Arquitectura de Redes II
Tema 1

Teoría de Colas

Objetivos del tema

1. Identificar los parámetros de un sistema de colas
2. Enunciar y aplicar el teorema de Little
3. Utilizar la notación de Kendall para identificar los distintos modelos de colas
4. Aplicar los procesos de Poisson para obtener las expresiones de distintos modelos de colas.
5. Identificar y resolver problemas con colas de un único servidor
6. Identificar y resolver problemas con colas de múltiples servidores
7. Utilizar la fórmula de Erlang-C para dimensionar sistemas
8. Identificar y resolver problemas con colas con pérdidas
9. Utilizar la fórmula de Erlang-B para dimensionar sistemas
10. Calcular parámetros de una cola con tiempo de servicio general
11. Enunciar y aplicar el teorema de Burke
12. Enunciar y aplicar el teorema de Jackson
13. Identificar y resolver problemas con redes abiertas de colas

Escuela Politécnica Superior

Contenido

- **Introducción a la teoría de colas**
  - Introducción
  - Definiciones de un sistema de colas
  - Teorema de Little
  - Modelos de sistemas de colas. Notación de Kendall
  - Procesos de nacimiento-muerte
  - Procesos de Poisson

  Bibliografía: Gross 1
- **Cola con un único servidor: M/M/1**
  - Proceso de nacimiento-muerte para M/M/1
  - Cálculo de probabilidades de estados para M/M/1
  - Cálculo de parámetros para M/M/1: valores medios de unidades y tiempos de estancia
  - Distribución del tiempo de estancia para M/M/1

  Bibliografía: Gross 2.1, 2.2

Escuela Politécnica Superior

Contenido

- **Cola con múltiples servidores: M/M/c**
  - Proceso de nacimiento-muerte para M/M/c
  - Cálculo de probabilidades de estados para M/M/c
  - Fórmula de Erlang-C
  - Cálculo de parámetros para M/M/c: valores medios de unidades y tiempos de estancia

  Bibliografía: Gross 2.3
- **Colas con pérdidas: M/M/c/c y M/M/1/k**
  - M/M/c/c
    - Proceso de nacimiento-muerte para M/M/c/c
    - Cálculo de probabilidades de estados para M/M/c/c
    - Fórmula de Erlang-B
    - Cálculo de parámetros para M/M/c/c: valores medios de unidades y tiempos de estancia
  - M/M/1/k
    - Proceso de nacimiento-muerte para M/M/1/k
    - Cálculo de probabilidades de estados para M/M/1/k
    - Cálculo de parámetros para M/M/1/k: valores medios de unidades y tiempos de estancia

  Bibliografía: Gross 2.5, 2.4

Escuela Politécnica Superior

Contenido

- **Colas con tiempo de servicio general: M/G/1**
  - Cálculo de parámetros para M/G/1: valores medios de tiempos de estancia y unidades en el sistema
  - Aproximación a distribuciones exponenciales

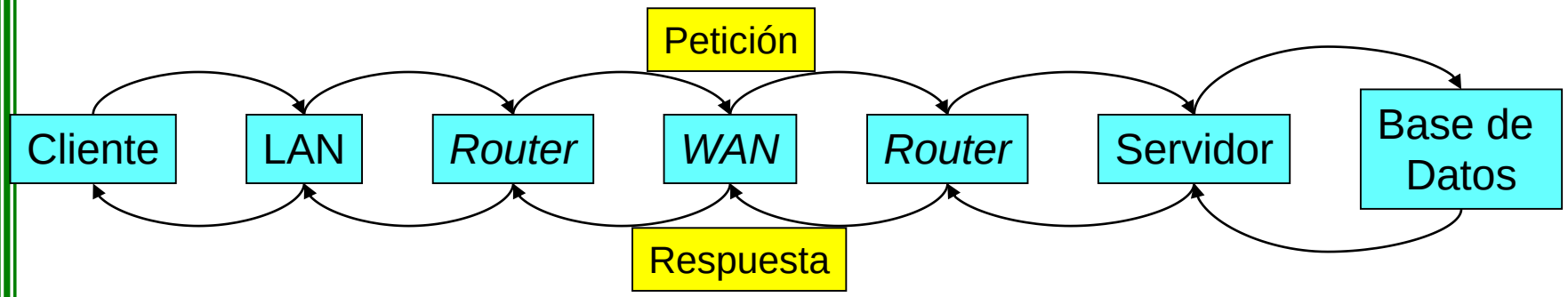
  Bibliografía: Gross 5.1
- **Redes de Colas**
  - Representación de una red de colas
  - Teorema de Burke
  - Teorema de Jackson

  Bibliografía: Gross 4.2

Escuela Politécnica Superior

Introducción

- La transmisión de datos en una red de comunicaciones es el resultado del trabajo de múltiples elementos en secuencia.

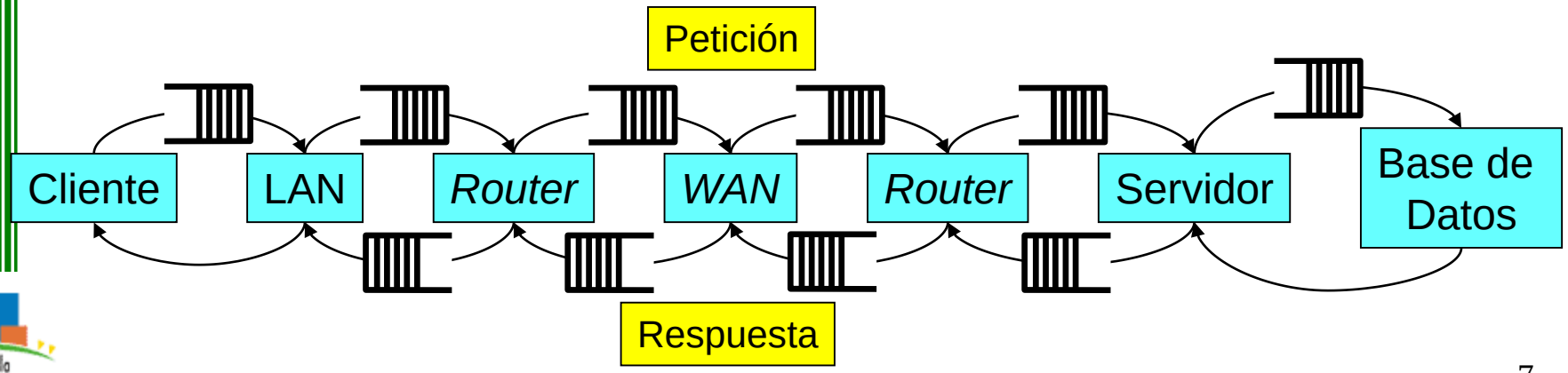


- Estos elementos constituyen la cadena de procesamiento.
- Cada elemento en la cadena tiene una capacidad de procesamiento propia, que debe ser conocida.
- El rendimiento de un sistema distribuido es una característica extremo a extremo (*end-to-end*). Se ve afectado por todos los componentes que intervienen en la cadena de procesamiento.
- El rendimiento total del sistema siempre será peor que el rendimiento del elemento de menor capacidad de procesamiento.

Escuela Politécnica Superior

Introducción

- Algunos elementos de la cadena de procesamiento son de uso común para múltiples elementos.
- Cuando se desea utilizar uno de estos recursos, es posible que se encuentre ocupado en atender a otra petición. Se produce entonces:
  - Un rechazo de la petición. Es necesario reintentar para lograr su ejecución.
  - Una espera a que termine de procesar y quede libre para atender a una nueva petición. Es necesario esperar en una cola.
- En cualquiera de los casos se produce un retardo adicional en la cadena de procesamiento.

