

Modelado de usuarios

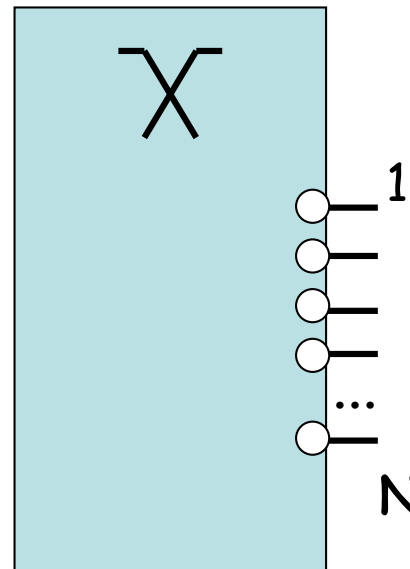
Area de Ingeniería Telemática
<http://www.tlm.unavarra.es>

Arquitectura de Redes, Sistemas y Servicios

Carga o tráfico

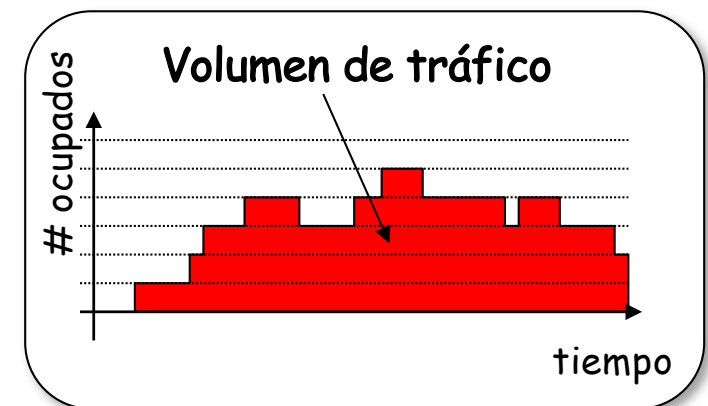
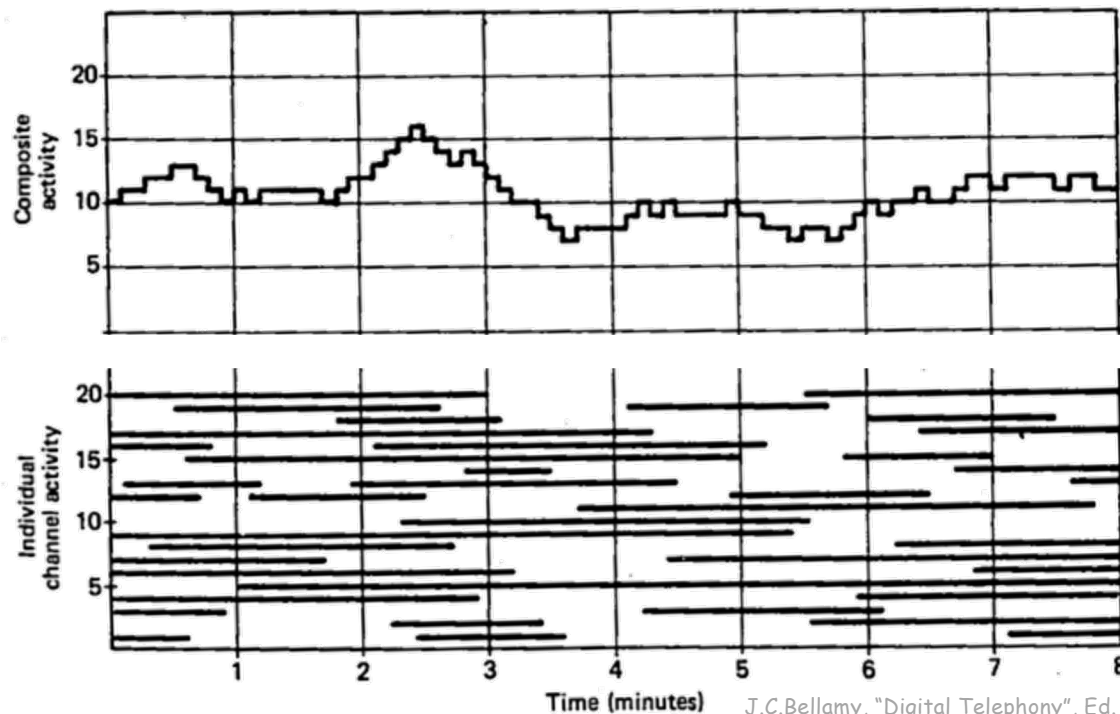
Capacidad

- Medida de la habilidad del sistema para proporcionar servicio
- Típicamente se mide como el número de "servidores" (líneas de salida, puertos de un conmutador...)
- Variable de diseño del sistema
- Proporcional al coste
 - Más capacidad = más coste y más calidad de servicio



