

# Ingeniería de Tráfico: Introducción

Jhon Jairo Padilla Aguilar, PhD.

# Definición de Teletráfico

---

- ▶ Es la aplicación de la teoría de Probabilidades a la solución de problemas concernientes a la planeación, evaluación del rendimiento, operación y mantenimiento de los sistemas de telecomunicaciones.



# Herramientas del Teletráfico

---

- ▶ Procesos estocásticos
- ▶ Teoría de colas
- ▶ Simulación por computador
- ▶ Mediciones de tráfico



# Areas de Aplicación del Teletráfico

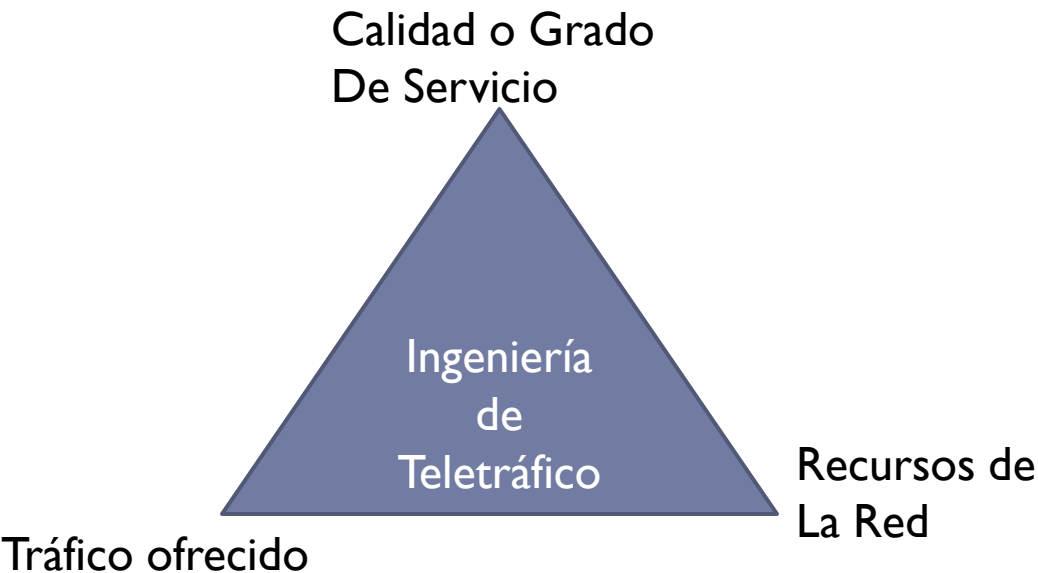
---

- ▶ El teletráfico se aplica tanto a redes de datos como de voz.
- ▶ En principio, este curso se enfoca a redes telefónicas y redes de datos.
- ▶ Sin embargo, estas herramientas son aplicables a otras áreas del conocimiento: tráfico de carreteras, tráfico aéreo, manufactura, líneas de ensamble, distribución, almacenamiento, etc.



# Objetivo del Teletráfico

---



- ▶ Diseñar sistemas con un costo mínimo y con una capacidad tal que cumplan con un grado de servicio pre-definido y satisfagan la demanda de tráfico futura
- ▶ Se deben obtener modelos que describan la relación entre los tres elementos.



# Aplicaciones del Teletráfico

---

- ▶ **Sistemas telefónicos convencionales (conmutación de circuitos)**
- ▶ **Sistemas de comunicación satelital**
- ▶ **Redes de conmutación de paquetes**
- ▶ **Métodos de control de Acceso en LANs**
- ▶ **Internet:**
  - ▶ **Búsqueda de direcciones en los Routers**
  - ▶ **Clasificación de paquetes en Routers**
  - ▶ **Planificación de paquetes en Routers**
  - ▶ **Algoritmos de Enrutamiento**
- ▶ **Redes Celulares:**
  - ▶ **Distribución del tráfico en las celdas**
  - ▶ **Criterios de asignación de recursos en celdas (Control de Acceso para llamadas nuevas y de Handoff)**
  - ▶ **Métodos de Acceso Múltiple**
  - ▶ **Modelos de movilidad de usuarios**