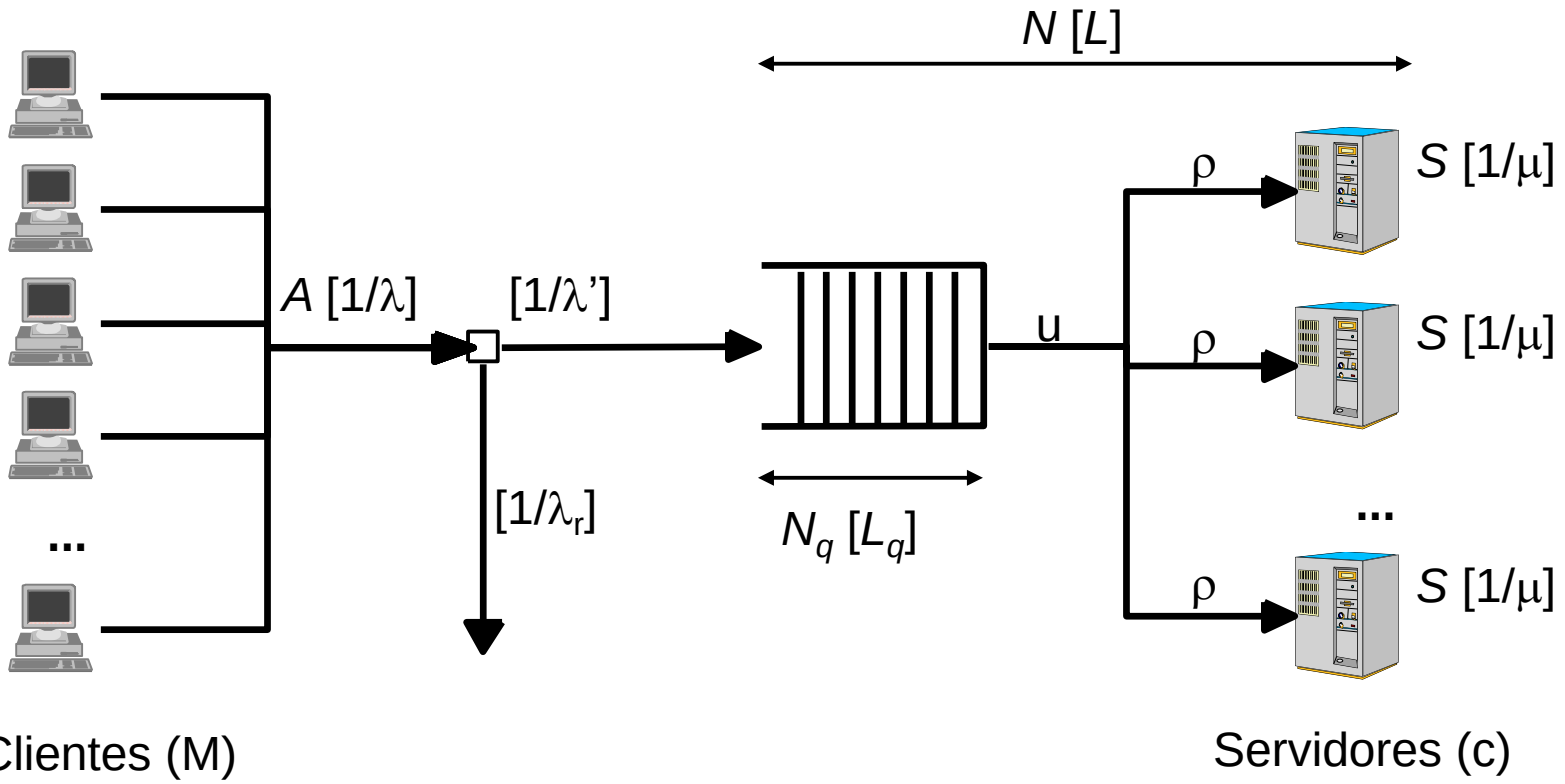


Teoría de colas

- La Teoría de Colas es la rama de la teoría de la probabilidad que estudia los retardos producidos por compartición de recursos entre un determinado número de clientes, y establece modelos para predecirlos.
- Generalmente hay tres tipos de resultados de interés:
  - Tiempo de espera de un cliente para utilizar un recurso.
  - Número de unidades en espera en un momento dado.
  - Una medida del tiempo en que el recurso está siendo utilizado.
- La naturaleza estocástica del problema hace que estas magnitudes sean casi siempre variables aleatorias.
- La teoría de colas proporciona los mecanismos para calcular sus funciones de distribución de probabilidad o, cuanto menos, sus valores esperados.

Escuela Politécnica Superior

Diagrama genérico de un sistema de colas



Definiciones y nomenclatura (I)

- Cliente: Solicitud de servicio recibida en un sistema.
  - La llegada de clientes al sistema es un proceso aleatorio.
  - El tiempo entre llegadas consecutivas es una variable aleatoria, **A**.
    - Se denomina **A(t)** a su función de distribución acumulada.
    - Su valor esperado o valor medio es E[A]=**Ta**, o *tiempo medio entre llegadas*.
    - El número medio de llegadas al sistema por unidad de tiempo, denominado *tasa de llegadas*, será **λ**= 1/Ta.
- Servidor: Elemento del sistema que atiende las solicitudes de servicio (clientes).
  - El tiempo de servicio de un servidor será una variable aleatoria, **S**.
    - Se denomina **S(t)** a su función de distribución acumulada.
    - Su valor esperado o valor medio es E[S]=**Ts**, o *tiempo medio de servicio*.
    - El número medio de clientes servidos por unidad de tiempo, denominado *tasa de servicio* será **μ**= 1/Ts .
- Cola: Elemento intermedio donde esperan los clientes a recibir el servicio.
- Un cliente que entra en el sistema estará siendo servido o esperando en cola.

Escuela Politécnica Superior

Definiciones y nomenclatura (II)

- **u**: Intensidad de tráfico. Relación entre la tasa de llegadas y la tasa de servicio.

$$u = \frac{\lambda}{\mu} = \frac{T_s}{T_a} \qquad \textbf{(1.1)}$$

  Su unidad es el **Erlang**.
- **ρ**: Factor de utilización del servidor. Fracción de tiempo en que se encuentra ocupado un servidor.
  - Equivale a la probabilidad de que el servidor esté activo en un instante dado. $0<\rho<1$
- La ocupación del sistema es un proceso aleatorio. El número de clientes en su estado estable es una variable aleatoria, **N**.
  - Su valor esperado, E[N]=**L**, es el *número medio de clientes en el sistema*.
  - $\mathbf{p_n}$ representa la probabilidad de que en el sistema haya n unidades.
- La ocupación de la cola del sistema es un proceso aleatorio. El número de clientes en su estado estable es una variable aleatoria, $\mathbf{N_q}$.
  - Su valor esperado, $E[N_q]=\mathbf{L_q}$, es el *número medio de clientes en espera*.

Escuela Politécnica Superior

Definiciones y nomenclatura (III)

- El tiempo de estancia en el sistema es una variable aleatoria, T.
  - **W(t)** es su función de distribución acumulada.
  - Su valor esperado, E[T]=**W**, es el *tiempo medio de estancia en el sistema*.
- El tiempo de espera en cola en el sistema es una variable aleatoria, T_q.
  - $\mathbf{W_q(t)}$ es su función de distribución acumulada.
  - Su valor esperado, $E[T_q]=\mathbf{W_q}$, es el *tiempo medio de espera en cola*.
- Entre ambos tiempos medios se verifica la siguiente relación:

$$W = W_q + T_s = W_q + 1/\mu \qquad \textbf{(1.2)}$$

- **Teorema de Little**: Relaciona el número medio de clientes con el tiempo medio de estancia, tanto en el sistema como en cola:

$$L = \lambda' W \qquad \textbf{(1.3)} \qquad\qquad L_q = \lambda' W_q \qquad \textbf{(1.4)}$$

Por tanto, de (1.2), (1.3) y (1.4) se obtiene que:

$$L = L_q + \lambda' T_s = L_q + \lambda'/\mu \qquad \textbf{(1.5)}$$

Escuela Politécnica Superior

Modelos de sistemas de colas

- Los modelos de colas se caracterizan mediante la notación de Kendall A/B/C/X/Y/Z, donde:
  - **A**: indica la distribución de probabilidades asumida para el tiempo entre llegadas.
  - **B**: indica la distribución de probabilidades asumida para el tiempo de servicio.
  - **C**: indica el número de servidores que contiene el sistema.
  - **X**: indica el máximo número de clientes que puede contener el sistema.
  - **Y**: indica la población del sistema.
  - **Z**: Es la disciplina de servicio de la cola.
- Para caracterizar las distribuciones se emplea la siguiente nomenclatura:
  - **M**: Distribución exponencial (Markoviana).
  - **G**: Distribución general (no específica).
  - **D**: Tiempo de servicio constante (Deterministic).
  - $\mathbf{H_n}$: Distribución hiperexponencial de orden n.
  - $\mathbf{E_m}$: Distribución de Erlang-m
- En el caso más habitual, **Y** = ∞, y **Z** = FCFS (first-come, first-served), y se omiten.

Escuela Politécnica Superior

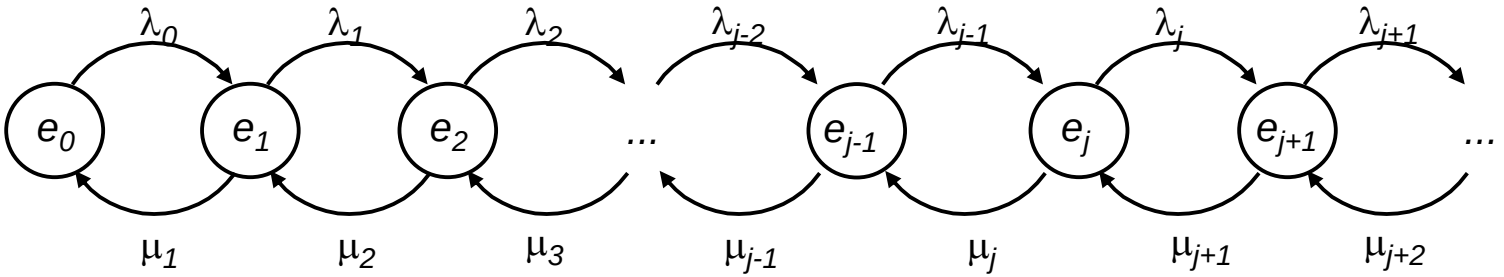
Procesos aleatorios o estocásticos

- Una variable aleatoria x es una función que asigna a cada resultado ζ de un experimento aleatorio S un número, x (ζ).
- Un proceso aleatorio x(t), $t \in T$, es una función que asigna a cada resultado ζ de un experimento aleatorio S una función, x(t, ζ). Habitualmente, T representa el tiempo, y el proceso aleatorio es una familia de funciones temporales con parámetro ζ.
  - Si se fija ζ, x(t) es una función del tiempo, llamada muestra o realización del proceso aleatorio.
  - Si se fija t pero ζ es variable, x(ζ) es una variable aleatoria denominada el estado del proceso aleatorio en el instante t.
- Tipos de procesos aleatorios:
  - Parámetro discreto (*discrete-parameter* o *discrete-time*): T es un conjunto numerable.
  - Parámetro continuo (*continous-parameter* o *continous-time*): $T \in \Re$
  - Estado discreto (*discrete-state*): El conjunto de estados es numerable,
    $\{e_1, e_2, e_3 \ldots e_j \ldots\}$
  - Estado continuo (*continous-state*): Estados toman valor en un conjunto continuo.