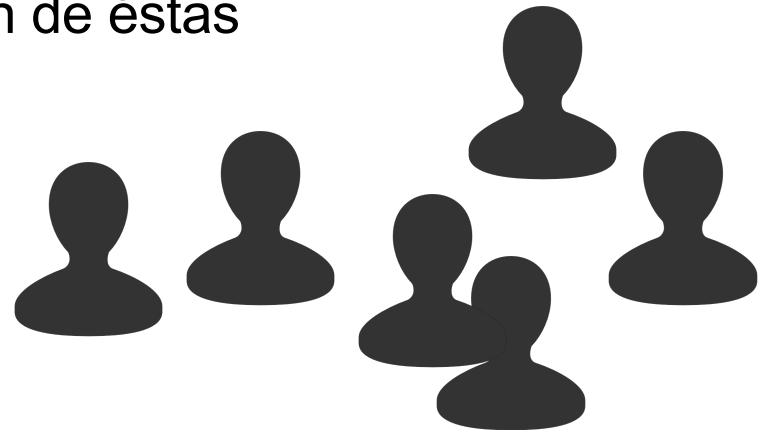
Carga o Tráfico

- Medida de la demanda de servicio al sistema
- Agregación de todas las peticiones de servicio de los usuarios
- = recursos en uso del sistema bajo condiciones de servicio ideales
- **Variables aleatorias**
 - Peticiones de servicio llegan de forma aleatoria
 - Solicitan servicio durante una cantidad de tiempo no predecible
- Volumen de tráfico: suma de las duraciones de los servicios



Carga o Tráfico

- Depende de
 - Número de usuarios (n)
 - Tasa a la que generan llamadas (λ_i)
 - Duración de las llamadas (s)
- No distingue el efecto de n del efecto de λ_i
 - Ej: 600 usuarios, cada uno con una petición por hora, es equivalente a 10 usuarios con una petición por minuto cada uno
- Normalmente lo reducimos a:
 - Tasa de generación de llamadas de todos los usuarios (λ)
 - Duración de las llamadas (s)
- El primer paso del análisis de tráfico es la caracterización de las llegadas de peticiones y la duración de éstas



Recordatorio: Medida del Tráfico

- Intensidad de tráfico

$$I = \frac{\text{Volumen de tráfico}}{\text{Tiempo de observación}} = \frac{\text{Tiempo acumulado de ocupación}}{\text{Tiempo de observación}}$$



- Sin unidades físicas. Se mide en *Erlangs* (*E*) (*Agner Krarup Erlang 1878-1929*)
- **1 Erlang** = el tráfico que mantiene ocupada completamente una línea durante el tiempo de observación
- Intensidad de tráfico media: empleando el volumen *medio* de tráfico en el intervalo de observación
- La máxima cantidad de tráfico que pueden cursar *N* líneas es de *N Erlangs*

Pregunta

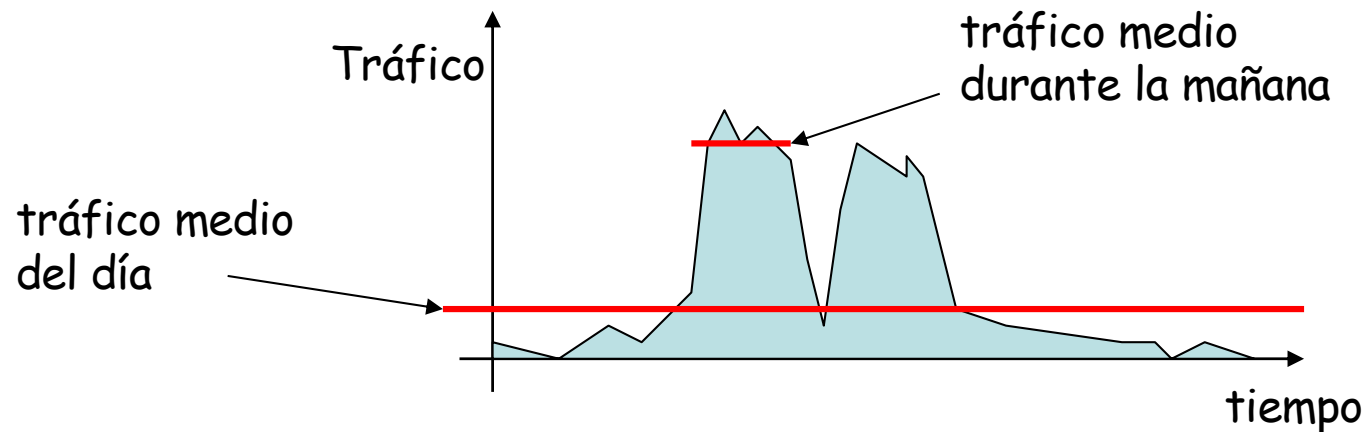
- ¿Cuál es la intensidad de tráfico que generan 200 usuarios si entre todos, a lo largo del día, hacen 80 llamadas, en media de 2 minutos de duración?