# Componentes del Modelo de un sistema de Telecomunicaciones

---

## ▶ Estructura:

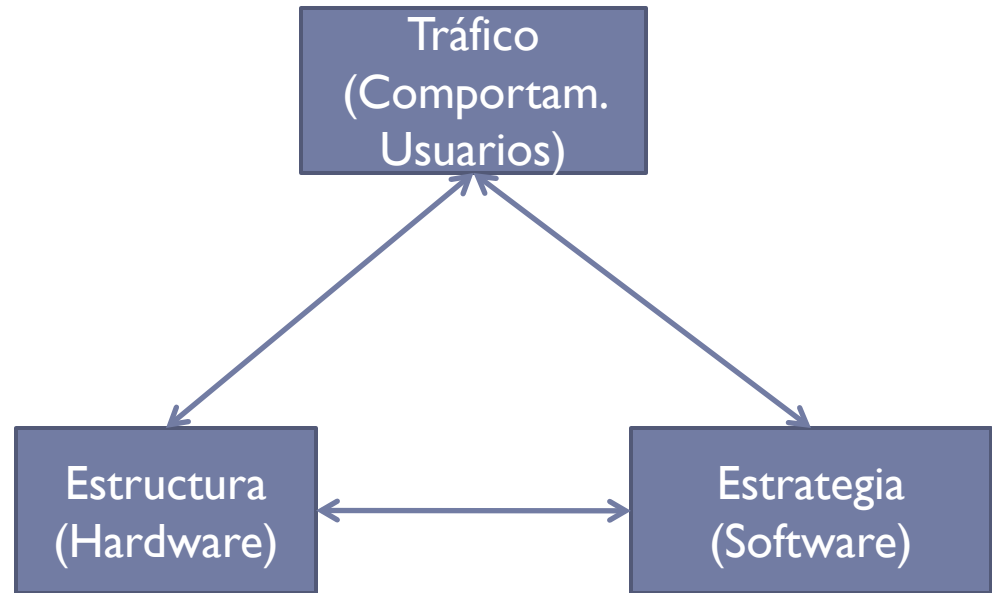
- ▶ Es la parte técnica. Tiene que ver con la forma de construcción del dispositivo o sistema a modelar. Se puede obtener de manuales, estándares, etc.

## ▶ Estrategia:

- ▶ Tiene que ver con el manejo de los elementos a tratar: llamadas, paquetes, flujos de paquetes, etc. Por lo general, se materializan como programas de control de los dispositivos del sistema.

## ▶ Tráfico:

- ▶ La generación de elementos a procesar por el sistema (llamadas, paquetes) depende en gran medida del comportamiento de los usuarios. Este comportamiento suele ser aleatorio.
- ▶ También debe modelarse el tiempo que tarda el servicio que se le da al elemento. Este tiempo puede ser aleatorio. Ej: tiempo de duración de una llamada.



# Conceptos de Tráfico

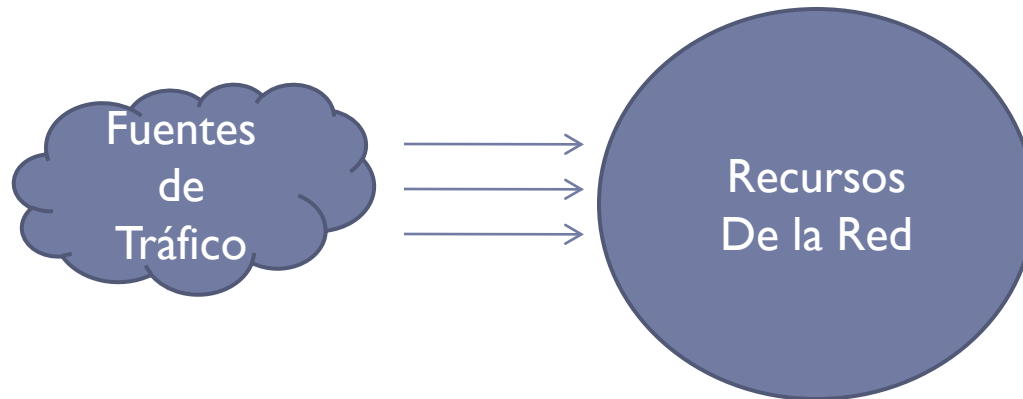
Jhon J. Padilla A., PhD.



# Situación a Modelar

---

- ▶ Fuentes: Suelen ser los usuarios (teléfonos, computadores, etc) que generan solicitudes de comunicación (llamadas, paquetes, sesiones, etc)
- ▶ Recursos: Dan servicio (atienden) las solicitudes. Pueden ser frecuencias, ranuras de tiempo, servidores, etc. Son limitados en cantidad y capacidad debido a sus altos costos.



# Concepto de Tráfico

---

- ▶ **Definición (ITU-T, 1993):**

- ▶ La intensidad de tráfico instantánea en un conjunto de recursos es el número de recursos ocupados en un instante de tiempo dado.

