Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Modelado de usuarios

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

Arquitectura de Redes, Sistemas y Servicios

Carga o tráfico

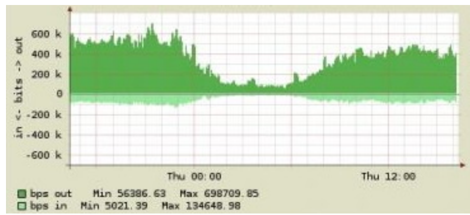
Definiciones

Capacidad

- Recursos de un sistema para dar un servicio, número de líneas de salida…
- Ej: nuestra centralita tiene 5 líneas para llamadas salientes

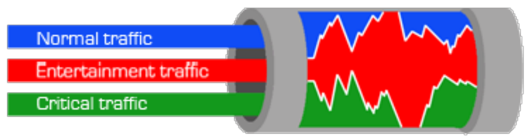
Carga (Intensidad de tráfico)

- Cantidad de servicio demandada al sistema, medida como cantidad de recursos necesarios en un determinado momento
- Ej: nuestra centralita recibe en media 3.2 llamadas por minuto



Calidad de servicio

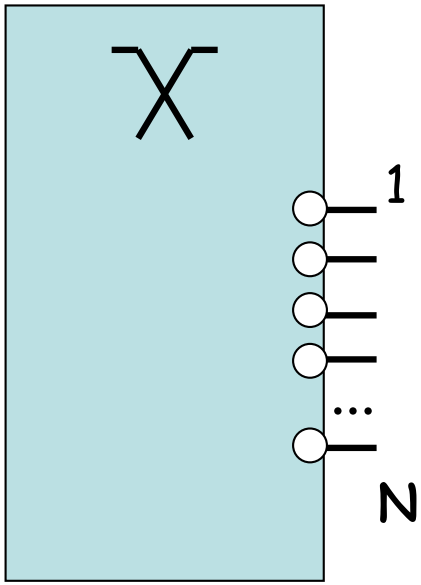
- Medida del servicio obtenido del sistema
- Ej: nuestra centralita, con las líneas de entrada que tenemos y la carga típica que soporta, pierde menos del 0.1% de las llamadas