Hora cargada

Tráfico ofrecido y cursado

Medida del Tráfico

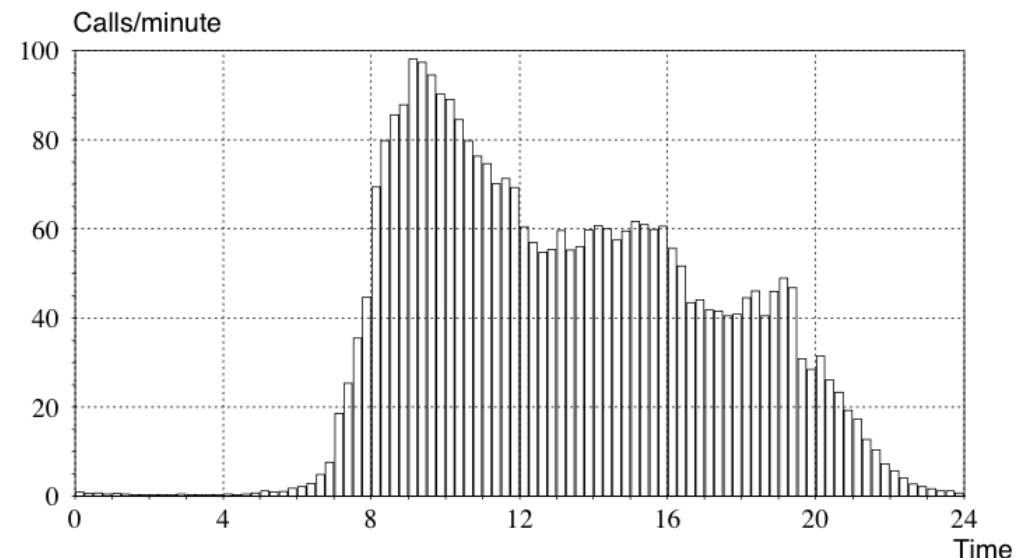
- Normalmente la intensidad del tráfico varía con el tiempo (no es un proceso estocástico estacionario) pero se puede considerar estable en un tiempo limitado



- En telefonía se caracteriza por horas
- Varía entre meses, entre días y entre horas del mismo día (y dentro de la hora)
- Suele haber patrones semanales
- Días de fiesta, el clima, etc. afectan al patrón

Hora cargada (“busy hour”)

- Periodo de 60 minutos consecutivos durante los cuales el volumen de tráfico es máximo
- Los análisis para dimensionamiento de equipos se efectúan siempre sobre la **hora cargada**
- Para determinarla se toman medidas en **intervalos de 15min** y entonces es el periodo de tiempo de 4 intervalos consecutivos con mayor volumen de tráfico
- Se calcula la hora cargada en un periodo largo (unas semanas) en la época del año de mayor tráfico
- Diferentes patrones usuarios residenciales y empresariales
- No es el volumen de tráfico mayor del año (nochevieja, día de la madre,...) pues llevaría a un sobredimensionamiento para la mayor parte del tiempo
- 1 teléfono en hora cargada approx. 0.05-0.1 E y 3-4min duración



Calidad de servicio

- Medida de la bondad del servicio proporcionado
- En telefonía:
 - Probabilidad de bloqueo = probabilidad de que el sistema no pueda aceptar una llamada entrante.

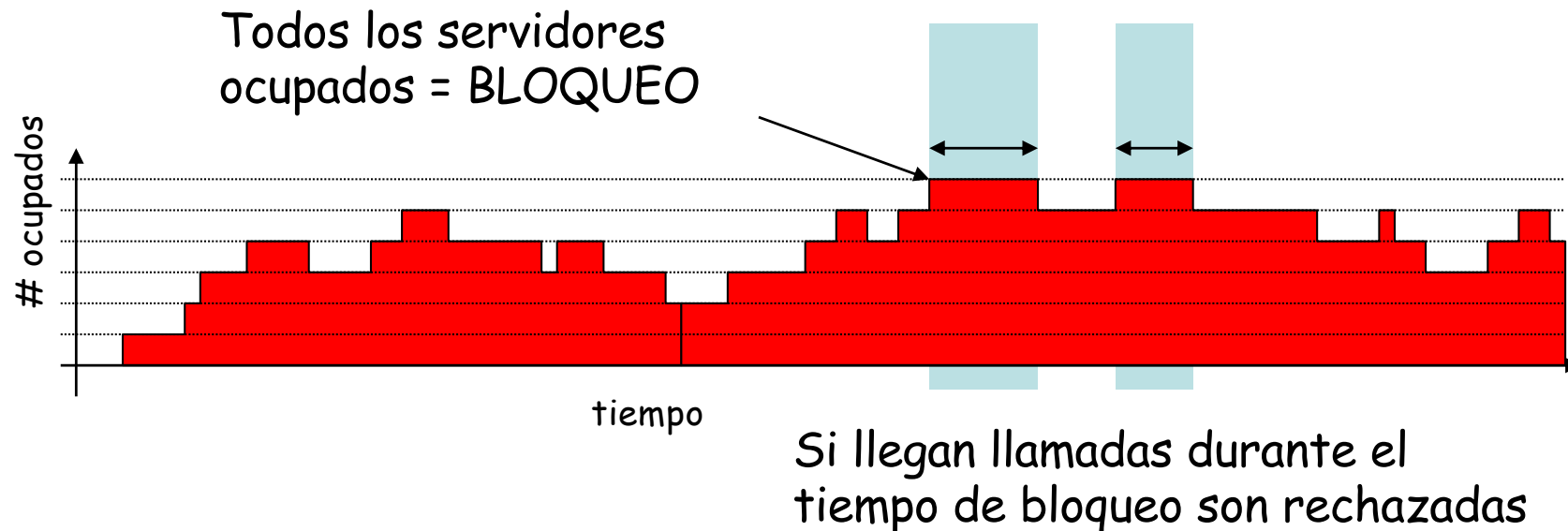
En ese caso:

- Se descarta: La llamada es rechazada → Menos calidad de servicio (*congestion theory*)
 - Se hace esperar la llamada hasta que se libere una línea → Menos calidad de servicio (*queueing theory*)
- Analizaremos el caso con descarte
- Requisito de diseño del sistema: probabilidad de bloqueo objetivo
- Dimensionar la capacidad para conseguirla