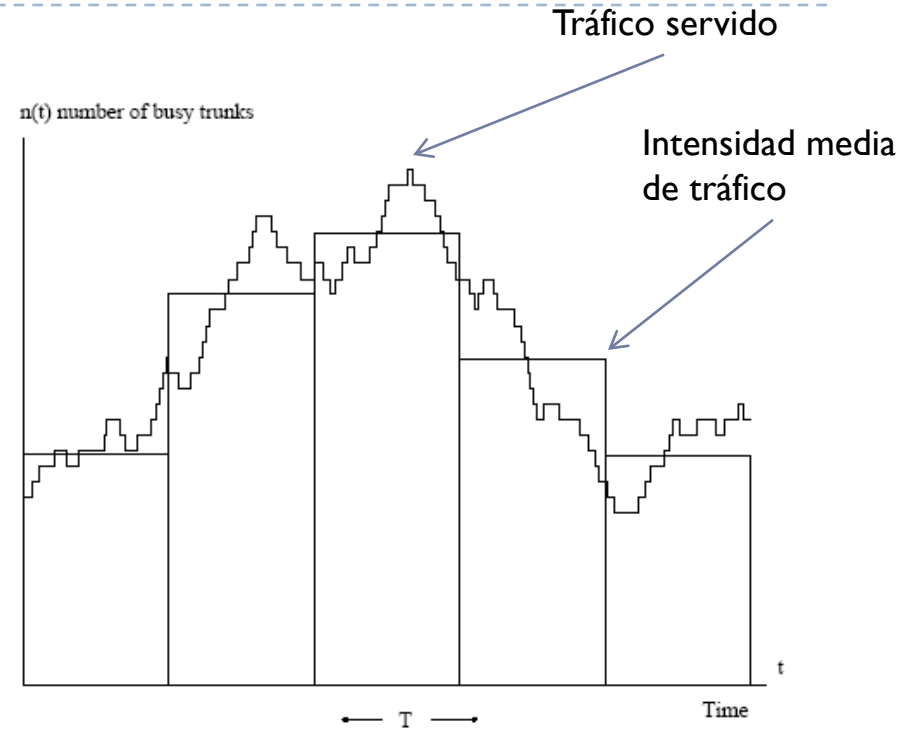
- ▶ **Recursos:**

- ▶ Servidores
- ▶ Líneas troncales
- ▶ Radio-canales
- ▶ etc



# Tráfico Servido

- ▶ El tráfico Servido (Carried Traffic) es el tráfico transportado por el grupo de servidores durante un intervalo de tiempo  $T$ .
- ▶ Usualmente, el término “Intensidad de tráfico” tiene el significado de “Intensidad Media de Tráfico” y se denotará por  $A_c$



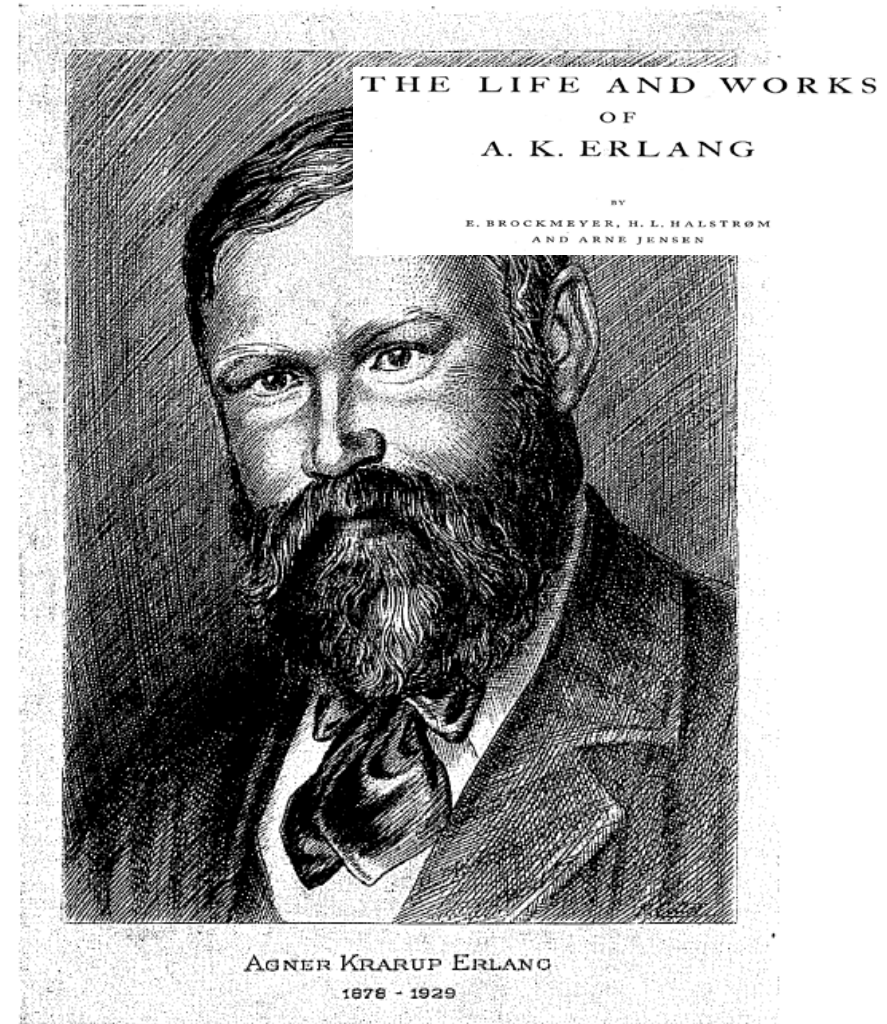
Fuente: Iversen (2001)

$$Y(T) = \frac{1}{T} \cdot \int_0^T n(t) dt.$$

$$A_c = Y = A'$$

# Unidad de medida del Tráfico

- ▶ La unidad utilizada por la ITU-T es el Erlang (E) y es adimensional
- ▶ Nombre dado en 1946 en honor al matemático Danés A. K. Erlang, fundador de la teoría de tráfico telefónico con su trabajo “*The Theory of Probabilities and Telephone Conversations*”, *Nyt Tidsskrift for Matematik B*, Vol 20, 1909.