A continuación, en más detalle…

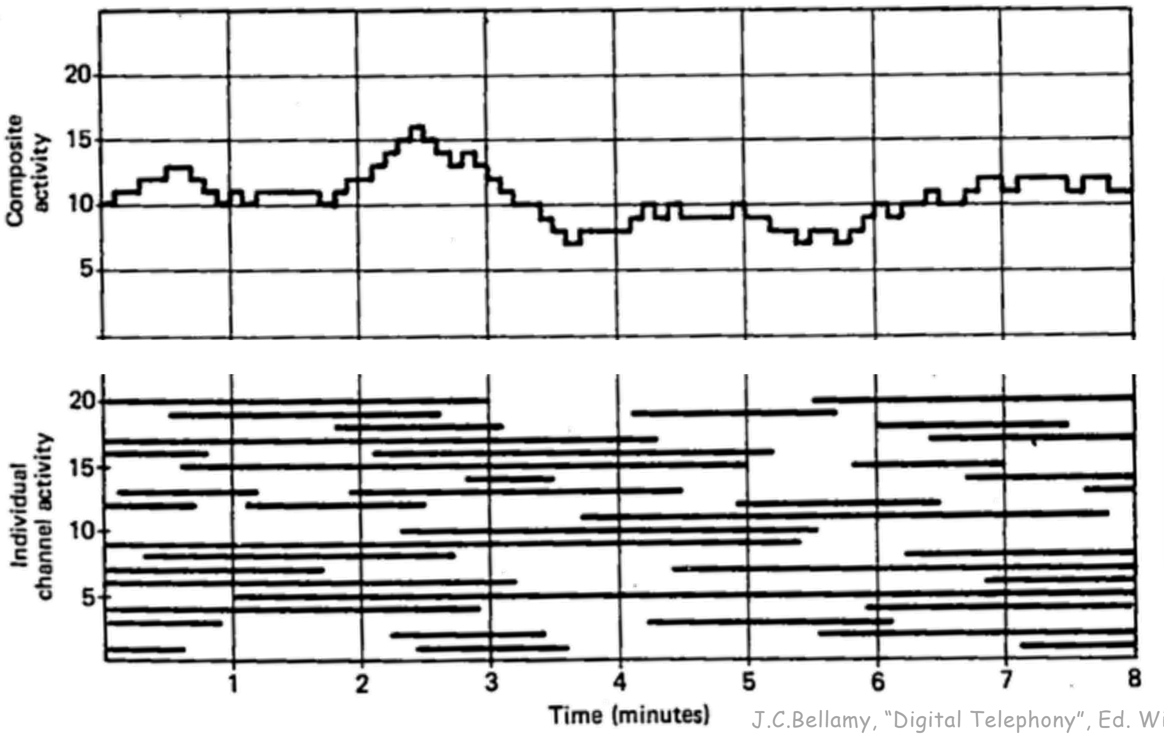
Capacidad

- 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

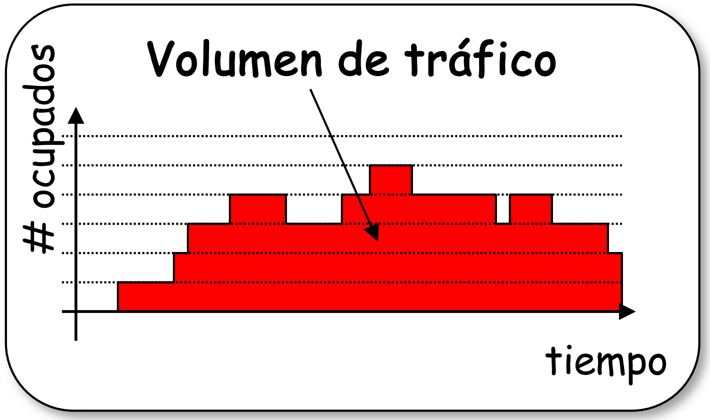


Carga o Tráfico

- 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

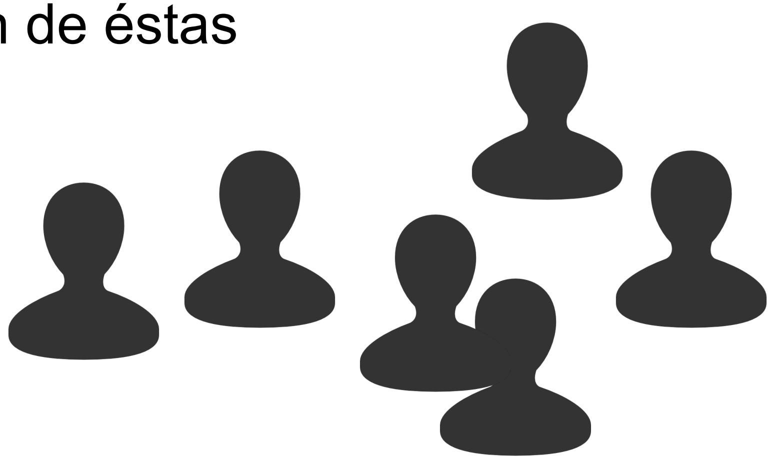


J.C.Bellamy, "Digital Telephony", Ed. Wiley Interscience



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



Medida del Tráfico

- Intensidad de tráfico

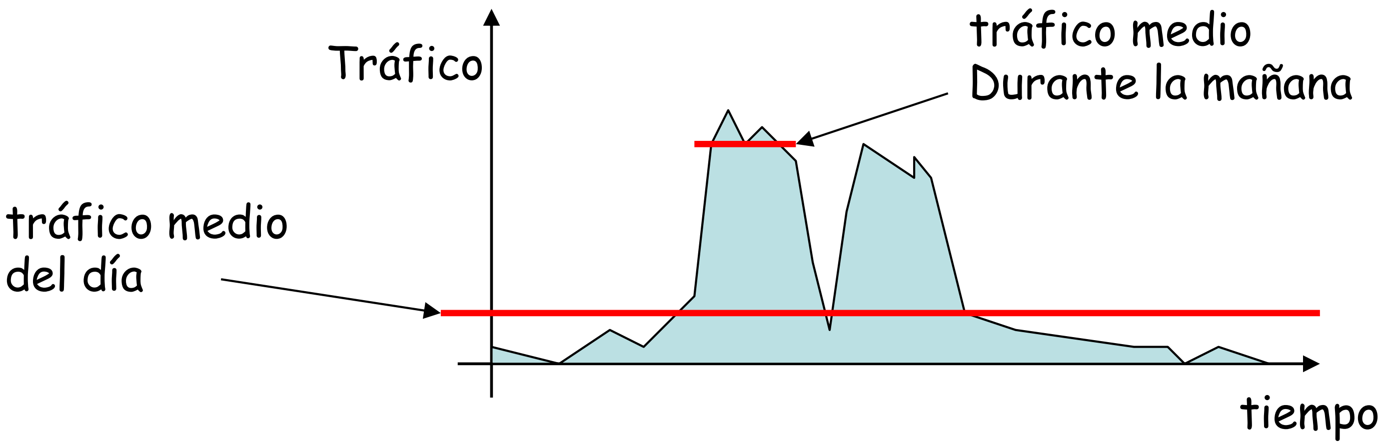
$$I = \frac{\text{Volumen de tráfico}}{\text{Tiempo de observación}} = \frac{\text{Tiempo acumulado de ocupación}}{\text{Tiempo de observación}}$$

- 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

Hora cargada
Tráfico ofrecido y cursado

Medida del Tráfico

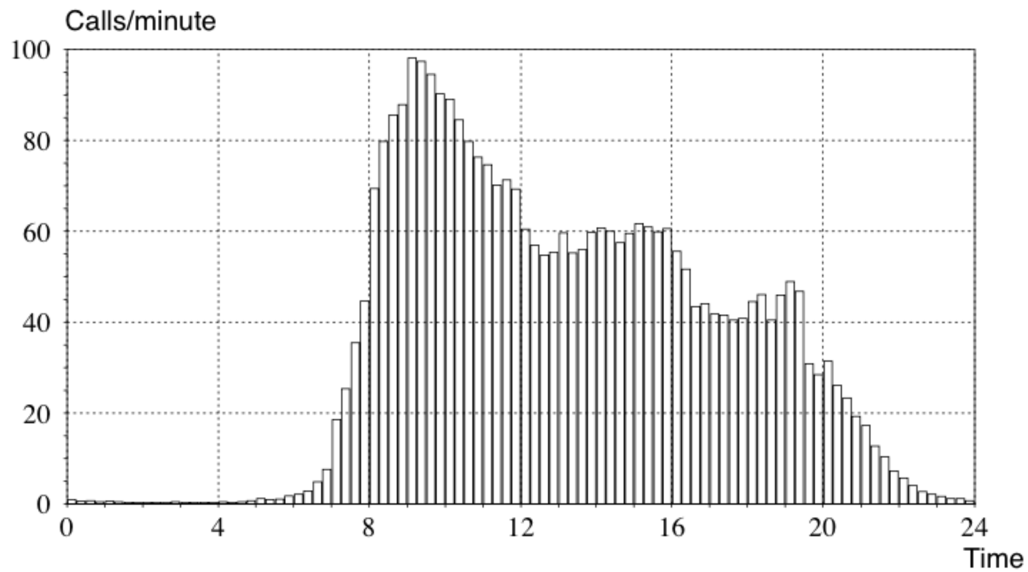
- Normalmente la intensidad del tráfico varía con el tiempo (no es un proceso estocástico estacionario) pero se puede considerar estable en un tiempo limitado



- 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



http://www.itu.int/ITU-D/study_groups/SGP_1998-2002/SG2/StudyQuestions/Question_16/RapporteursGroupDocs/teletraffic.pdf

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 $\rightarrow$ Menos calidad de servicio (*congestion theory*)
  - Se hace esperar la llamada hasta que se libere una línea $\rightarrow$ 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

Calidad de servicio

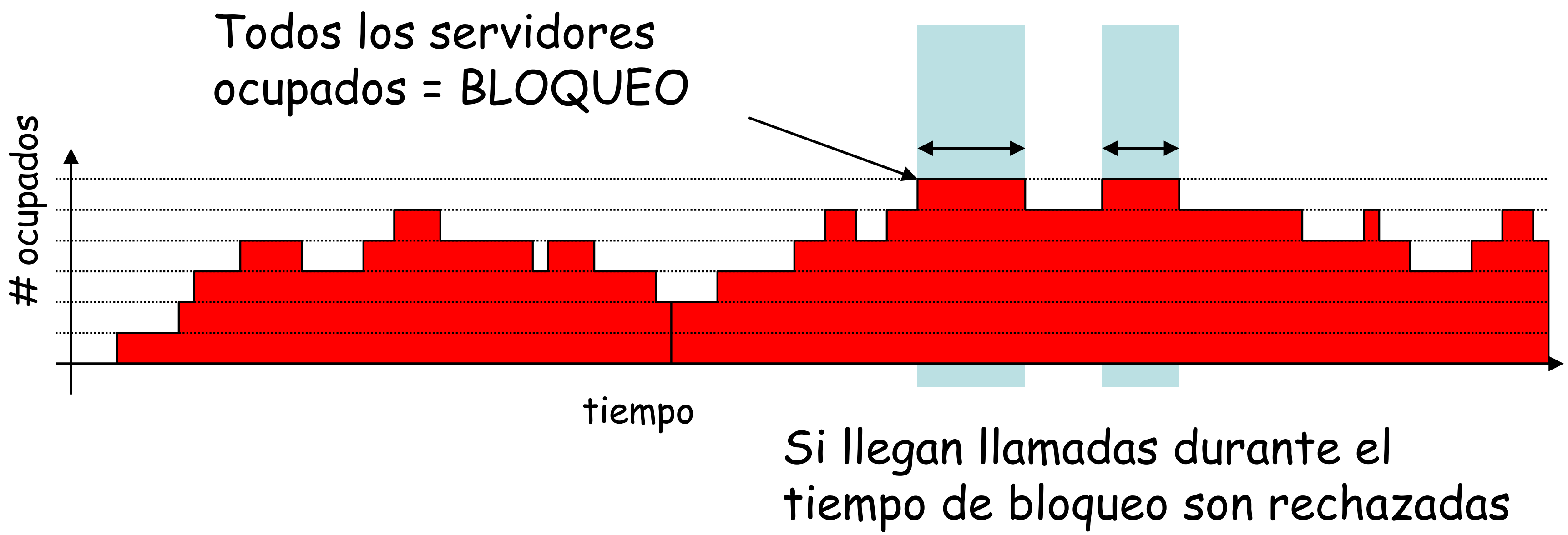
- Se suele distinguir:
  - Sistema en **situación de Bloqueo**