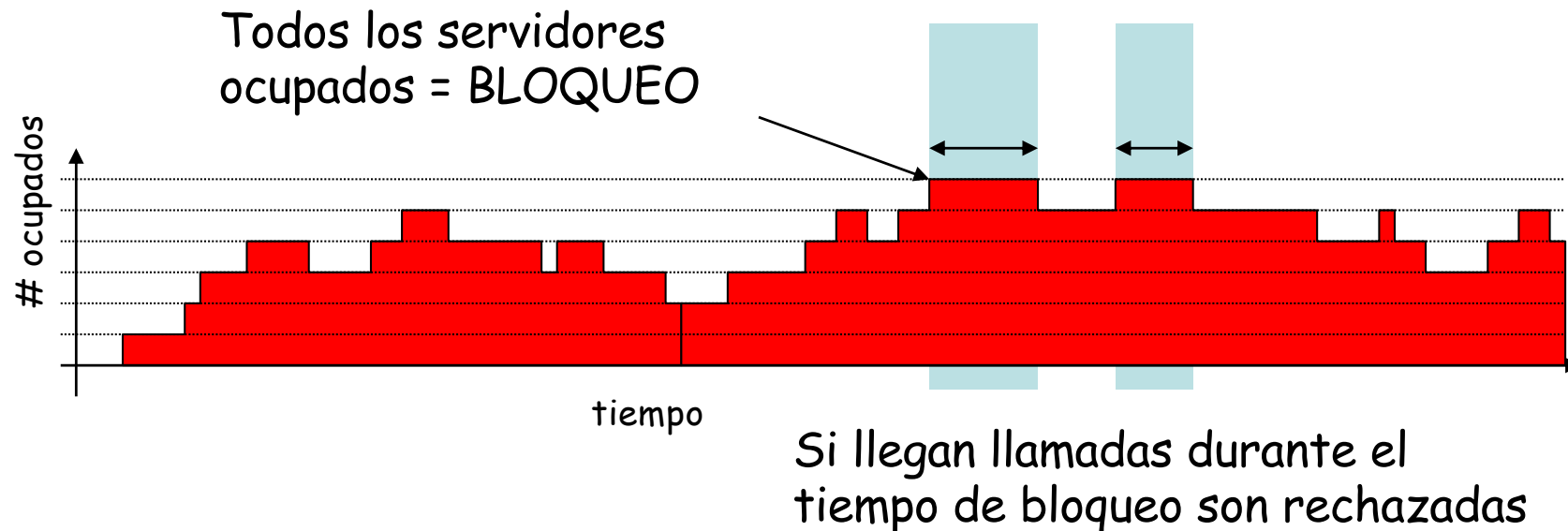
Tráfico ofrecido vs cursado

- Tráfico ofrecido: el tráfico total que sería cursado por una red que pudiera dar servicio a todas las peticiones
- Normalmente por economía dispondremos de recursos finitos
- Cuando el número de llamadas en curso en la troncal = número de líneas, el sistema está en BLOQUEO
- (líneas = “servidores” en terminología de teoría de colas)
- Buscaremos calcular la probabilidad de que una nueva llamada (una “llegada”) encuentre al sistema en bloqueo



Tráfico ofrecido vs cursado

- El tráfico cursado tiene en cuenta solo las llamadas que se han podido establecer
- Es decir, las “peticiones de servicio” que se han podido “atender”
- Es siempre menor o igual al tráfico ofrecido



Tráfico cursado

- Un conjunto k de líneas tiene un tráfico ofrecido de I Erlangs y una probabilidad de bloqueo P_b
- ¿Cuál es la intensidad de tráfico que atraviesa las líneas?
- Esto será el **tráfico cursado** y será a su vez el tráfico ofrecido al siguiente sistema al que lleguen las líneas