# Unidades del Tráfico

---

- ▶ El tráfico total servido en un período de tiempo  $T$  es un **Volúmen de Tráfico** y es medido en *Erlangs-hora (Eh)*
- ▶ Según la ISO, la medida estándar es *Erlangs-segundo*, pero las unidades por hora son más adecuadas al orden natural de la práctica del Teletráfico.
- ▶ Advertencias:
  - ▶ El tráfico servido nunca puede exceder el número de recursos
  - ▶ Un recurso sólo puede transportar un Erlang como máximo



# Tráfico Ofrecido (Tráfico Potencial)

---

- ▶ Es el tráfico que podría ser transportado si no se rechazaran comunicaciones debido a la falta de capacidad, es decir, si el número de servidores fuese ilimitado.
- ▶ Es un valor teórico y no puede ser medido. Sólo es posible estimarlo del tráfico Servido.
- ▶ Sin embargo, puede verse como el tráfico que generan todos los usuarios del sistema a modelar

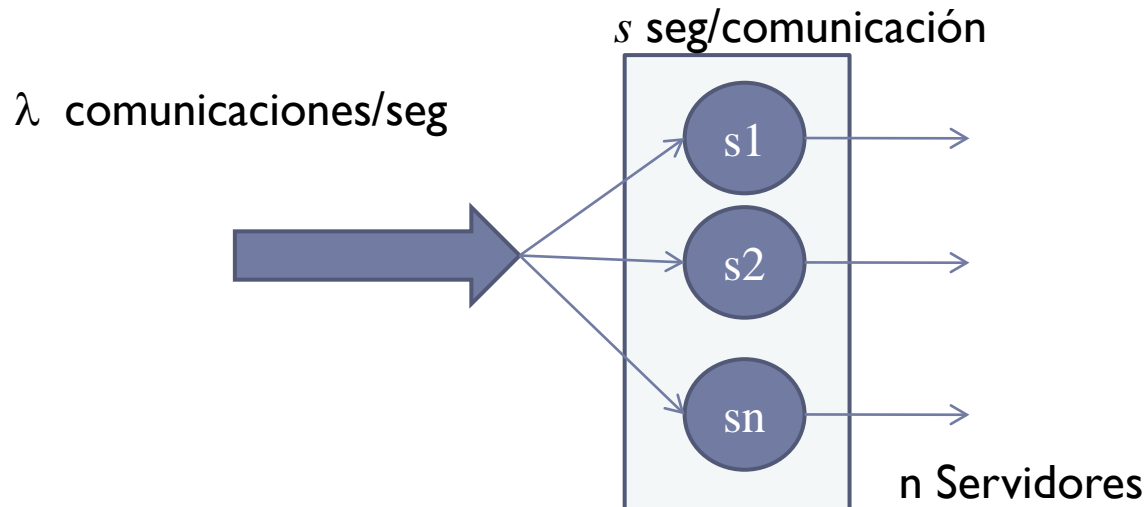


# Tráfico Ofrecido en sistemas telefónicos

- ▶ El tráfico ofrecido se calcula como

$$A = \lambda \cdot s$$

- ▶ Teóricamente, trabajamos con la intensidad de llamadas  $\lambda$ , que es el número medio de llamadas ofrecidas por unidad de tiempo, y el tiempo medio de servicio  $s$  por llamada.
- ▶ Se observa que es adimensional.
- ▶ Realmente, esta definición es más apropiada para el tráfico servido, pero se utiliza como si fuera el tráfico ofrecido (sería más apropiado decir para el tráfico ofrecido que  $s$  es la duración de las llamadas según la voluntad de los usuarios).



# Ejemplo

---

- ▶ Intensidad de llamadas: 5 llamadas por minuto
- ▶ Tiempo medio de la llamada: 3 minutos
- ▶ Tráfico ofrecido:
  - ▶  $A=5*3=15$  Erlangs

