3.5395	3.6274	3.7080	3.7825	4.3447	4.7479	5.3702	6.5464	8.5217	13.045	9
10	4.1911	4.2889	4.3784	4.4612	5.0840	5.5294	6.2157	7.5106	9.6850	14.677	10
11	4.8637	4.9709	5.0691	5.1599	5.8415	6.3280	7.0764	8.4871	10.857	16.314	11
12	5.5543	5.6708	5.7774	5.8760	6.6147	7.1410	7.9501	9.4740	12.036	17.954	12
13	6.2607	6.3863	6.5011	6.6072	7.4015	7.9667	8.8349	10.470	13.222	19.598	13

8 Ejemplo: Dimensionado del Número de Líneas Necesarias en una PBX

- Consideremos una Private Branch Exchange (PBX) que recoge las llamadas telefónicas de una compañía privada donde hay 1000 usuarios cada uno contribuyendo con una intensidad de tráfico ofrecido de 30mErlangs
- Supongamos que la duración media de la llamada es 3min
- Dimensione el número de líneas necesarias desde la PBX a la oficina central de la red pública para garantizar una probabilidad de bloqueo menor o igual del 3%



8 Ejemplo: Dimensionado del Número de Líneas Necesarias en una PBX

- El número medio de llamadas por usuario es:

$$\lambda_i = I_i \cdot \mu_i = 30 \cdot 10^{-3} \text{ Erlangs x } 1/3 \text{ min} = 10^{-2} \text{ llamadas/min}$$

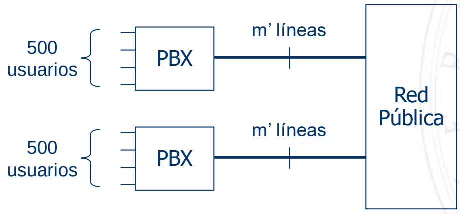
- La intensidad de tráfico total ofrecido es:

$$I = 1000 \cdot I_i = 30 \text{ Erlangs}$$

- El número de líneas para alcanzar un bloqueo del 3% es

8 Ejemplo: Dimensionado del Número de Líneas Necesarias en una PBX

- Supongamos que los 1000 usuarios son servidos mediante dos PBX que recogen las llamadas de 500 usuarios cada una



- $\Gamma' = 500 \cdot 30 \cdot 10^{-3}$ Erlangs = 15Erlangs $\Rightarrow$ m' = 22 para alcanzar un bloqueo menor o igual del 3% $\Rightarrow$ 2m'= 44 líneas > m = 38 líneas
- Con un incremento del número de usuarios del 30% $\Rightarrow \Gamma'$= 19.5Erlangs $\Rightarrow$ m' = 27 para alcanzar un bloqueo menor o igual del 3% $\Rightarrow$ 2m'= 54 líneas

8 Resumen: Sistemas de Pérdidas y Espera

- **Sistemas de Pérdidas**
  - Modelan Sistemas de Conmutación de Circuitos
  - Dimensionado del número de canales: Erlang-B
- **Sistemas de Espera**
  - Modelan Sistemas de Conmutación de Paquetes
  - En C. de Paquetes el proceso de llegada de paquetes no suele ser Poisson y el tiempo de servicio no suele ser exponencialmente distribuido
  - Dimensionado de 1 enlace: M/M/1, M/G/1, y M/M/m como aprox. de primer orden
  - Dimensionado de red: Teorema de Jackson como aprox. de primer orden
  - No obstante, por la naturaleza a ráfagas $\lambda < \mu$. Si $\lambda \to \mu \Rightarrow \rho \to 1$ y $T \to \infty$

Anexo al Tema 1

8 ## Ejemplo: Sistema M/M/1/m

- Consideremos un sistema mixto (de espera y pérdidas) que no permite más de m clientes en el sistema
- Supongamos que proceso de llegadas es de Poisson (tasa media λ), y con un tiempo de servicio exponencialmente distribuido $\Rightarrow$ podemos modelar el sistema como un proceso de nacimiento y muerte M/M/1/m

8 **Ejemplo: Sistema M/M/1/m**

- De la ecuación de balance:

$$p_n = p_0 \cdot \prod_{i=0}^{n-1} \lambda/\mu = p_0\left(\lambda/\mu\right)^n = p_0 \cdot \rho^n \qquad n = 0,1,\ldots m$$

- La condición de equilibrio exige que $\lambda/\mu < 1$, es decir que la tasa de servicio exceda la tasa de llegadas

$$\sum_{n=0}^{\infty} p_n = 1 \Rightarrow \sum_{n=0}^{m} p_0 \cdot \rho^n = 1 \Rightarrow p_0 = \frac{1}{\sum_{n=0}^{m} \rho^n} \Rightarrow p_0 = \frac{1-\rho}{1-\rho^{m+1}}$$

- Sustituyendo:

$$p_n = \frac{1-\rho}{1-\rho^{m+1}} \cdot \rho^n \qquad n = 0,1,\ldots,m$$

Sistemas de Pérdidas y Espera

8 **Ejemplo: Sistema M/M/1/m**

- Por tanto la probabilidad de bloquear un cliente es:

$$B = \frac{1-\rho}{1-\rho^{m+1}} \cdot \rho^m$$

- La intensidad de tráfico cursado es:

$$I_c = \frac{\lambda_c}{\mu} = (1-B) \cdot \frac{\lambda}{\mu}$$