$$I_c = I_{in} (1 - P_b)$$

I_c : tráfico cursado

I_{in} : tráfico ofrecido o de entrada

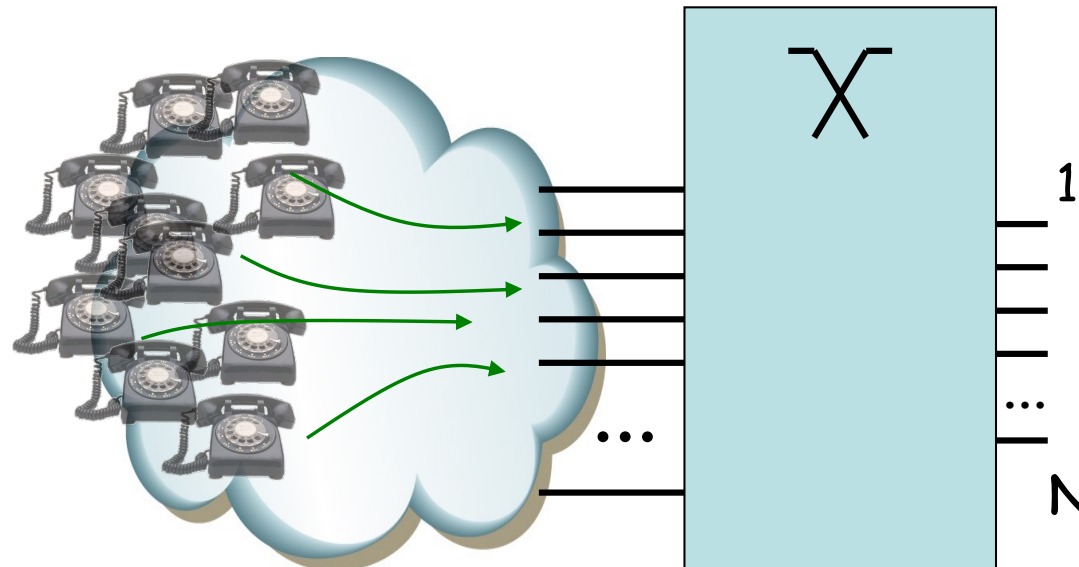
P_b : Probabilidad de bloqueo



Tipos de problemas

Problema tipo a resolver

- Conmutador con líneas de entrada y de salida
- Entradas usuarios finales o troncales: lo que nos importará es la cantidad de llamadas que llegan al conmutador
- Salidas troncales (máximo N llamadas simultáneas salen)
- Decidir N para poder cursar las llamadas con una probabilidad de bloqueo máxima objetivo
- o decidir la cantidad de llamadas que puede cursar para un N y ese máximo bloqueo



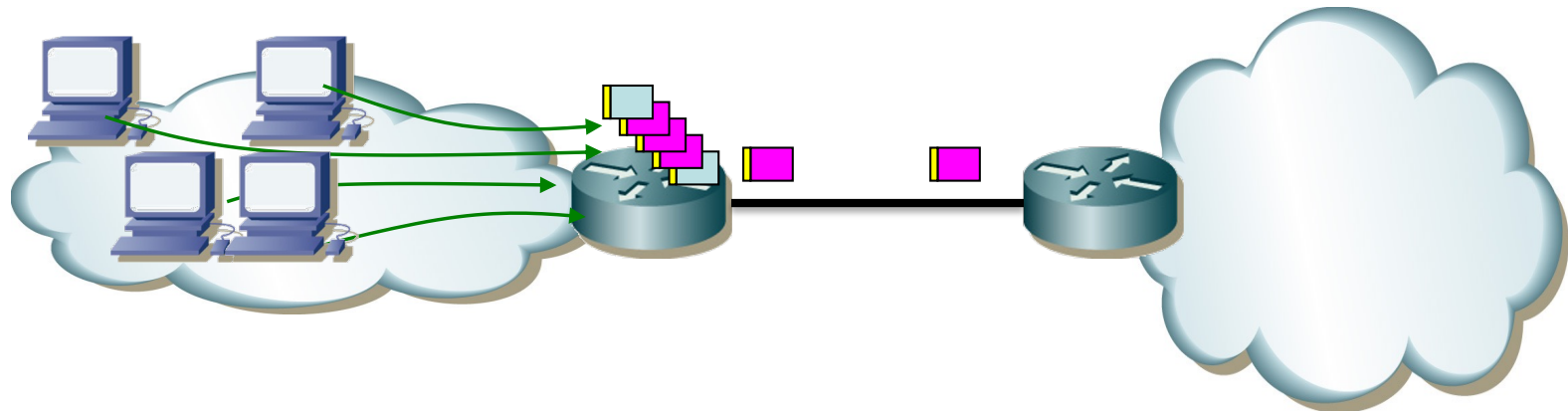
Otro tipo de problema

- Un servidor web *single-threaded*
- Recibe peticiones de ficheros que debe obtener del disco duro
- El S.O. atiende las peticiones en serie, completando una antes de atender a la siguiente
- Si el disco está ocupado el hilo del servidor web se bloqueará a la espera de que el disco finalice
- El disco es capaz de servir datos a C Mbps
- Normalmente en estos casos el S.O **encola** peticiones que no pueden ser atendidas en el momento (teoría de colas)



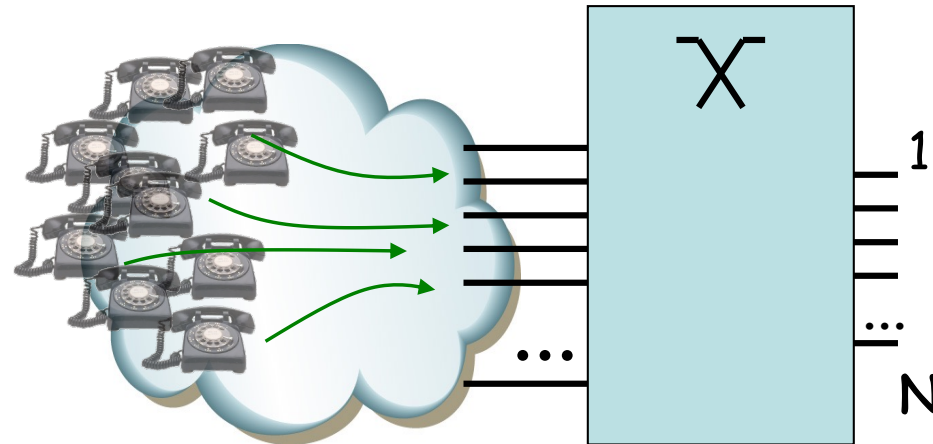
Otro tipo de problema

- Enlace entre dos conmutadores de paquetes
- Los usuarios envían paquetes
- Cada paquete monopoliza el enlace durante un tiempo proporcional a su tamaño
- Hay una memoria donde los paquetes acumulan retardo
- Si se excede la ocupación de la misma se descartan
- Teoría de colas