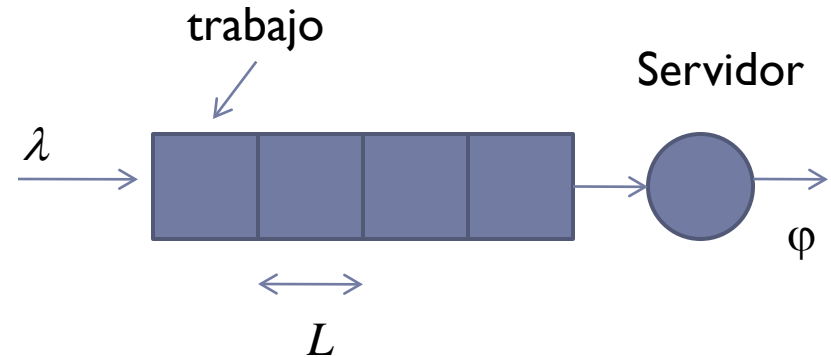




# Tráfico Ofrecido en Sistemas de transmisión de datos

- ▶ No se habla de llamadas sino de “trabajos” o “tareas” que son puestos en una cola para darles servicio
- ▶ Un trabajo puede ser la transferencia de  $L$  unidades (p.ej. bits o Bytes)
- ▶ La capacidad del sistema,  $\varphi$ , se mide en bits/segundo
- ▶ El tiempo de servicio ( $t_s$ ) es el tiempo de transmisión para un trabajo:  
$$t_s = \frac{L}{\varphi}$$
- ▶ Si en media llegan  $\lambda$  trabajos por unidad de tiempo, entonces la utilización del sistema es (tráfico en Erlangs):

$$\rho = \lambda t_s = \lambda \frac{L}{\varphi}$$



# Tráfico multi-tasa

---

- ▶ Sucede cuando una comunicación (llamada) ocupa más de un canal
- ▶ Una comunicación de tipo  $i$  ocupa  $d_i$  canales
- ▶ El tráfico ofrecido se expresa como:

$$A = \sum_{i=0}^N \lambda_i \cdot s_i \cdot d_i$$



# Tráfico perdido o rechazado ( $A_\ell$ )

---

- ▶ Es la diferencia entre el tráfico ofrecido y el tráfico servido.

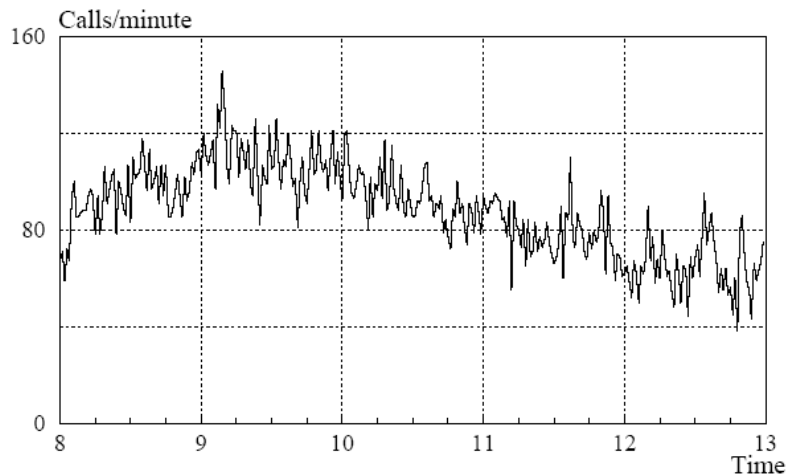
$$A_\ell = A - Y$$

- ▶ El tráfico rechazado puede reducirse si se aumenta la capacidad del sistema (se aumenta el número de recursos ó se reduce el tiempo de servicio)



# Variaciones de Tráfico

---



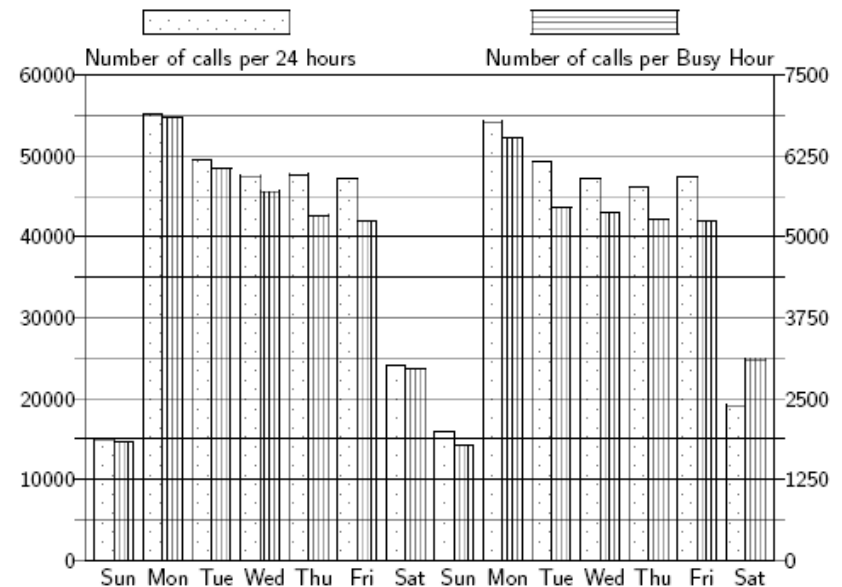
Fuente: Iversen (2008)

- ▶ El tráfico varía con dos componentes:
- ▶ **Aleatoria:** No es seguro exactamente cuántos usuarios harán llamadas en un determinado instante de tiempo
- ▶ **Determinista:** El comportamiento general de los usuarios diariamente es predecible debido a los ciclos de la vida cotidiana (negocios, vida personal), según las actividades de la sociedad.