Escuela Politécnica Superior

Proceso nacimiento-muerte (I)

- Caso de proceso de Markov (el estado del proceso $x(t_n)$ depende sólo del estado en el instante anterior, $x(t_{n-1})$). Estado: número de individuos.
  - El estado del proceso en el futuro es independiente de su pasado.
  - En un estado e_i, la tasa de nacimientos es λ_i y la tasa de defunciones, μ_i.
  - No se producen nacimientos ni muertes simultáneos.
  - El diagrama de transiciones de estados es de la siguiente forma:



  - Si λ_{jk} es la tasa de paso del estado j al estado k, en este modelo se verifica:

$$\lambda_{jk}\begin{cases}\lambda_j & (k=j+1)\\ \mu_j & (k=j-1,\ j>0)\\ 0 & (k\neq j,\ j+1,\ j-1)\end{cases} \quad \textbf{(1.6)}$$

$$\lambda_{j,j+1}=\lambda_j \qquad \lambda_{j,j-1}=\mu_j \qquad \lambda_{jj}=-(\lambda_{j,j-1}+\lambda_{j,j+1})=-(\lambda_j+\mu_j) \quad \textbf{(1.7)}$$

Proceso nacimiento-muerte (II)

- Al ser un proceso de Markov, para el caso estacionario se obtiene:

$$\begin{cases} \lambda_{j-1,j}\,p_{j-1} - \left(\lambda_{j,j-1} + \lambda_{j,j+1}\right)p_j + \lambda_{j+1,j}\,p_{j+1} = 0 \quad (j > 0) \\ -\lambda_{0,1}p_0 + \lambda_{1,0}p_1 = 0 \end{cases} \tag{1.8}$$

- Introduciendo (1.6) y (1.7) en (1.8):

$$\begin{cases} \lambda_{j-1}p_{j-1} - \left(\lambda_j + \mu_j\right)p_j + \mu_{j+1}p_{j+1} = 0 \quad (j > 0) \\ -\lambda_0 p_0 + \mu_1 p_1 = 0 \end{cases} \tag{1.9}$$

- Reescribiendo las ecuaciones se obtiene la siguiente identidad iterativa:

$$\mu_{j+1}p_{j+1} - \lambda_j p_j = \mu_j p_j - \lambda_{j-1}p_{j-1} = \ldots = \mu_1 p_1 - \lambda_0 p_0 = 0 \tag{1.9}$$

Escuela Politécnica Superior

Proceso nacimiento-muerte (III)

- Despejando y sustituyendo recursivamente se llega a la expresión:

$$p_{j+1} = \frac{\lambda_j}{\mu_{j+1}} p_j = \frac{\lambda_j \lambda_{j-1}}{\mu_{j+1}\mu_j} p_{j-1} = \ldots = \frac{\lambda_j \lambda_{j-1} \ldots \lambda_0}{\mu_{j+1}\mu_j \ldots \mu_1} p_0 \qquad \textbf{(1.10)}$$

$$p_n = \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}} p_0 \quad (n > 0) \qquad \textbf{(1.11)}$$

- p_0 se calcula para que se cumpla el segundo axioma de la probabilidad:

$$\sum_{n=0}^{\infty} p_n = 1 \Rightarrow \left(1 + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}}\right) p_0 = 1 \qquad \textbf{(1.12)}$$

- Por lo tanto, la solución estacionaria de esta cadena de Markov existe si converge la serie infinita de la expresión (1.12).

Escuela Politécnica Superior

Procesos de Poisson (I)

- Un proceso de Poisson es un proceso de Markov que sólo puede cambiar desde el estado e_i al estado e_{i+1} con una probabilidad que es independiente del estado en que se encuentre.
- Representan sucesos que ocurren en instantes aleatorios con las siguientes condiciones:
  - El número de ocurrencias es independiente del tiempo (no existen horas punta).
  - Una nueva ocurrencia del suceso es independiente de sucesos anteriores.
  - La probabilidad de dos ocurrencias simultáneas se puede considerar nula.
- En estas condiciones, si los sucesos ocurren a un ritmo de λ s^{-1}, la probabilidad de n llegadas en t segundos se expresa por:

$$p_n(t) = P\{x(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t};\ \lambda > 0,\ n = 0,1,2... \qquad \textbf{(1.13)}$$

Procesos de Poisson (II)

- El número medio de llegadas en un intervalo t y su varianza vienen dados por:

$$\mathrm{E}[x(t)] = \lambda t \quad \mathrm{Var}[x(t)] = \lambda t \tag{1.14}$$

- La probabilidad de que el tiempo entre llegadas sea menor que t sigue una distribución exponencial, dada por la expresión:

$$A(t) = P\{A \le t\} = 1 - e^{-\lambda t} \tag{1.15}$$

- La función densidad de probabilidad del tiempo entre llegadas será:

$$a(t) = \lambda e^{-\lambda t} \tag{1.16}$$

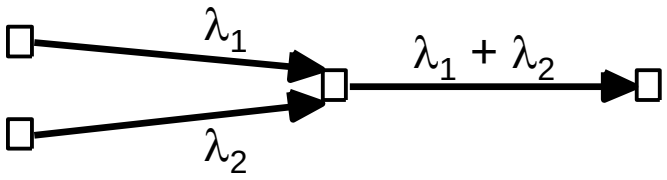
- El valor medio del tiempo entre llegadas es:

$$\mathrm{E}[A] = \frac{1}{\lambda} \quad \mathrm{Var}[A] = \frac{1}{\lambda^2} \tag{1.17}$$

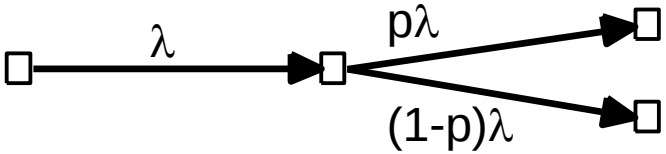
Escuela Politécnica Superior

Procesos de Poisson (III)

- Propiedades:
  - La suma de varios procesos de Poisson independientes es un proceso de Poisson, con tasa de ocurrencias igual a la suma de las de los procesos componentes

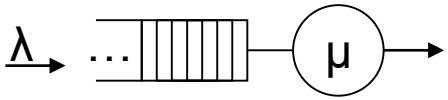


  - Los procesos resultantes de la partición de un proceso de Poisson asignando cada llegada a los procesos resultantes de modo aleatorio (independiente de las asignaciones previas) son procesos de Poisson.
    - La tasa de ocurrencias de cada uno es proporcional a la probabilidad de reparto. Por ejemplo, para una partición en dos procesos con probabilidad p y 1-p:



    - La partición con asignaciones fijas no es un proceso de Poisson.

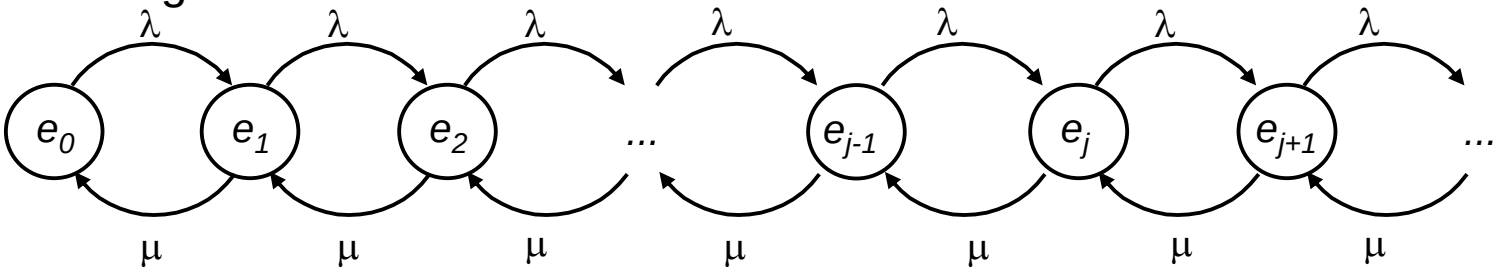
Modelo M/M/1 (I)



- Las llegadas al sistema siguen un proceso de Poisson con tasa de llegadas λ.
- El tiempo de servicio está distribuido de modo exponencial con media 1/μ.
- Hay un único servidor para atender las peticiones.
- En estas condiciones, el número de unidades en el sistema es una cadena de Markov, equivalente al proceso nacimiento – muerte, en la que:

$$\lambda_j = \lambda \quad \mu_j = \mu \quad (\forall j) \qquad \textbf{(1.18)}$$

- El diagrama de transiciones de estados será:



- Sustituyendo (1.18) en (1.11) se obtiene la distribución de probabilidad del número de unidades en el sistema:

$$p_n = p_0(\lambda/\mu)^n \qquad \textbf{(1.19)}$$

Modelo M/M/1 (II)

- p_0 (probabilidad de 0 clientes, sistema inactivo) se puede calcular sabiendo que la suma de todas las probabilidades debe ser 1 (segundo axioma probabilidad).

$$\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} p_0 (\lambda/\mu)^n = 1 \qquad \textbf{(1.20)}$$

Es una serie geométrica, que será convergente si su razón, λ/μ, es menor que 1:

$$\lambda/\mu < 1 \Rightarrow \sum_{n=0}^{\infty} p_0 (\lambda/\mu)^n = \frac{p_0}{1-\lambda/\mu} = 1 \Rightarrow p_0 = 1-\lambda/\mu \qquad \textbf{(1.21)}$$

El factor de utilización del servidor será:

$$\rho = 1 - p_0 = \lambda/\mu \qquad \textbf{(1.22)}$$

Sustituyendo (1.22) en (1.19) se obtiene:

$$p_n = (1-\rho)(\rho)^n \qquad \textbf{(1.23)}$$

Escuela Politécnica Superior

Modelo M/M/1 (III)

- El número medio de unidades en el sistema será el valor medio de (1.23):

$$L = \mathrm{E}[N] = \sum_{n=0}^{\infty} n p_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n = (1-\rho)\sum_{n=0}^{\infty} n\rho^n \qquad \textbf{(1.24)}$$

Considerando que:

$$\sum_{n=0}^{\infty} n\rho^n = \rho\sum_{n=0}^{\infty} n\rho^{n-1} = \rho\frac{\partial}{\partial\rho}\sum_{n=0}^{\infty}\rho^n = \rho\frac{\partial}{\partial\rho}\left(\frac{1}{1-\rho}\right) = \rho\frac{1}{(1-\rho)^2} \qquad \textbf{(1.25)}$$