    Todos los recursos están ocupados y una llamada nueva que llegue será rechazada
  - Sistema en **situación de Congestión**

    Se han empezado a rechazar llamadas

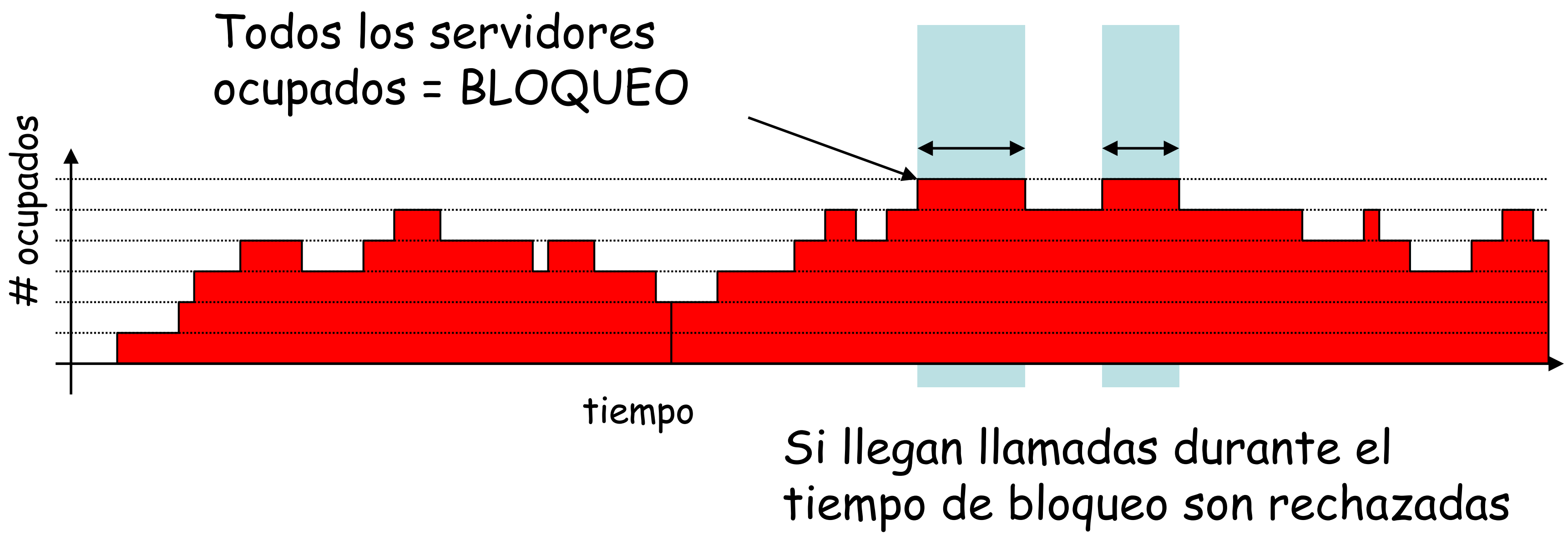
Tráfico ofrecido vs cursado

- 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

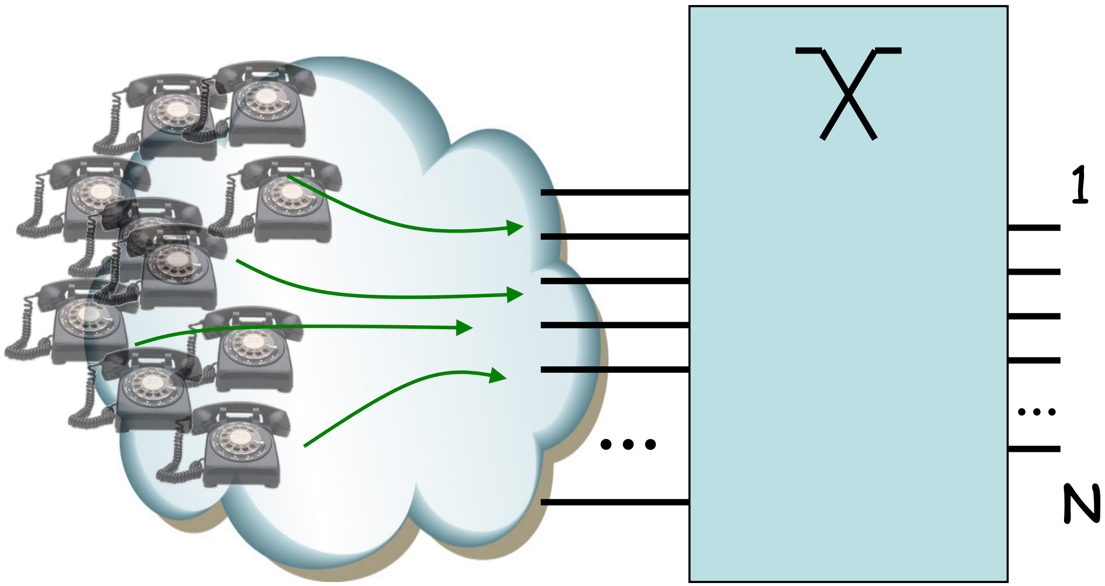
I_{in} : tráfico ofrecido o de entrada

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

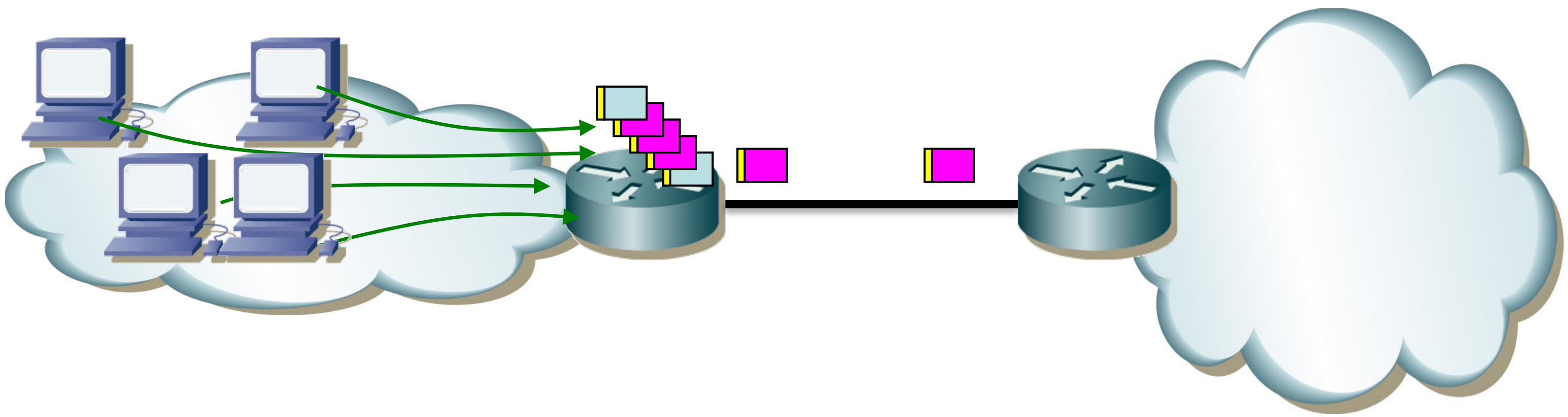


Otro tipo de problema

- 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)

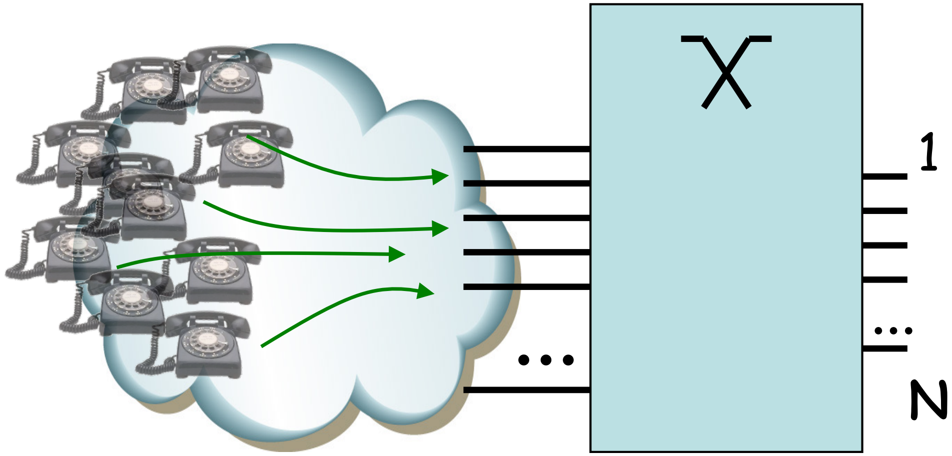
Otro tipo de problema

- 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

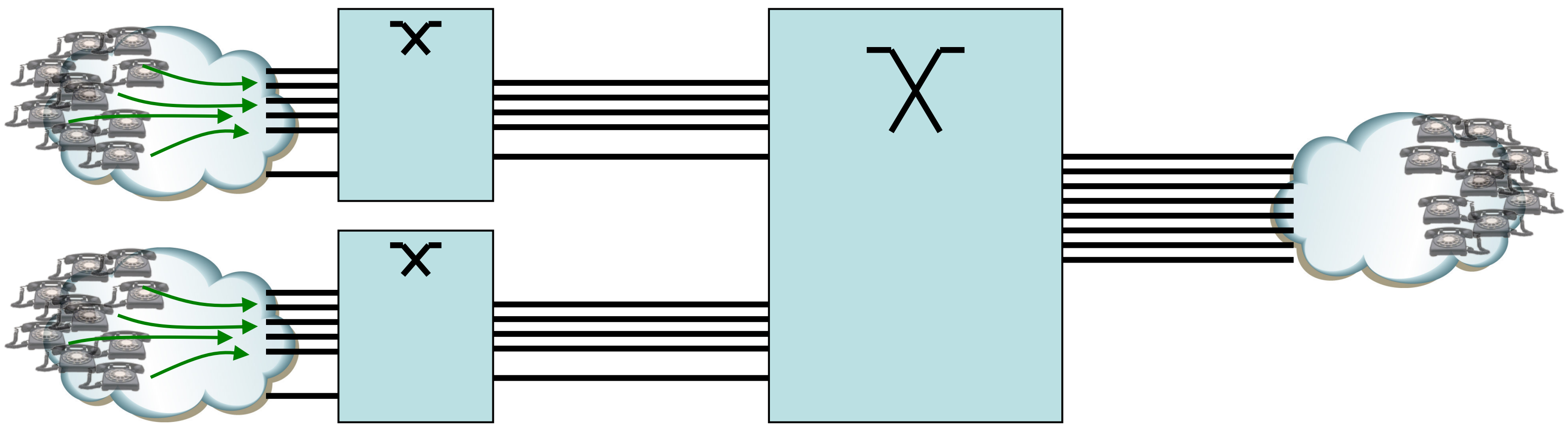