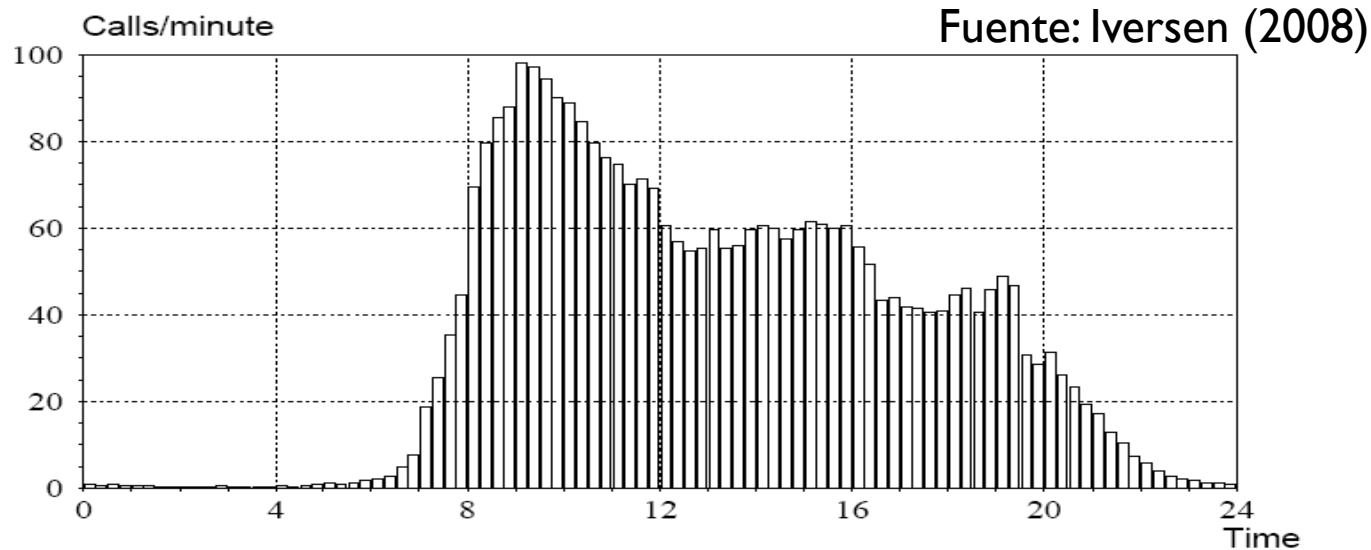


# Variaciones deterministas

- ▶ Pueden observarse en diferentes escalas de tiempo:
  - ▶ 24 horas:
    - ▶ Mayor tráfico en la hora pico.
  - ▶ Semanales :
    - ▶ Lunes (mayor tráfico), sigue viernes; luego martes, miércoles, jueves.
  - ▶ Anual: (Mayor tráfico)
    - ▶ Al inicio de cada mes
    - ▶ Después de vacaciones



# Hora Pico (Busy Hour)



- ▶ El comportamiento diario en general (la tendencia) del tráfico es predecible.
- ▶ Existe una hora de mayor tráfico (Busy Hour). En el ejemplo es aproximadamente :
  - ▶ Entre 9 am y 10 am en la mañana
  - ▶ Entre 3pm a 4pm en la tarde
  - ▶ Alrededor de las 7pm en la noche
- ▶ TCBH:Time Consistent Busy Hour
  - ▶ Es el período de 60 minutos consecutivos del día durante los cuales en media hay la más alta tasa de tráfico