Sustituyendo (1.25) en (1.24):

$$L = \rho(1-\rho)\frac{1}{(1-\rho)^2} = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda} \qquad \textbf{(1.26)}$$

Escuela Politécnica Superior

Modelo M/M/1 (IV)

- El tiempo medio de estancia en el sistema se calcula aplicando el Teorema de Little (1.3)

$$W = \frac{L}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu-\lambda} \qquad \textbf{(1.27)}$$

- Para el tiempo medio de espera en cola, aplicando (1.2):

$$W_q = W - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)} = \frac{\lambda}{\mu(\mu-\lambda)} \qquad \textbf{(1.28)}$$

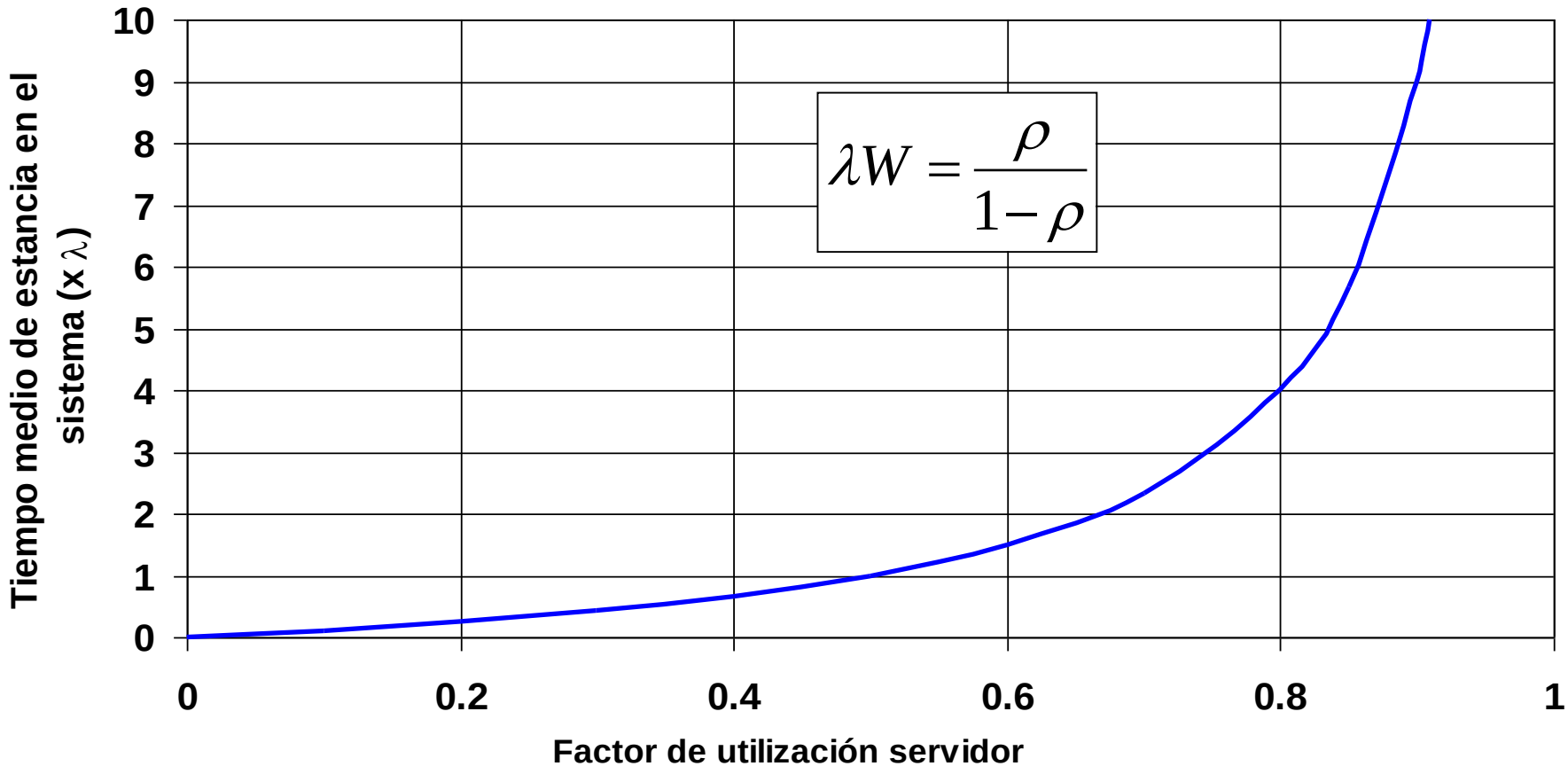
- La ocupación media de la cola, aplicando el Teorema de Little (1.4):

$$L_q = \lambda W_q = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)} \qquad \textbf{(1.29)}$$

Escuela Politécnica Superior

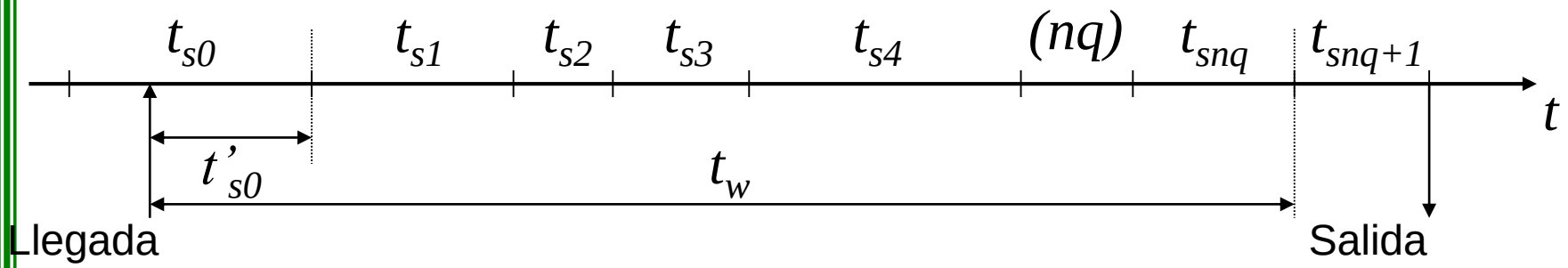
Modelo M/M/1 (V)

Representación gráfica del tiempo medio de estancia en el sistema



Modelo M/M/1 (VI)

- El tiempo de estancia en el sistema para un cliente se puede calcular considerando que es la suma de los siguientes tiempos:
  - El tiempo de servicio residual del cliente que está siendo atendido al producirse la llegada.
  - El tiempo de servicio de los n_q clientes en cola al producirse la llegada.
  - El tiempo de servicio propio.



$$t_{Wn} = t'_{s0} + t_{s1} + t_{s2} + \ldots + t_{snq} + t_{snq+1} \qquad \textbf{(1.30)}$$

  - A partir de esta expresión se puede obtener la función de distribución para el tiempo de estancia en el sistema

$$W(t) = 1 - e^{-(\mu-\lambda)t} \qquad \textbf{(1.31)}$$

Modelo M/M/1 (VII)

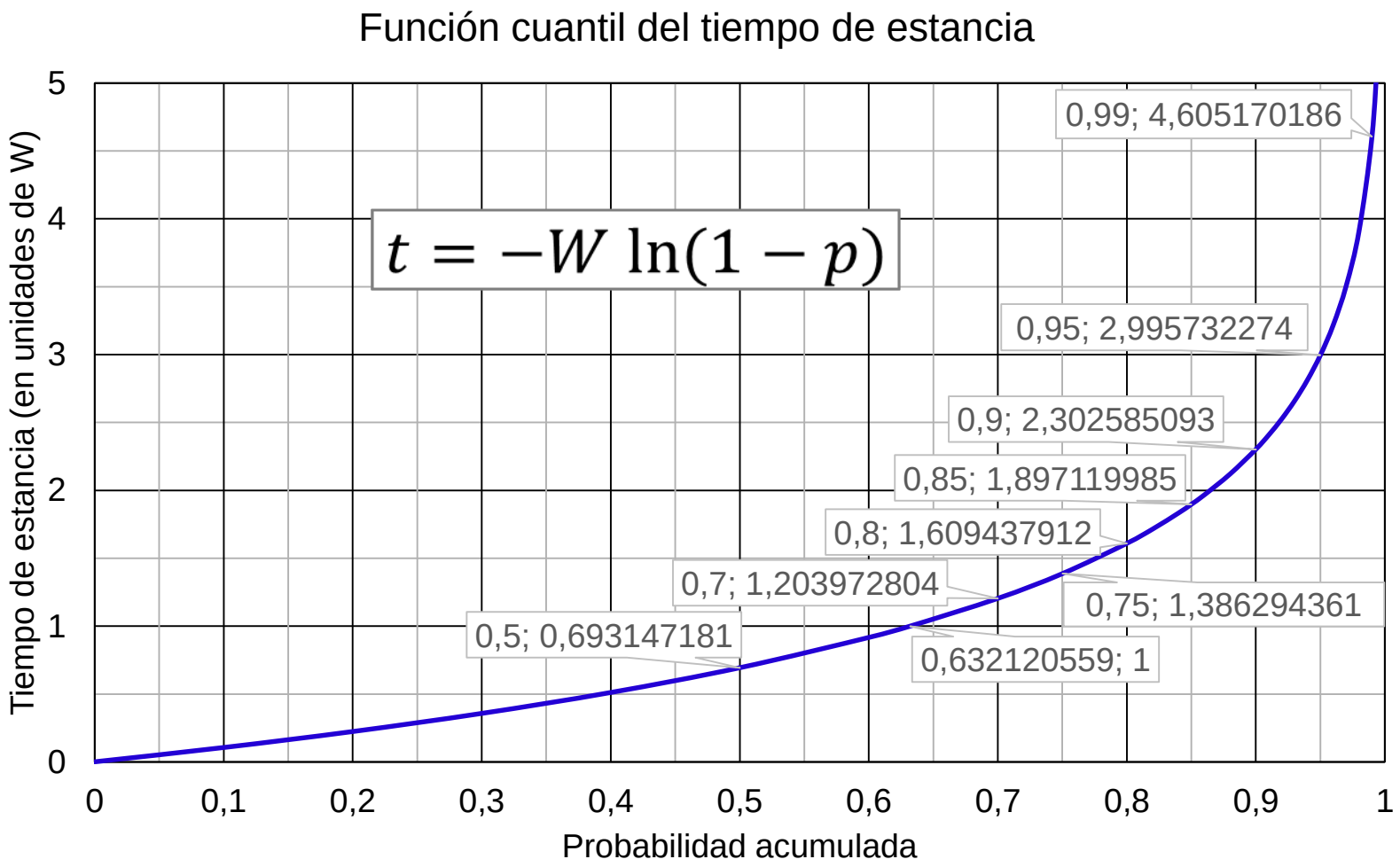
- A partir de esta expresión se pueden calcular los percentiles del tiempo de estancia en el sistema.
- Percentil *p* de una variable aleatoria es el valor de la variable aleatoria para el cual su función de densidad acumulada vale *p*.
- Para un percentil *p*:

$$p = 1 - e^{-(\mu-\lambda)t} \Rightarrow t = \frac{1}{\mu-\lambda}\ln\frac{1}{1-p} = -W\ln(1-p) \qquad \textbf{(1.32)}$$