Problema tipo a resolver

- Conmutador con líneas de entrada y de salida



- Extensión:



Caracterización estadística

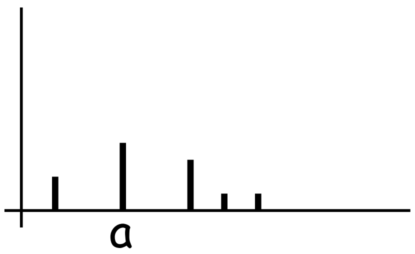
Modelando la carga

Variable aleatoria (V)

- No tiene un valor sino que describe el resultado aleatorio de un experimento
- Se caracteriza por la descripción de los posibles resultados que puede tomar en términos de probabilidad
- **Función de distribución / densidad de probabilidad**

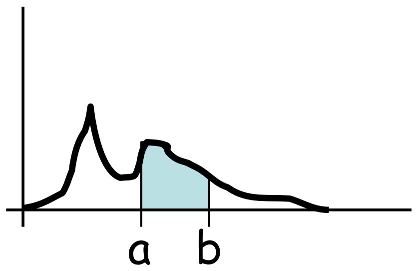
Variable discreta

$$p(a) = P[V = a]$$



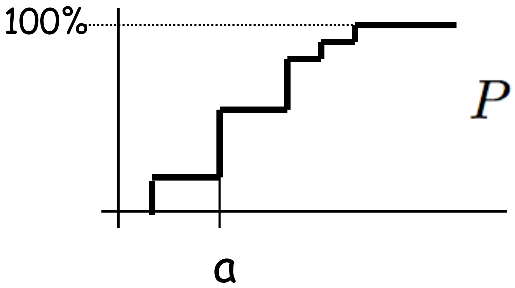
Variable continua

$$P[a < V < b] = \int_a^b p(t)dt$$



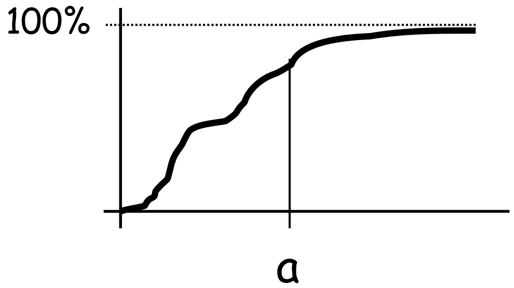
- **Función acumulada de probabilidad / distribución**