# Hora Pico

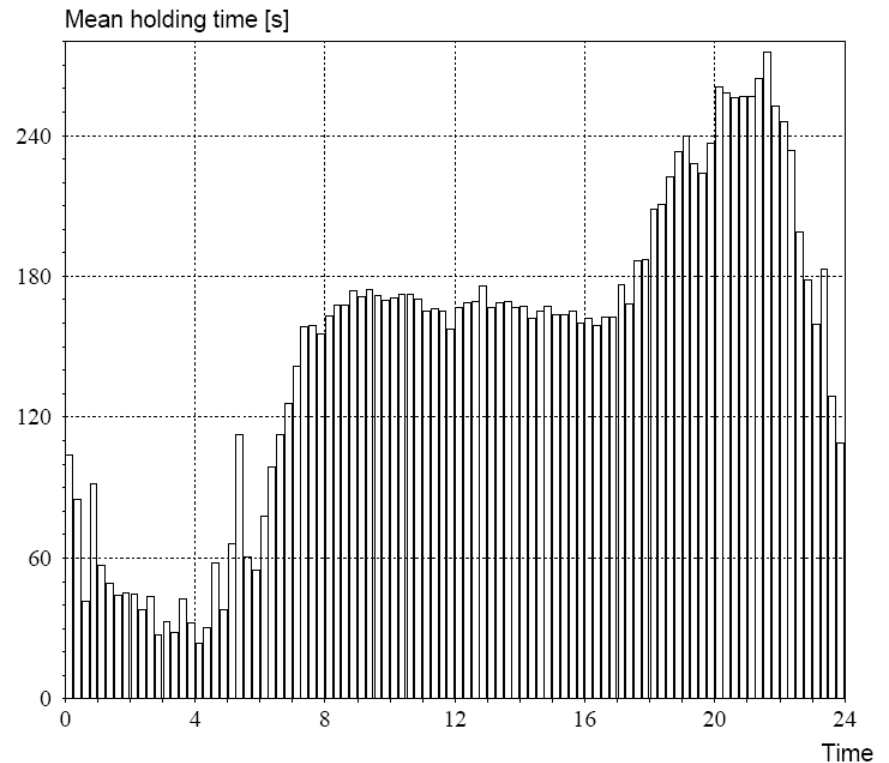
---

- ▶ Para sistemas telefónicos, aproximadamente el 5-8% realizan llamadas simultáneamente durante la hora pico (cada teléfono es usado el 10-16% del tiempo)
- ▶ Para llamadas internacionales, sólo el 1% de los usuarios realizan llamadas simultáneamente en la hora pico.



# Duración de las llamadas

- ▶ El tiempo de duración de las llamadas también varía dependiendo de la hora del día.
- ▶ En el ejemplo, las llamadas de la mañana son más cortas que las de la noche
- ▶ El tiempo de duración de las llamadas también cambia según si el servicio es telefonía fija o celular (las llamadas de celular son más cortas, por los costos)

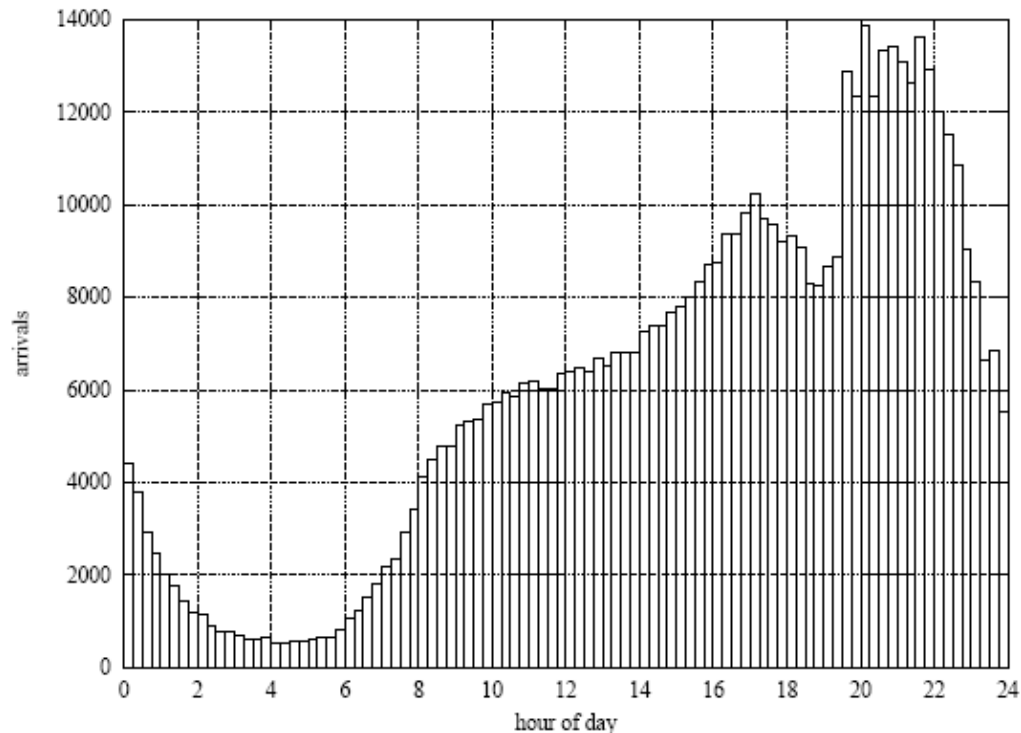


Fuente: Iversen (1973)



# Número de llamadas a un operador de Internet

- ▶ Las solicitudes de acceso fijo a Internet en el hogar se incrementan en las horas de la noche (cuando las personas están en casa)



Fuente: Iversen (2001)

Figure 2.6: Number of calls per 15 minutes to a modem pool of Tele Danmark Internet. Tuesday 1999.01.19.

# Congestión

---

- ▶ Las redes de telecomunicaciones poseen recursos limitados que realizan un multiplexado estadístico de las comunicaciones de los usuarios aprovechando que no todos llaman al mismo tiempo y que sus comunicaciones no duran lo mismo.
- ▶ La congestión sucede cuando el número de comunicaciones que se transmite sobre una red comienza a acercarse al límite de la capacidad de gestión de esta.
- ▶ Como consecuencia, las prestaciones de la red disminuyen drásticamente



# Tipos de Redes

---

▶ Según la forma en que reacciona la red ante una situación de congestión, estas se clasifican en:

## ▶ **Sistemas de pérdidas:**

▶ Una comunicación (llamada) que encuentra una situación de congestión abandona el sistema. A esta se le denomina una llamada perdida. **Ejemplo:** Redes telefónicas

## ▶ **Sistemas de Espera:**

▶ La comunicación (llamada, tarea o trabajo) se pone en una cola de espera y se atenderá cuando se libere algún recurso.