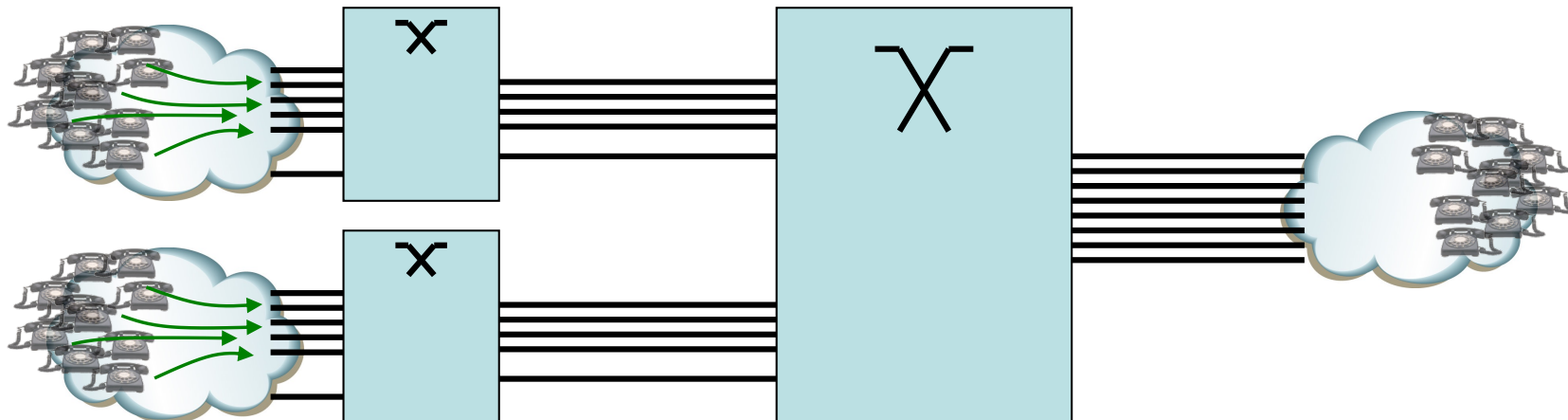


Problema tipo a resolver

- Conmutador con líneas de entrada y de salida



- Extensión:



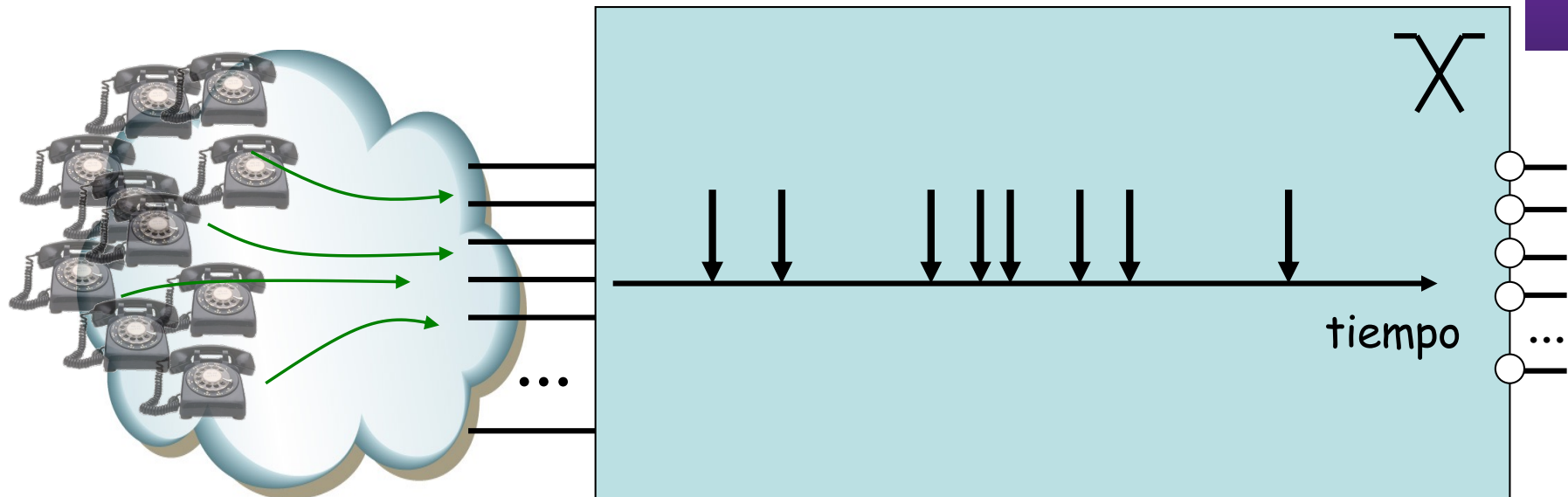
Proceso de llegadas

Proceso de llegadas

- Hipótesis fundamental en teoría clásica: llegadas independientes
- Tasa media de llegadas de llamadas de una gran población de fuentes (usuarios) independientes: λ
- Proceso de Poisson