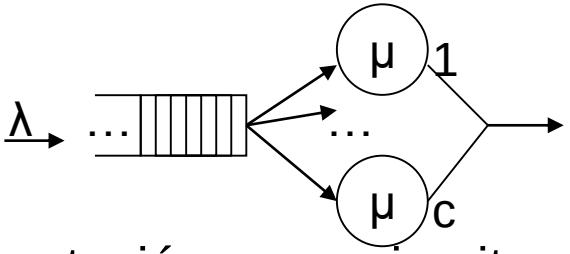
- La siguiente tabla muestra algunos valores de los percentiles:

p	t
.63	W
.70	1.20W
.80	1.61W
.85	1.90W
.90	2.30W
.95	3.00W
.99	4.61W

Modelo M/M/1 (VIII)



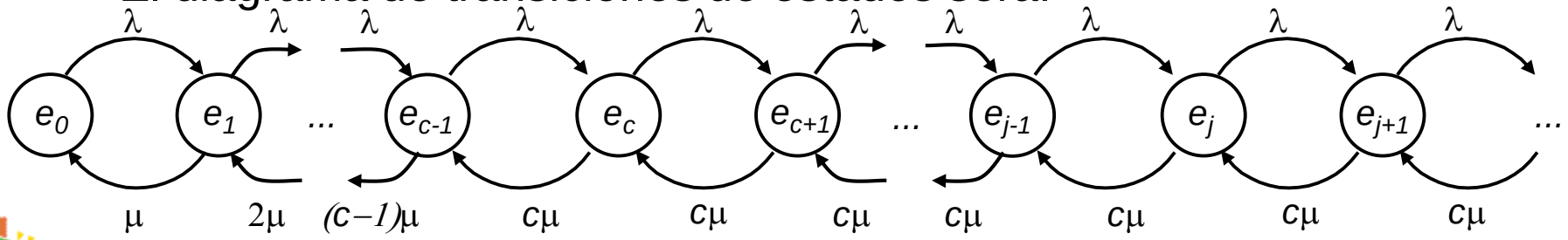
Modelo M/M/c (I)



- Útil para modelar una central de conmutación con c circuitos.
- Las llegadas al sistema siguen un proceso de Poisson con tasa de llegadas λ.
- El tiempo de servicio de cada servidor está distribuido de modo exponencial con media 1/μ.
- Hay c servidores para atender peticiones. El tráfico se reparte por igual entre todos los servidores.
- En estas condiciones, el número de unidades en el sistema es una cadena de Markov, equivalente al proceso nacimiento – muerte, en la que:

$$\lambda_j = \lambda \quad \forall j \quad \mu_j = \begin{cases} j\mu & (j < c) \\ c\mu & (j \geq c) \end{cases} \qquad \textbf{(1.33)}$$

- El diagrama de transiciones de estados será:



Modelo M/M/c (II)

- Sustituyendo (1.33) en (1.11) se obtiene la función de distribución de probabilidad del número de unidades en el sistema:

$$p_n = \begin{cases} p_0 \dfrac{(\lambda/\mu)^n}{n!} & (n < c) \\ p_0 \dfrac{c^c}{c!}\left(\dfrac{\lambda}{c\mu}\right)^n & (n \geq c) \end{cases} \qquad \textbf{(1.34)}$$

- p_0 se calcula aplicando el segundo axioma de la probabilidad:

$$\sum_{n=0}^{\infty} p_n = 1 \Rightarrow p_0 = \left[\left(\sum_{n=0}^{c-1} \frac{(\lambda/\mu)^n}{n!}\right) + \left(\sum_{n=c}^{\infty} \frac{c^c}{c!}\left(\frac{\lambda}{c\mu}\right)^n\right)\right]^{-1} \qquad \textbf{(1.35)}$$

Serie geométrica, que converge si su razón es menor que 1.

Modelo M/M/c (III)

- En este caso, por tanto

$$\frac{\lambda}{c\mu}<1 \Rightarrow p_0=\left[\left(\sum_{n=0}^{c-1}\frac{(\lambda/\mu)^n}{n!}\right)+\left(\frac{(\lambda/\mu)^c}{c!(1-\rho)}\right)\right]^{-1} \quad \textbf{(1.36)}$$

- El factor de utilización de cada servidor será una fracción 1/c del tráfico total:

$$\rho = u/c = \lambda/c\mu \quad \textbf{(1.37)}$$

- La probabilidad de que al llegar un cliente tenga que esperar en cola es:

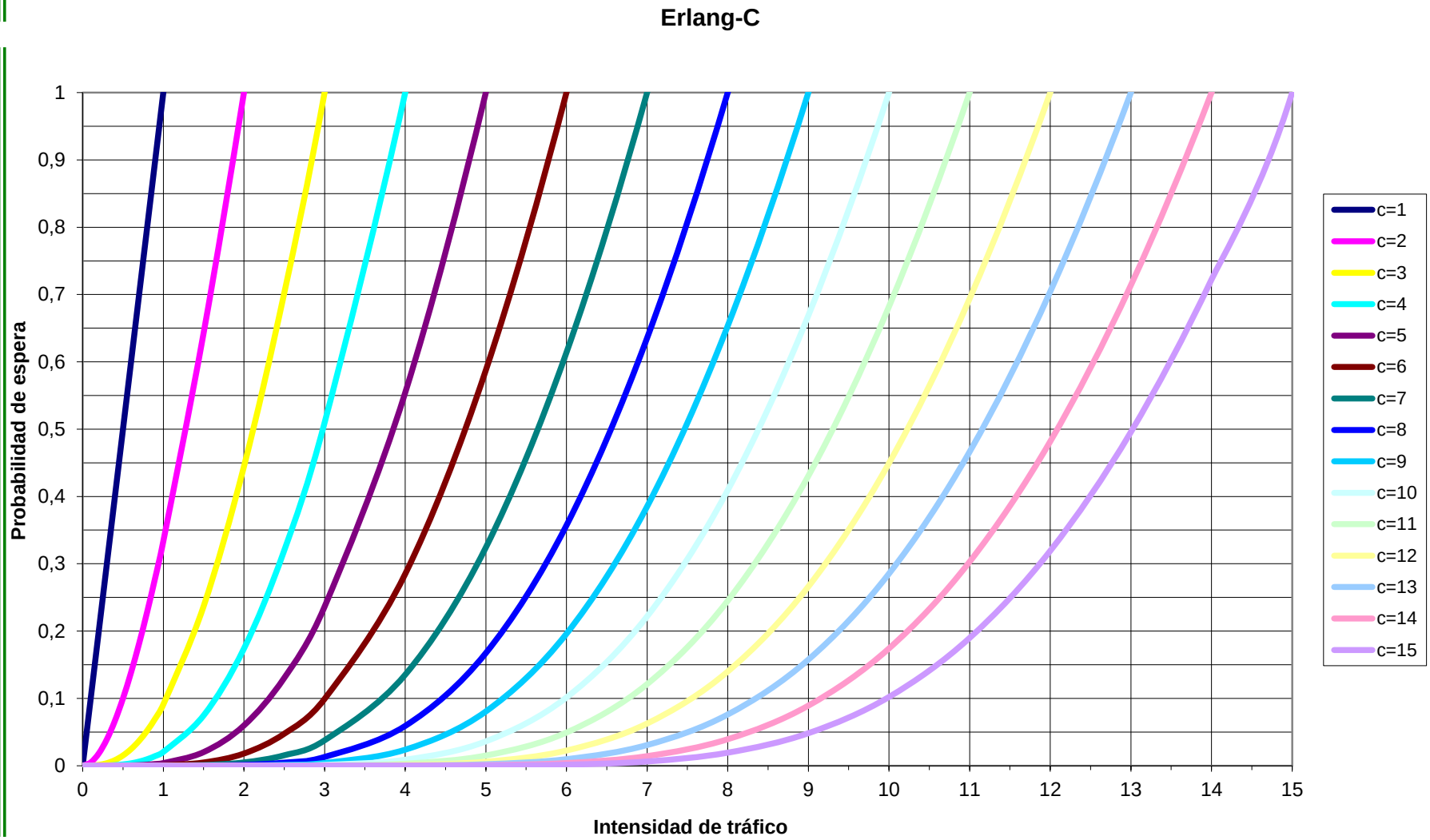
$$P_q = P\{N(t)\geq c\} = \sum_{n=c}^{\infty} p_n = \sum_{n=c}^{\infty} p_0 \frac{c^c}{c!}\left(\frac{\lambda}{c\mu}\right)^n = p_0 \frac{c^c}{c!}\left(\frac{\lambda}{c\mu}\right)^c \sum_{n=c}^{\infty}\left(\frac{\lambda}{c\mu}\right)^{n-c} =$$

$$= p_0 \frac{c^c}{c!}\left(\frac{\lambda}{c\mu}\right)^c \frac{1}{1-\lambda/c\mu} = \frac{p_c}{1-\rho} = E_c(c,u) \quad \textbf{(1.38)}$$

conocida como Fórmula C de Erlang, Erlang-C, o probabilidad de llamada en espera.