Variable discreta



$$P[V \leq a] = F(a)$$

Variable continua



Modelando la carga

Procesos estocásticos (V)

- Una familia de variables aleatorias

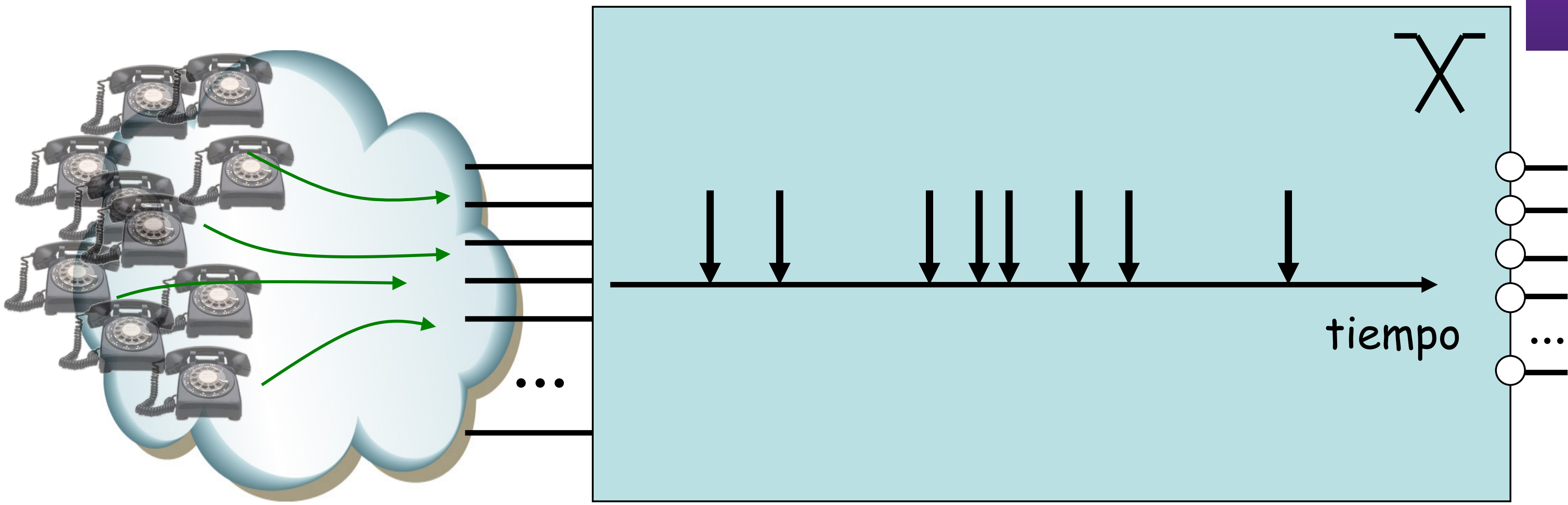
$$\left\{X_t : t \in T\right\}$$

- Hablaremos de
  - “Tiempo continuo” cuando T es real, por ejemplo T = [0,∞]
  - “Tiempo discreto” cuando T es numerable, por ejemplo T = {0,1,2…}

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

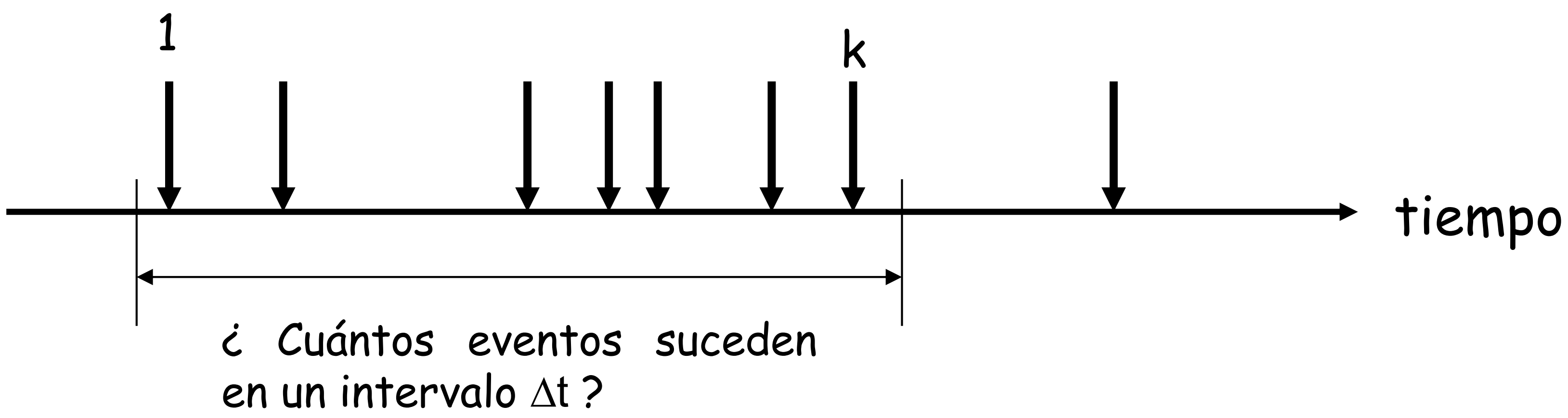


upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Número de llegadas

- Hipótesis:
  - En un intervalo suficientemente pequeño solo puede producirse una llegada
  - La probabilidad de una llegada en un intervalo suficientemente pequeño es directamente proporcional a la longitud del mismo (probabilidad $\lambda\Delta t$)
  - La probabilidad de una llegada en un intervalo es independiente de lo que suceda en otros intervalos
- Se demuestra que el número de llegadas en un intervalo sigue una distribución de Poisson

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



Tiempos entre llegadas

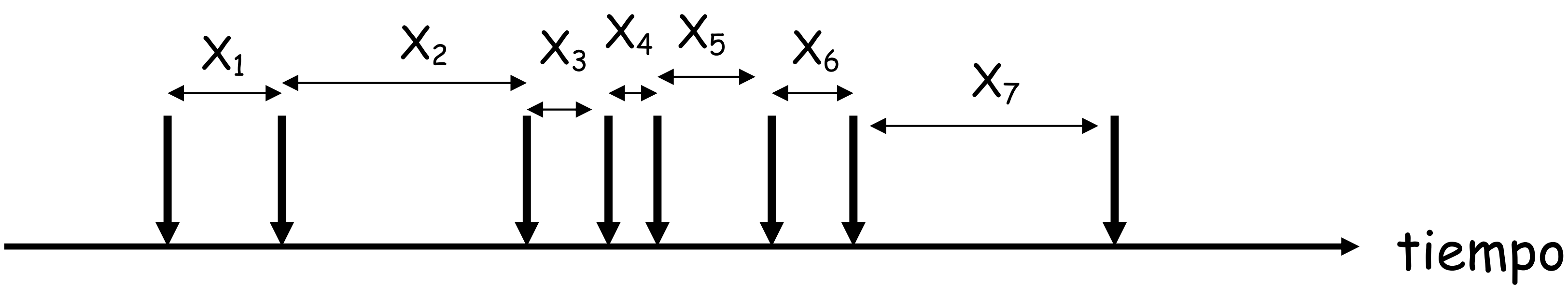
- Se demuestra que: si el número de eventos que ocurren en un intervalo cualquiera sigue una distribución de Poisson, los tiempos entre llegadas de eventos siguen una distribución exponencial
- El tiempo entre llegadas sigue una v.a. exponencial de parámetro λ
- X_i variables aleatorias independientes e idénticamente distribuidas (i.i.d.) ('X')