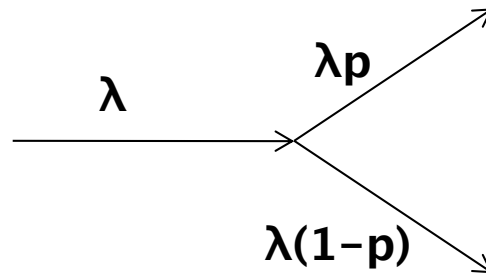
$$P_{\lambda\Delta t}[N = k] = \frac{(\lambda\Delta t)^k}{k!} e^{-\lambda\Delta t}$$

$$p_X(t) = \lambda e^{-\lambda t} \quad (t > 0) \quad P[X < t] = 1 - e^{-\lambda t}$$



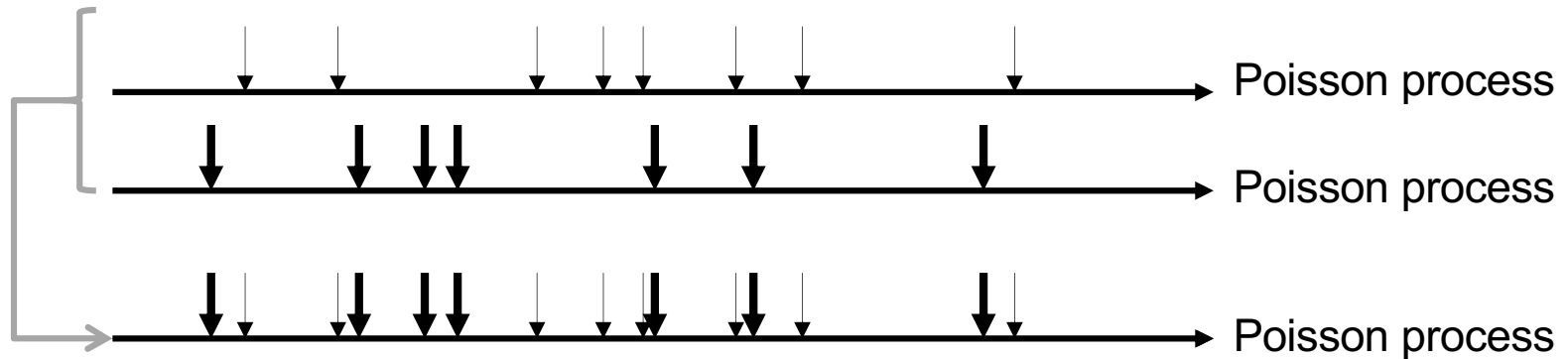
Random splitting

- Proceso de Poisson con tasa λ
- Repartidas las llegadas en dos grupos mediante Bernoulli de parámetro p
- Los procesos resultantes son procesos de Poisson de tasas λp y $\lambda(1-p)$



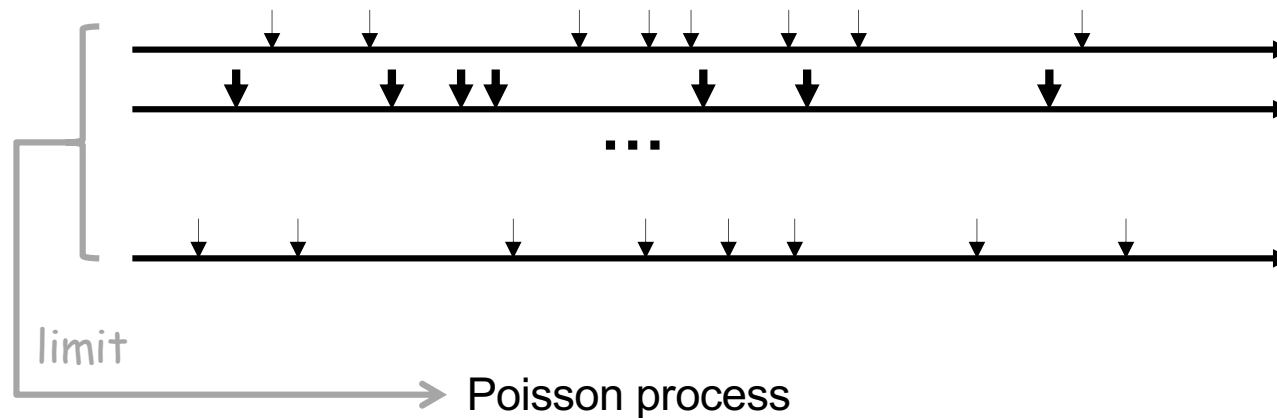
Superposición

- La superposición de dos procesos de Poisson es un proceso de Poisson de tasa la suma de las dos (...)



Superposición

- Para ciertos procesos muy comunes (independientes), la superposición de un gran número de ellos tiende a un proceso de Poisson

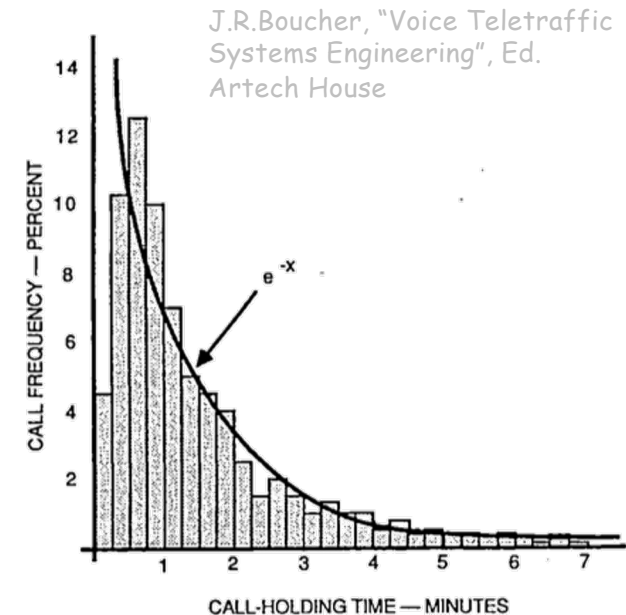


- Las peticiones de usuarios individuales es probable que no se puedan modelar con un proceso de Poisson
- El múltiplex de un gran número de usuarios independientes sí