Escuela Politécnica Superior

Modelo M/M/c (IV)



Escuela Politécnica Superior

Modelo M/M/c (V)

- El número medio de clientes en cola será el valor medio de (1.34) para $n \geq c$:

$$L_q = \mathrm{E}[N_q] = \sum_{n=c}^{\infty}(n-c)p_n = \sum_{n=c}^{\infty}(n-c)p_0 \frac{c^c}{c!}\left(\frac{\lambda}{c\mu}\right)^n =$$

$$= p_0 \frac{c^c}{c!}\left(\frac{\lambda}{c\mu}\right)^c \sum_{n=c}^{\infty}(n-c)\left(\frac{\lambda}{c\mu}\right)^{n-c} = p_c \sum_{m=0}^{\infty} m\rho^m \qquad \textbf{(1.39)}$$

Sustituyendo en (1.39) los valores obtenidos en (1.25) y (1.38), sucesivamente, se obtiene:

$$L_q = p_c \frac{\rho}{(1-\rho)^2} = P_q \frac{\rho}{1-\rho} \qquad \textbf{(1.40)}$$

Modelo M/M/c (VI)

- El tiempo medio de espera en cola, aplicando la forma (1.4) del Teorema de Little a (1.40), y considerando (1.37), viene dado por la relación:

$$W_q = \frac{L_q}{\lambda} = P_q \frac{\rho}{\lambda(1-\rho)} = \frac{P_q}{c\mu - \lambda} \qquad \textbf{(1. 41)}$$

- Y para el tiempo medio de estancia en el sistema, por (1.2):

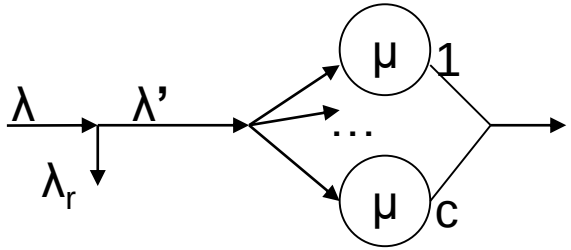
$$W = W_q + T_s = \frac{P_q}{c\mu - \lambda} + \frac{1}{\mu} \qquad \textbf{(1.42)}$$

- El número medio de clientes en el sistema se obtiene aplicando la forma (1.3) del Teorema de Little a (1.42):

$$L = \lambda W = \frac{\lambda P_q}{c\mu - \lambda} + \frac{\lambda}{\mu} = \frac{P_q \rho}{1-\rho} + c\rho = L_q + c\rho \qquad \textbf{(1.43)}$$

Escuela Politécnica Superior

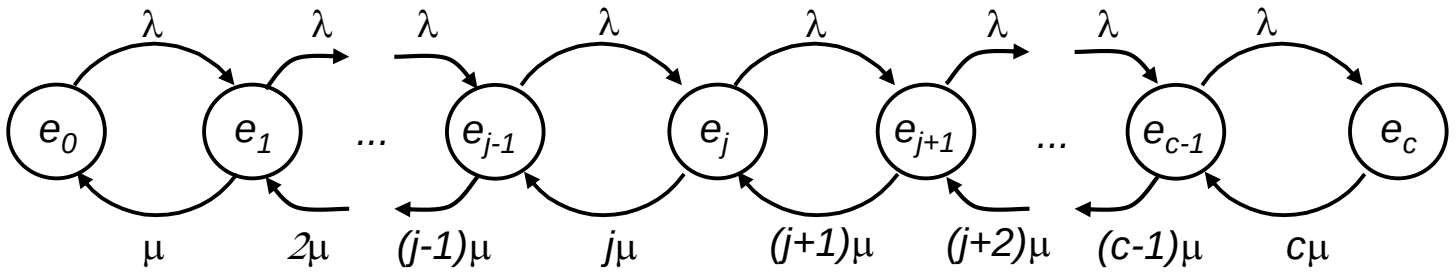
Modelo M/M/c/c (I)



- Las condiciones son iguales que en el caso M/M/c, pero no hay colas de espera: Cualquier cliente adicional (> *c*) se rechaza.
- Útil para modelar una central de conmutación con c circuitos.
- En estas condiciones, el número de unidades en el sistema es una cadena de Markov, equivalente al proceso nacimiento – muerte, en la que:

$$\lambda_j = \begin{cases} \lambda & (j < c) \\ 0 & (j \geq c) \end{cases} \qquad \mu_j = \begin{cases} j\mu & (j \leq c) \\ 0 & (j > c) \end{cases} \tag{1.44}$$

- El diagrama de transiciones de estados será:



Modelo M/M/c/c (II)

- Sustituyendo (1.44) en (1.11) se obtiene la función de distribución de probabilidad del número de unidades en el sistema:

$$p_n = p_0 \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} \quad (0 \le n \le c) \qquad \textbf{(1.45)}$$

- p_0 se calcula aplicando el segundo axioma de la probabilidad, y se obtiene:

$$p_0 = \left[\sum_{n=0}^{c} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}\right]^{-1} \qquad \textbf{(1.46)}$$

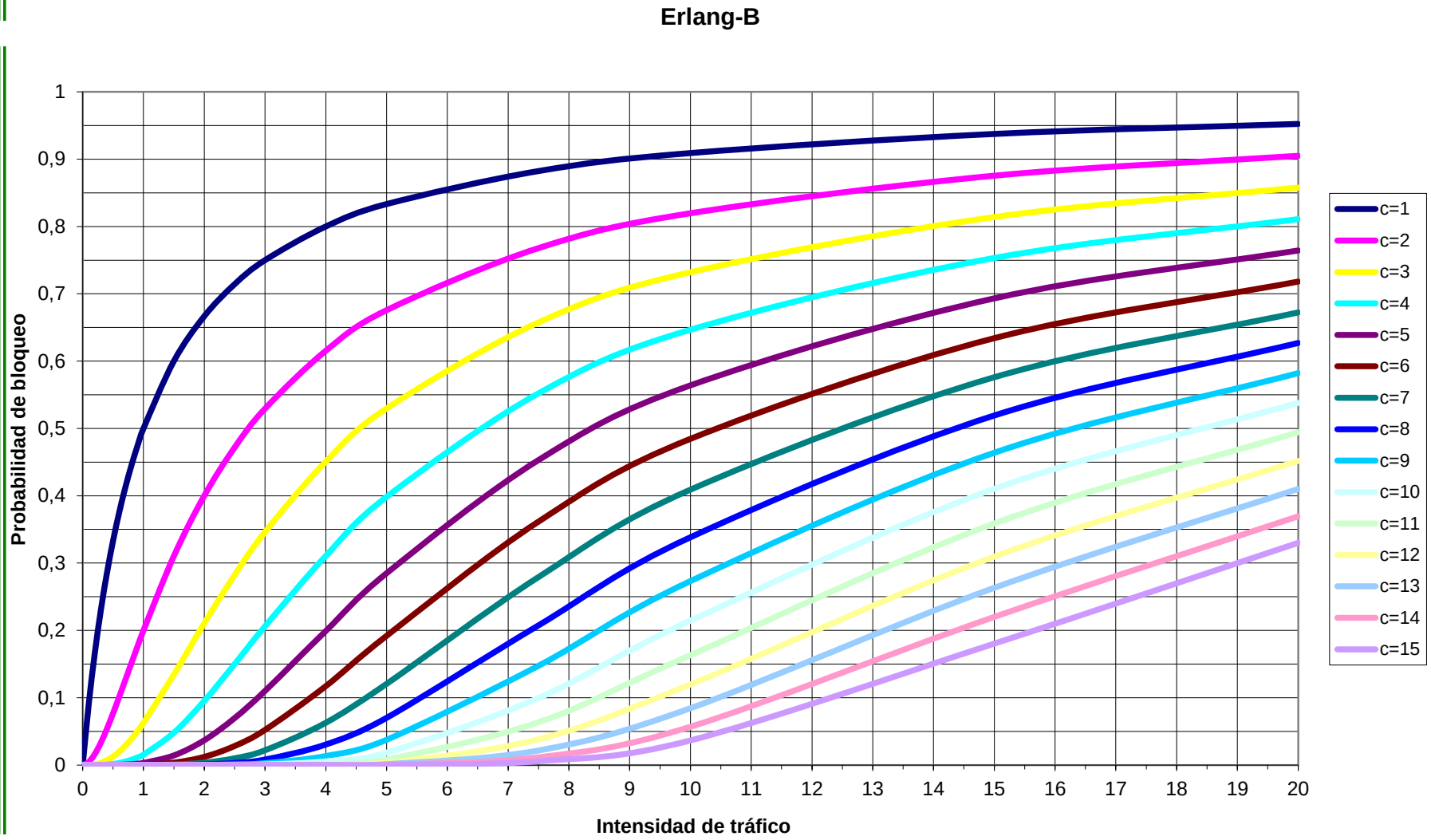
- La probabilidad de perder una petición será la probabilidad de que haya *c* unidades en el sistema:

$$p_c = \frac{(\lambda/\mu)^c / c!}{\sum_{i=0}^{c} \left[(\lambda/\mu)^i / i!\right]} = E_B(c,u) \qquad \textbf{(1.47)}$$

conocida como fórmula B de Erlang, Erlang-B, o probabilidad de bloqueo de Erlang.