$$\boxed{p_X(t) = \lambda e^{-\lambda t} \quad \text{(t>0)}}$$

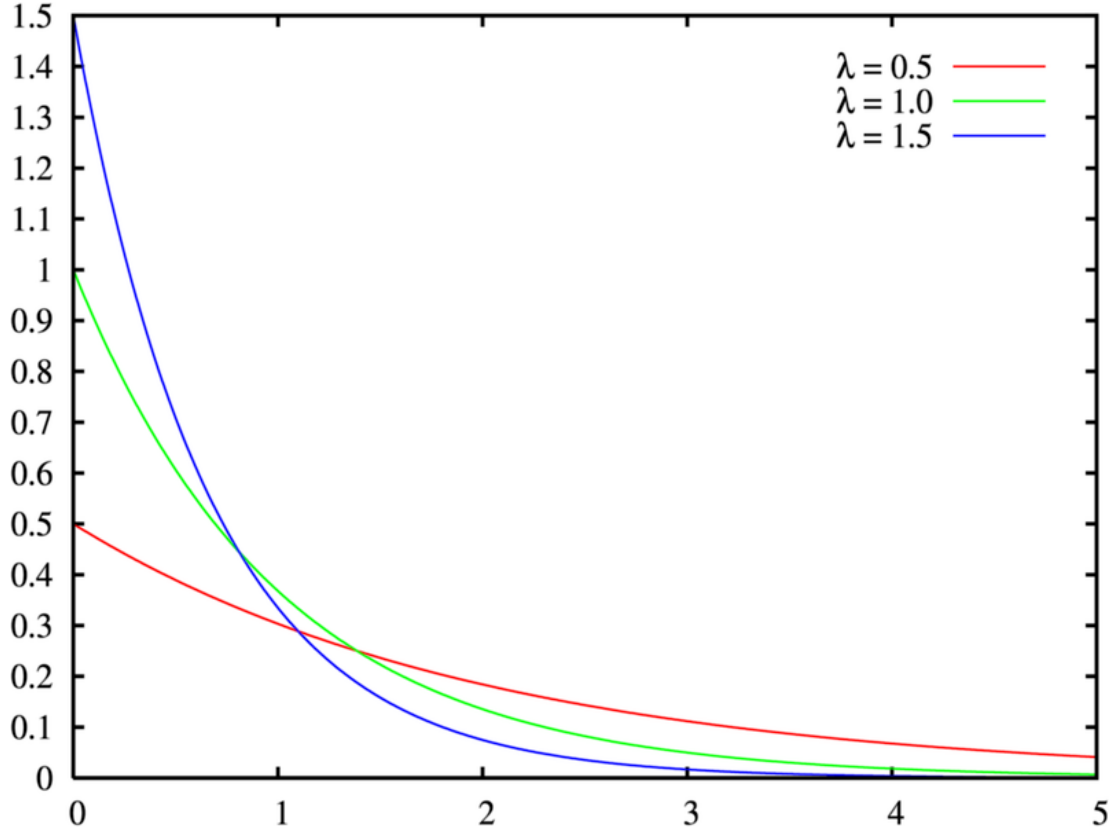
$$P[X < t] = 1 - e^{-\lambda t}$$

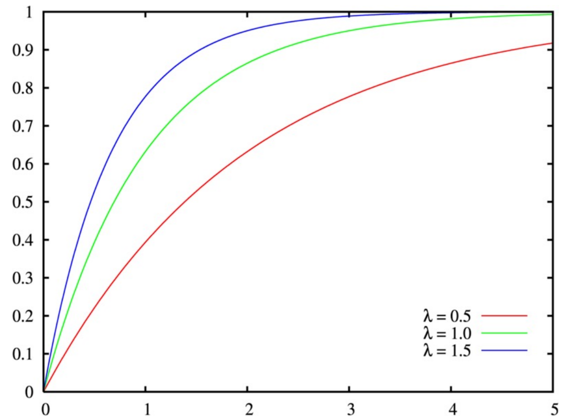
- Media: $E[X] = \int_0^\infty t\lambda e^{-\lambda t} = 1/\lambda$

- Tiempo medio entre llegadas $1/\lambda \Rightarrow$ en media λ llegadas por segundo