Duración de las llamadas

Tiempo de ocupación

- Duración de las llamadas
- Lo más simple: tiempo constante
 - Poco realista para llamadas
 - Actividades automáticas: reproducción de mensajes, procesamiento de señalización, etc.
- Tiempo exponencial
 - Variables aleatorias (continuas) 's'
 - i.i.d. ('s')
 - Tiempos menores de la media muy comunes
 - Cada vez menos comunes tiempos mayores que la media
 - Propiedad: el tiempo restante de una llamada es independiente de lo que haya durado hasta ahora
- Duración exponencial: 's' caracterizada por su función de densidad



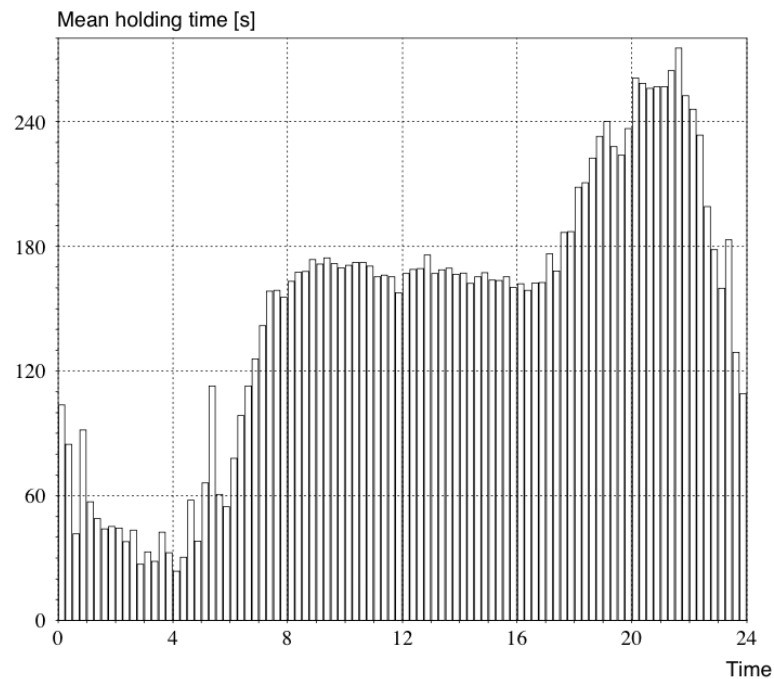
$$p_s(t) = \mu e^{-\mu t} \quad (t > 0)$$

$$\int_0^{\infty} \mu e^{-\mu t} dt = 1 \quad \text{es una fdp}$$

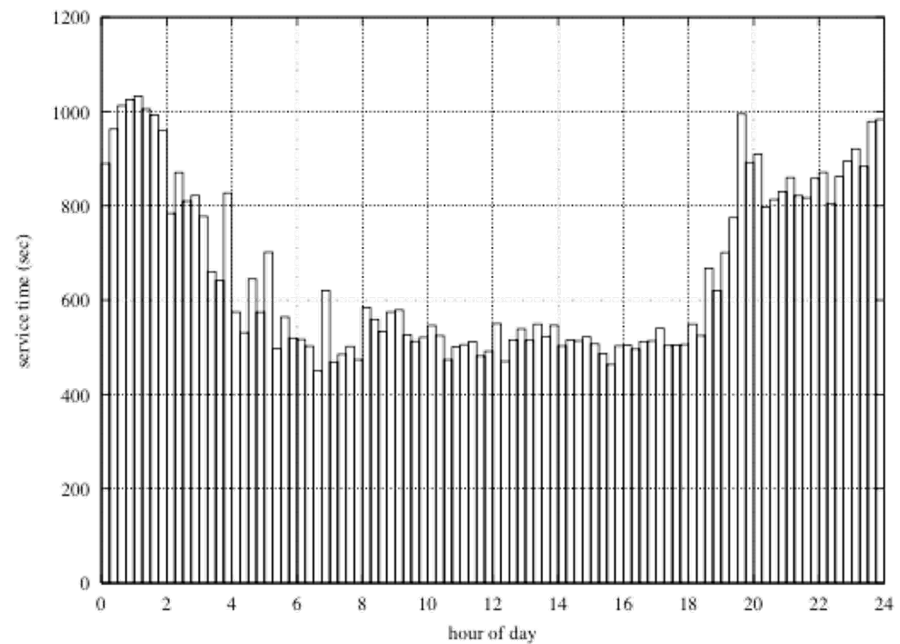
$$\bar{s} = E[s] = 1/\mu$$

Tiempo de ocupación

- La duración media de las llamadas también tiene cierta variación con la hora del día



Llamadas de voz



Llamadas de modem