Escuela Politécnica Superior

Modelo M/M/c/c (III)



Escuela Politécnica Superior

Modelo M/M/c/c (III)

- La tasa de llegadas efectiva al sistema será:

$$\lambda' = \lambda(1 - p_c) \qquad \textbf{(1.48)}$$

- El factor de utilización de cada servidor viene dado por:

$$\rho = \frac{\lambda'}{c\mu} = \frac{\lambda}{c\mu}(1 - p_c) \qquad \textbf{(1.49)}$$

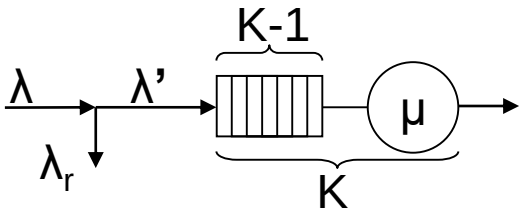
- En este modelo, al no haber cola de espera, W_q=0 y L_q=0, con lo que por (1.2) y (1.3):

$$W = W_q + \frac{1}{\mu} = \frac{1}{\mu} \qquad \textbf{(1.50)}$$

$$L = \lambda' W = \frac{\lambda'}{\mu} = \frac{\lambda}{\mu}(1 - p_c) = c\rho \qquad \textbf{(1.51)}$$

Escuela Politécnica Superior

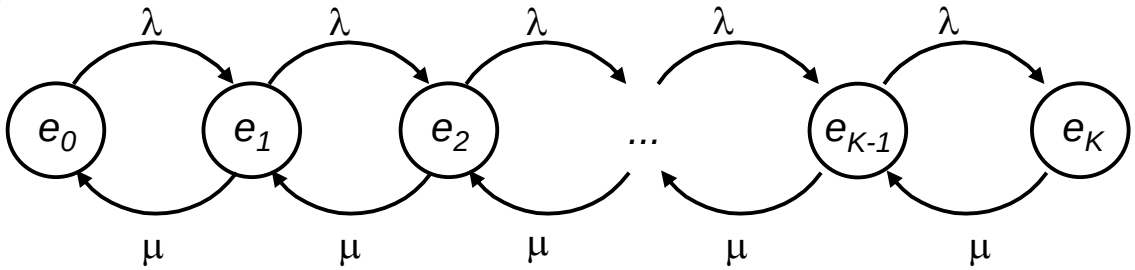
Modelo M/M/1/K (I)



- Las condiciones son iguales que en el caso M/M/1, pero el número de unidades en el sistema está limitado a *K* (cola finita).
- Útil para modelar un *router*
- En estas condiciones, el número de unidades en el sistema es una cadena de Markov, equivalente al proceso nacimiento – muerte, en la que:

$$\lambda_j = \begin{cases} \lambda & (j < K) \\ 0 & (j \geq K) \end{cases} \quad \mu_j = \begin{cases} \mu & (j \leq K) \\ 0 & (j > K) \end{cases} \qquad \textbf{(1.52)}$$

- El diagrama de transiciones de estados será:



- Sustituyendo (1.52) en (1.11) se obtiene la función de distribución de probabilidad del número de unidades en el sistema:

$$p_n = p_0 \left(\frac{\lambda}{\mu} \right)^n \quad (0 \leq n \leq K) \qquad \textbf{(1.53)}$$

Modelo M/M/1/K (II)

- p_0 se calcula a partir del segundo axioma de la probabilidad, y se obtiene:

$$p_0 = \left[\sum_{n=0}^{K} \left(\frac{\lambda}{\mu} \right)^n \right]^{-1} = \begin{cases} \dfrac{1-\lambda/\mu}{1-(\lambda/\mu)^{K+1}} & (\lambda \neq \mu) \\ \dfrac{1}{K+1} & (\lambda = \mu) \end{cases} \quad \textbf{(1.54)}$$

- La tasa de llegadas efectiva, λ' viene dada por la expresión:

$$\lambda' = \lambda(1 - p_K) = \begin{cases} \lambda \dfrac{1-(\lambda/\mu)^K}{1-(\lambda/\mu)^{K+1}} & (\lambda \neq \mu) \\ \lambda \dfrac{K}{K+1} & (\lambda = \mu) \end{cases} \quad \textbf{(1.55)}$$

Escuela Politécnica Superior

Modelo M/M/1/K (III)

- Por haber un único servidor, la probabilidad de que el sistema esté activo dará el factor de utilización del servidor, que será:

$$\rho = 1 - p_0 = \begin{cases} \frac{\lambda}{\mu}\left[\frac{1-(\lambda/\mu)^K}{1-(\lambda/\mu)^{K+1}}\right] & (\lambda \neq \mu) \\ \frac{K}{K+1} & (\lambda = \mu) \end{cases} \quad \textbf{(1.56)}$$

- Y en cualquiera de los dos casos se pude verificar que es igual a la intensidad de tráfico efectiva de entrada al servidor:

$$\rho = \frac{\lambda'}{\mu} = \frac{\lambda}{\mu}(1 - p_K) \quad \textbf{(1.57)}$$

- El número medio de unidades en el sistema será el valor esperado de (1.53).
  - Para $\lambda \neq \mu$:

$$L = \mathrm{E}[N] = \sum_{n=0}^{K} n p_n = \sum_{n=0}^{K} n p_0 \left(\frac{\lambda}{\mu}\right)^n = p_0 \sum_{n=0}^{K} n \left(\frac{\lambda}{\mu}\right)^n \quad \textbf{(1.58)}$$

Escuela Politécnica Superior

Modelo M/M/1/K (IV)

- Análogamente a (1.25), el sumatorio se puede calcular del siguiente modo.
  - Haciendo $u=\lambda/\mu$:

$$\sum_{n=0}^{K} nu^n = u\sum_{n=0}^{K} nu^{n-1} = u\frac{\partial}{\partial u}\sum_{n=0}^{K} u^n = u\frac{\partial}{\partial u}\left(\frac{1-u^{K+1}}{1-u}\right) =$$

$$= \frac{-(1-u)(K+1)u^K + (1-u^{K+1})}{(1-u)^2} = u\frac{1-(K+1)u^k + Ku^{K+1}}{(1-u)^2} \qquad \textbf{(1.59)}$$

  - Deshaciendo el cambio de variable y sustituyendo se obtiene:

$$L = \frac{\lambda/\mu}{1-\lambda/\mu}\left[\frac{1-(K+1)(\lambda/\mu)^k + K(\lambda/\mu)^{K+1}}{1-(\lambda/\mu)^{K+1}}\right] \quad (\lambda \neq \mu) \qquad \textbf{(1.60)}$$

Modelo M/M/1/K (V)

- Y en el caso en que $\lambda=\mu$:

$$L=\sum_{n=0}^{K} n\frac{1}{K+1}=\frac{1}{K+1}\sum_{n=0}^{K} n=\frac{1}{K+1}\frac{K+1}{2}K=\frac{K}{2}\quad(\lambda=\mu) \tag{1.61}$$

- Y para el resto de los valores medios, aplicando (1.2), (1.3) y (1.4), respectivamente:

$$W=\frac{L}{\lambda'}=\frac{L}{\lambda(1-p_K)} \tag{1.62}$$

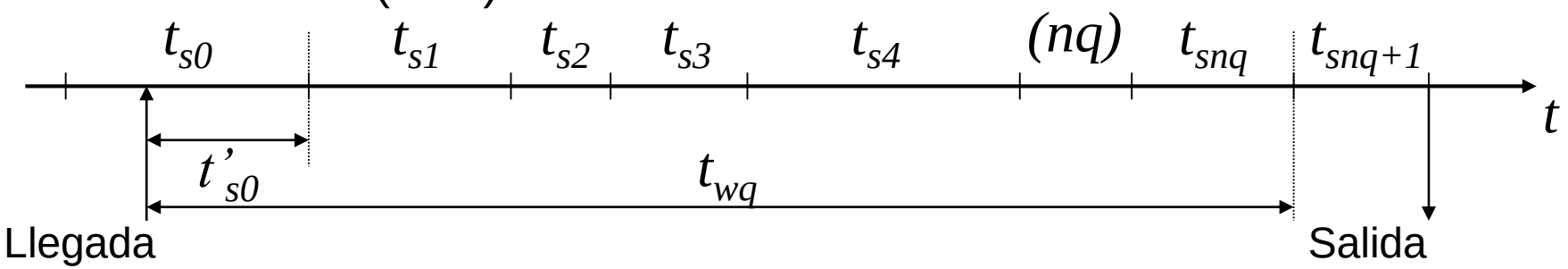
$$W_q=W-\frac{1}{\mu} \tag{1.63}$$

$$L_q=L-\frac{\lambda'}{\mu}=L-\frac{\lambda}{\mu}(1-p_K)=L-\rho \tag{1.64}$$

Escuela Politécnica Superior

Modelo M/G/1 (I)

- El tiempo de servicio tiene una distribución aleatoria cualquiera *S*, de la que se conocen sus momentos E[*S*] y E[S^2].
- Para calcular el tiempo medio de estancia en el sistema, recordando (1.30)



$$t_{Wn} = t'_{s0} + t_{s1} + t_{s2} + \ldots + t_{snq} + t_{snq+1}$$

- Para calcular su valor esperado, considerando que todos los tiempos son independientes:

$$W_q = \mathrm{E}\left[t'_{s0} + t_{s1} + t_{s2} + \ldots + t_{snq}\right] = \mathrm{E}\left[t'_{s0}\right] + \mathrm{E}\left[N_q\right]\mathrm{E}\left[S\right] \qquad \textbf{(1.65)}$$

- Por el teorema de Little (1.4):

$$W_q = \mathrm{E}\left[t'_{s0}\right] + \lambda W_q \,\mathrm{E}\left[S\right] \qquad \textbf{(1.66)}$$

Modelo M/G/1 (II)

- El valor esperado del tiempo residual se puede demostrar que viene dado por la expresión:

$$\mathrm{E}[t'_{s0}] = \frac{\lambda \mathrm{E}[S^2]}{2} \qquad \textbf{(1.67)}$$

- Sustituyendo en (1.66), y recordando que ρ=λ/μ= λE[S]:

$$W_q = \frac{\lambda \mathrm{E}[S^2]}{2} + W_q \rho \Rightarrow W_q = \frac{\lambda \mathrm{E}[S^2]}{2(1-\rho)} \qquad \textbf{(1.68)}$$

- El resto de los valores medios se obtienen a partir del teorema de Little y (1.2):

$$W = W_q + T_s = \frac{\lambda \mathrm{E}[S^2]}{2(1-\rho)} + \mathrm{E}[S] \qquad \textbf{(1.69)}$$

$$L_q = \lambda W_q = \frac{\lambda^2 \mathrm{E}[S^2]}{2(1-\rho)} \qquad \textbf{(1.70)}$$

$$L = \lambda W = \frac{\lambda^2 \mathrm{E}[S^2]}{2(1-\rho)} + \rho \qquad \textbf{(1.71)}$$

Escuela Politécnica Superior

Modelo M/G/1 (III)

- Verificando los resultados para una distribución exponencial, (M/M/1):

$$E[S]=\frac{1}{\mu} \quad E[S^2]=\frac{2}{\mu^2} \quad W_q=\frac{\lambda}{2(1-\rho)}\frac{2}{\mu^2}=\frac{\rho}{\mu(1-\rho)} \qquad \textbf{(1.72)}$$

  que es el valor obtenido en (1.18).