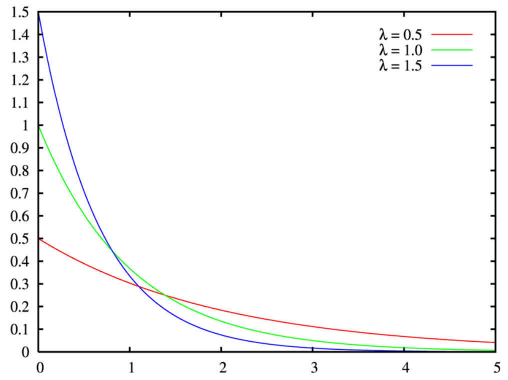


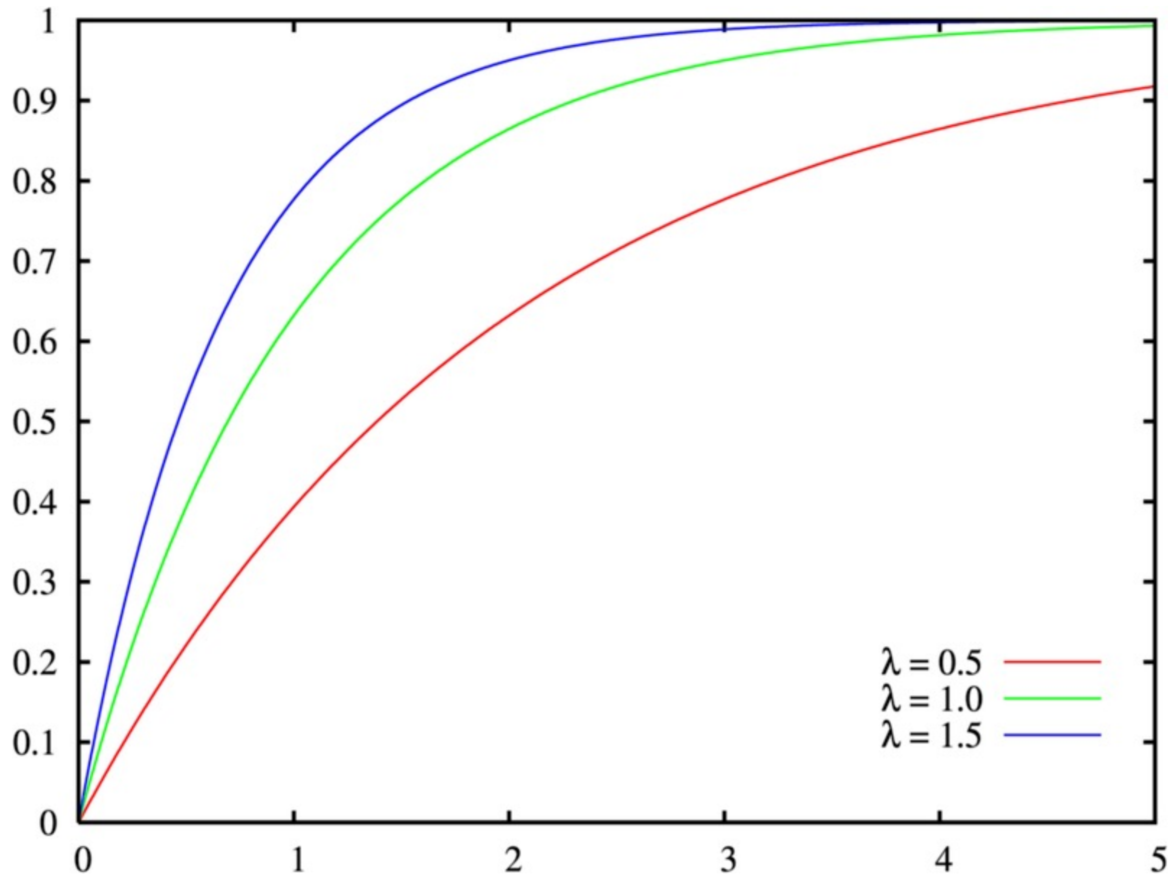
Ejemplo (exponencial)



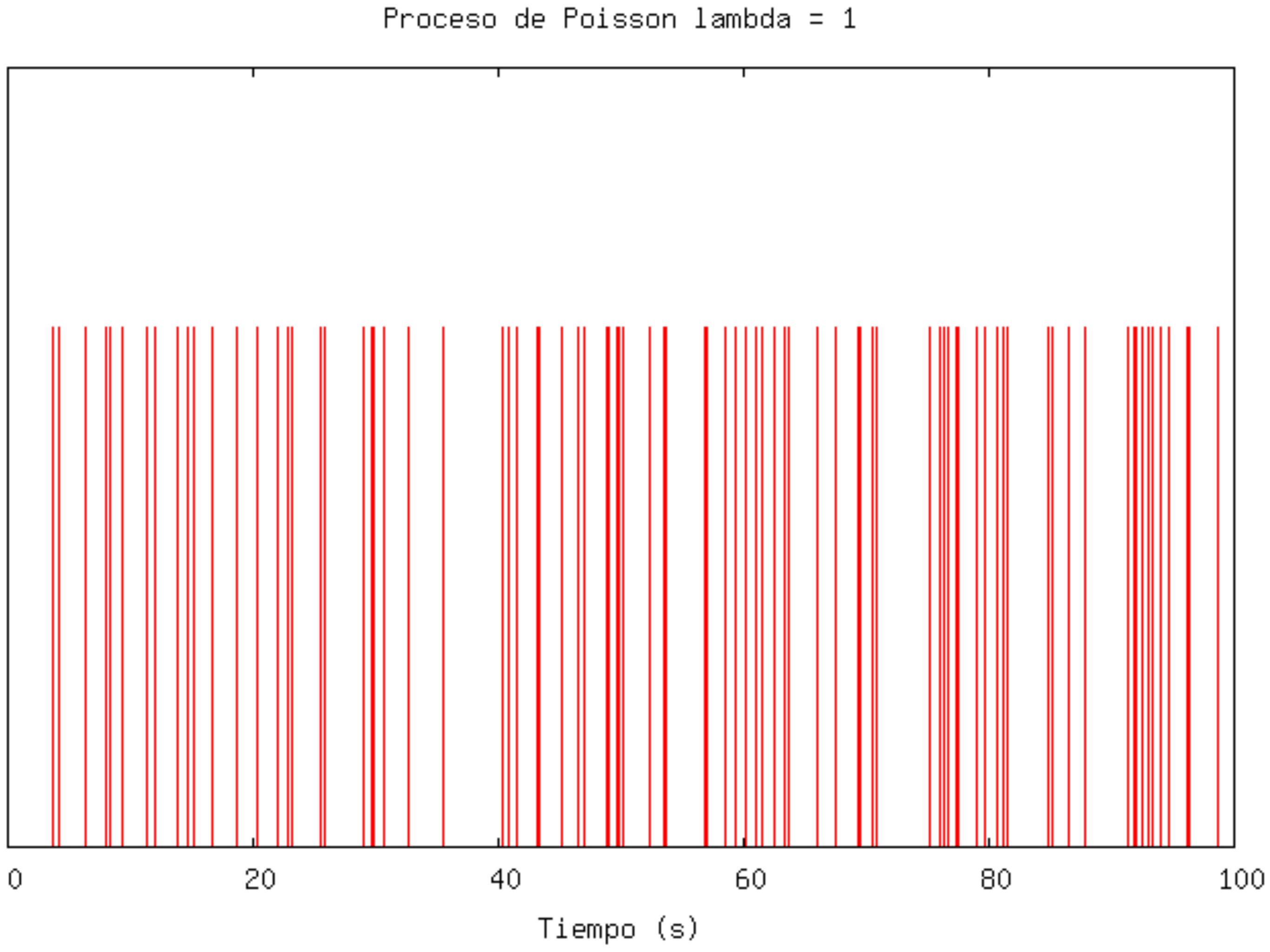


Ejemplo (exponencial)





Ejemplo (proceso de Poisson)



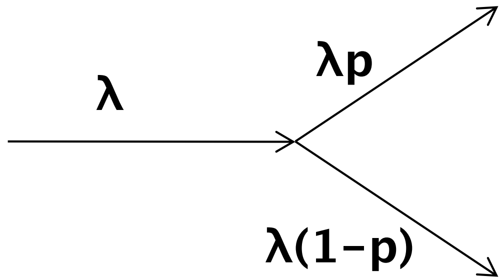
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