- Para una distribución constante, M/D/1:

$$E[S]=\frac{1}{\mu} \quad E[S^2]=\frac{1}{\mu^2} \quad W_q=\frac{\lambda}{2(1-\rho)}\frac{1}{\mu^2}=\frac{\rho}{2\mu(1-\rho)} \qquad \textbf{(1.73)}$$

  Son los valores mínimos para cualquier cola M/G/1 con los mismos valores de λ y μ. El valor obtenido para W es la mitad que el caso M/M/1.

- En caso de no conocer de modo analítico la función de distribución de probabilidad de *S*, se pueden emplear estimadores de E[*S*] y E[S^2] a partir de muestras del tiempo de servicio:

$$\{t_{s1}, t_{s2}, ..t_{sn}\} \qquad \overline{S}=\frac{\sum_{i=1}^{n} t_i}{n} \qquad \textbf{(1.74)} \qquad \overline{S^2}=\frac{\sum_{i=1}^{n} t_i^{\,2}}{n-1}=\sigma_s^{\,2}+\overline{S}^2 \qquad \textbf{(1.75)}$$

Aproximación a distribuciones exponenciales

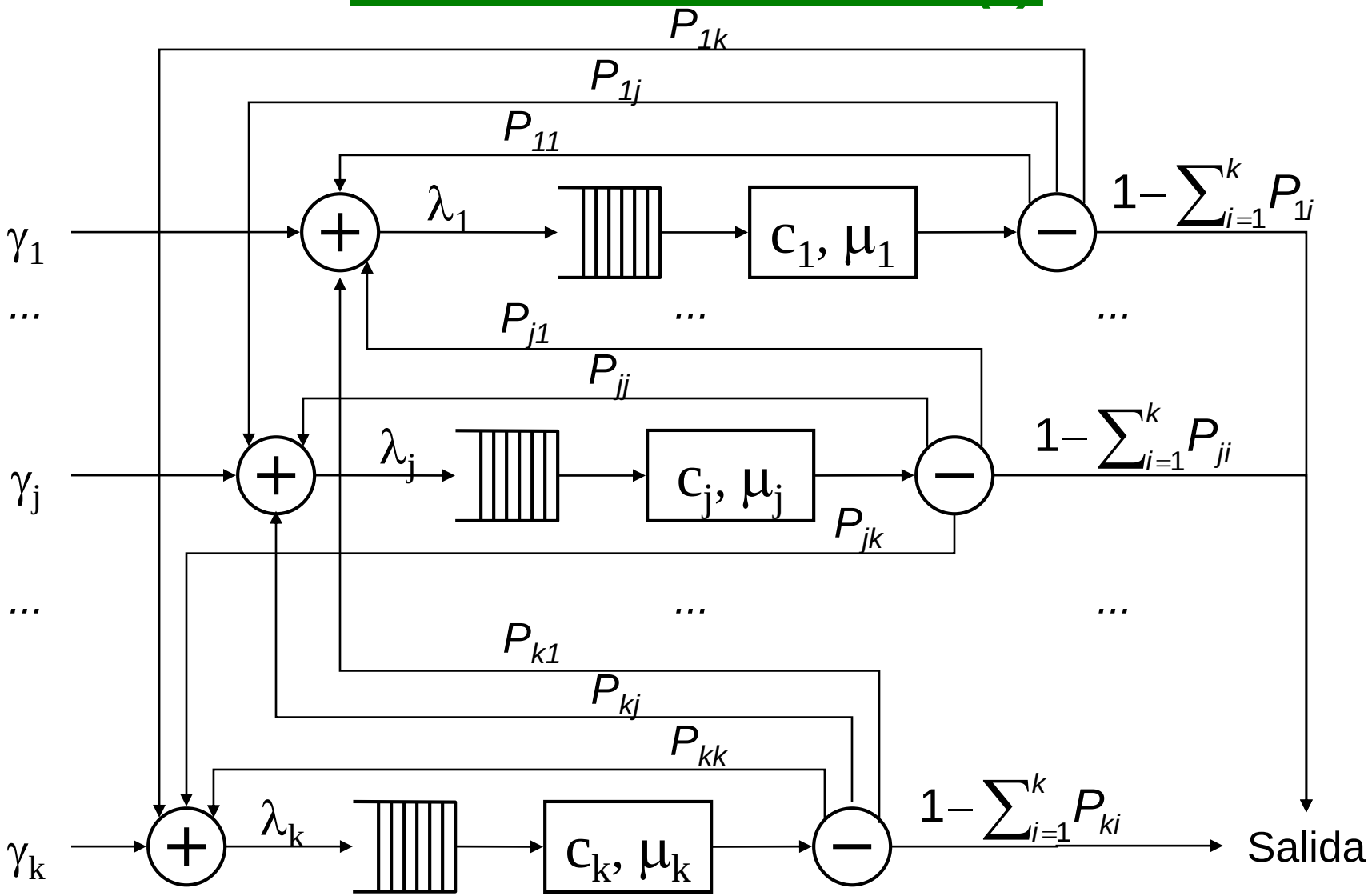
- Se define el Coeficiente Cuadrático de Variación de una variable aleatoria X como:

$$C^2 = \frac{\mathrm{Var}[S]}{\mathrm{E}[S]^2} = \frac{\mathrm{E}[S^2] - \mathrm{E}[S]^2}{\mathrm{E}[S]^2} = \frac{\mathrm{E}[S^2]}{\mathrm{E}[S]^2} - 1 \qquad \textbf{(1.76)}$$

  es decir, el cociente entre la varianza y el cuadrado del valor medio.
- Si X tiene distribución exponencial, $C^2 = 1$.
- Si se desconoce la distribución del tiempo de servicio se pueden estimar la media y la varianza, calcular C^2 y analizar según sus valores el tipo de proceso:
  - $0 < C^2 < 0{,}7$: Tendencia uniforme.
    - Se puede modelar mediante una distribución de tipo Erlang-m ($C^2 = 1/m$).
  - $0{,}7 < C^2 < 1{,}3$: Comportamiento Poissoniano.
  - $1{,}3 < C^2$: Tendencia al agrupamiento:
    - Se puede modelar mediante una distribución de tipo hiperexponencial.

Escuela Politécnica Superior

Redes de colas (I)



Redes de colas (II)

- Representan el flujo de peticiones de clientes a través de varios centros de servicio.
- **Teorema de Burke**: La salida en estado estacionario de un sistema M/M/c, con parámetro de entrada λ, es también un proceso de Poisson de parámetro λ.
- Redes de colas abiertas: Conjunto de K colas a las cuales:
  - Los clientes llegan del exterior en procesos de Poisson independientes con tasa γ_j.
  - Cada cola j tiene c_j servidores. El tiempo de servicio de cada servidor está distribuido exponencialmente con media μ_j.
  - Tras el proceso en el sistema *j*, el cliente pasa a la cola *i* con una probabilidad p_{ji}, o sale del sistema con probabilidad $1-\sum_{i=1}^{K} p_{ij}$
- La tasa total de llegadas a cada servidor es:

$$\lambda_j = \gamma_j + \sum_{i=1}^{K} \lambda_i p_{ij} \qquad \textbf{(1.77)}$$

- Tiene solución única si no hay clientes que permanecen indefinidamente en el sistema. El vector del número total de clientes en el sistema:

$$N = \left(N_1, N_2, \ldots, N_K\right) \qquad \textbf{(1.78)}$$

es un proceso de Markov.

Redes de colas (III)

- **Teorema de Jackson**: En cualquier red de colas abiertas como la descrita, si se verifica que

$$\lambda_j < c_j \mu_j \quad \forall j \tag{1.79}$$

entonces para cualquier posible estado

$$n = (n_1, n_2, ..., n_K) \tag{1.80}$$

la función de distribución de probabilidad del sistema viene dada por:

$$P[N = n] = P[N_1 = n_1]P[N_2 = n_2]...P[N_K = n_K] \tag{1.81}$$

donde $P[N_j = n_j]$ es la distribución del número de unidades en el sistema de un sistema M/M/c_j con tasa de llegadas λ_j y tasa de servicio μ_j

- Por tanto, se puede suponer que las colas son independientes y calcularlas por separado.
- **Corolario**: El tiempo medio de estancia en la red de colas se puede calcular como:

$$W_{red} = \frac{\sum L_j}{\sum \gamma_j} \tag{1.82}$$

Donde L_j es el número medio de unidades en cada cola, calculado según la suposición anterior, y γ_j son las distintas tasas de llegadas a la red.

Escuela Politécnica Superior