ARQUITECTURA DE REDES, SISTEMAS Y SERVICIOS
Área de Ingeniería Telemática

Propiedades del proceso de Poisson

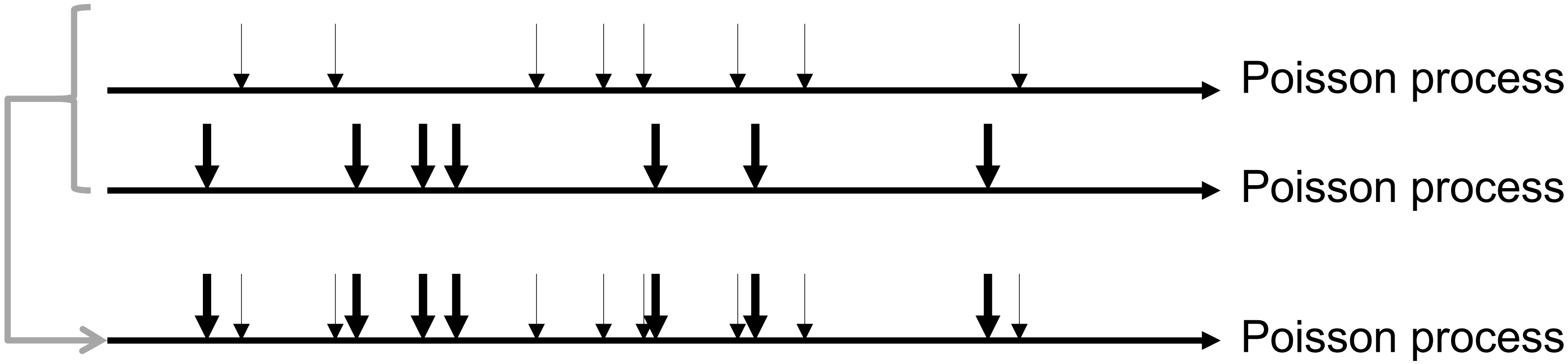
Random splitting

- 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 λ(1-p)



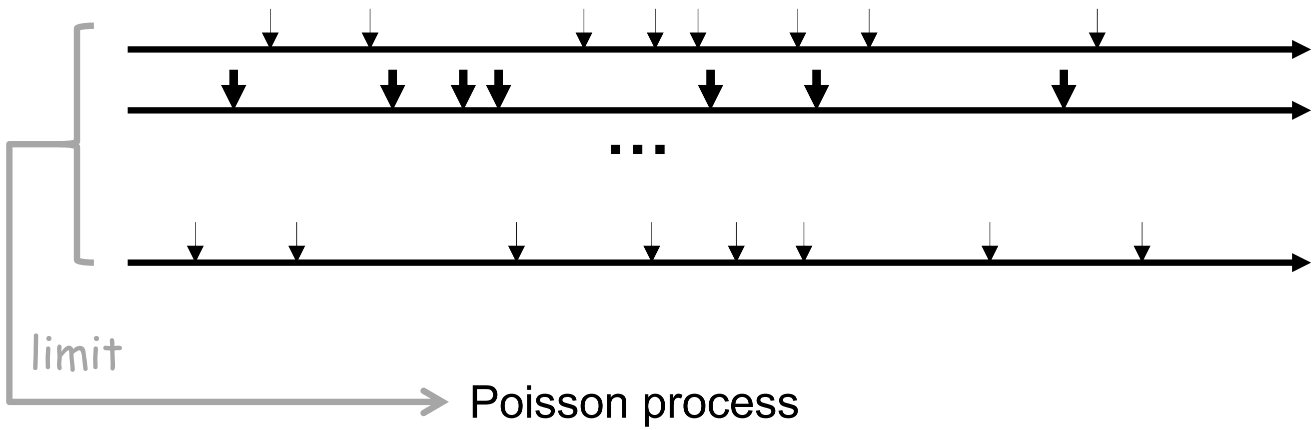
Superposición

- 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



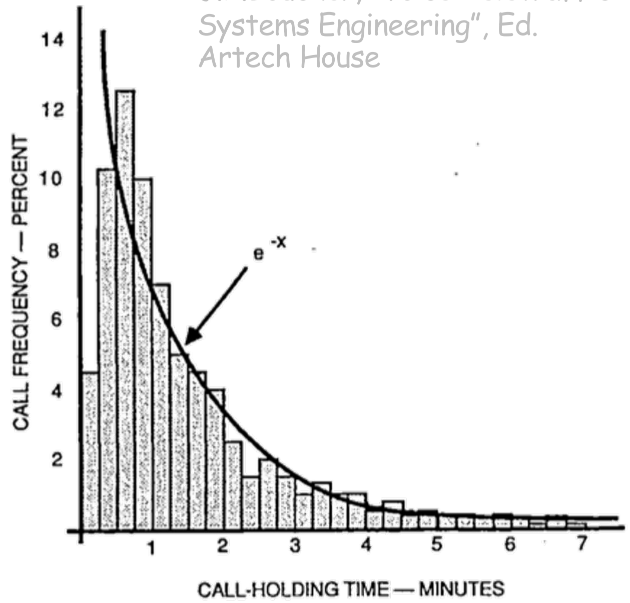
- Las peticiones de usuarios individuales es probable que no se puedan modelar con un proceso de Poisson
- El múltiplex de un gran número de usuarios independientes sí

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, procesado de señalización, etc.
- Tiempo exponencial
  - Variables aleatorias (continuas) 's_i'
  - 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

J.R.Boucher, "Voice Teletraffic Systems Engineering", Ed. Artech House



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

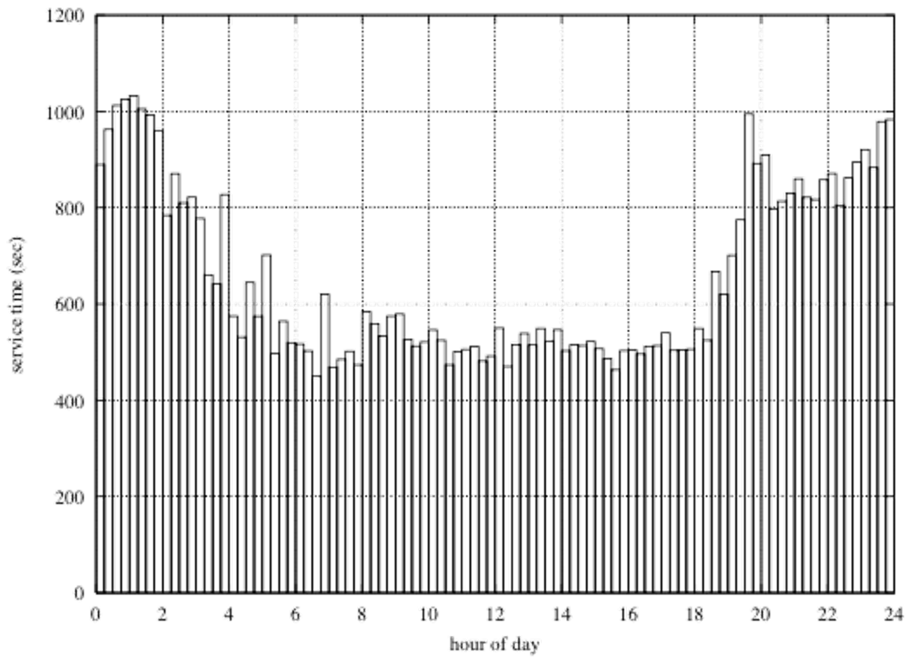
$$\int_0^\infty \mu e^{-\mu t} = 1 \qquad \text{es una fdp}$$

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

Tiempo de ocupación

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

Llamadas de voz



Llamadas de modem

http://www.itu.int/ITU-D/study_groups/SGP_1998-2002/SG2/StudyQuestions/Question_16/RapporteursGroupDocs/teletraffic.pdf