▶ La cola de espera puede ser finita o infinita.

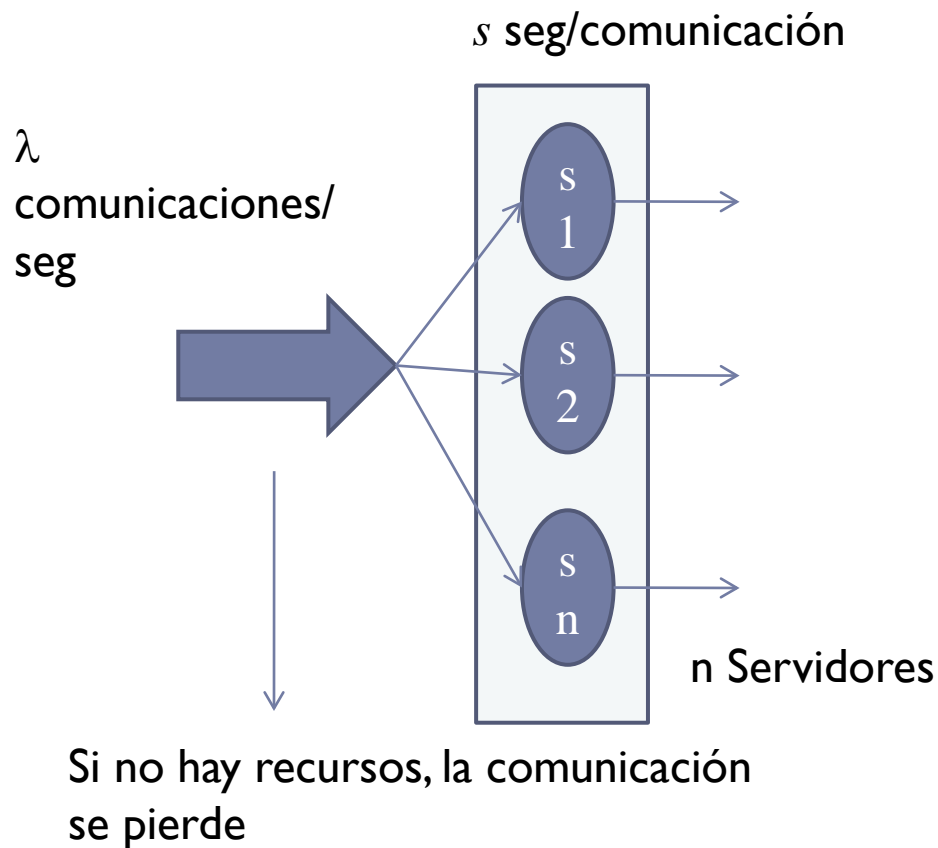
▶ **Ejemplo:** Redes de Paquetes

---

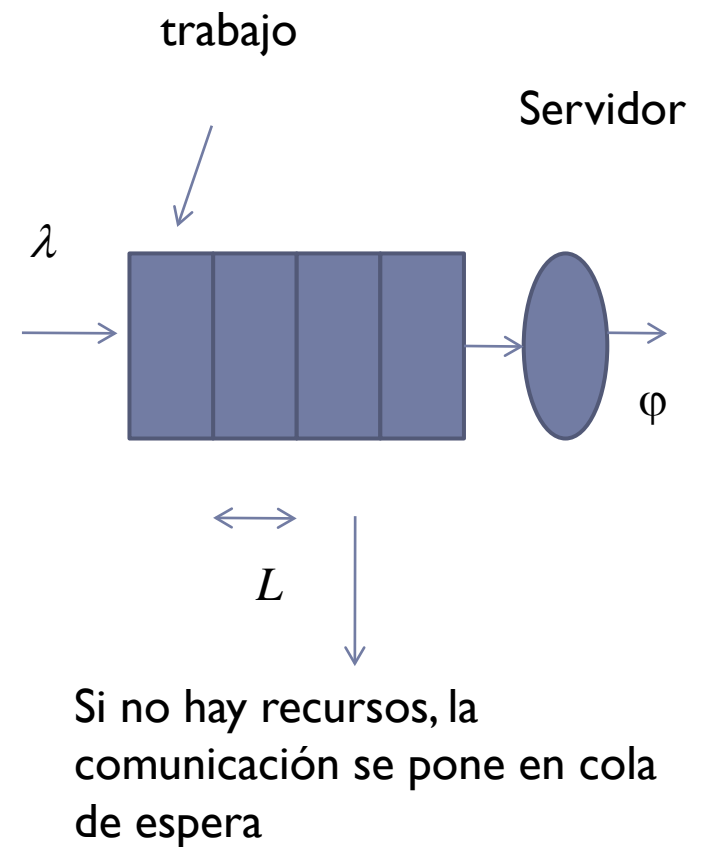


# Tipos de Redes

## Sistemas de Pérdidas



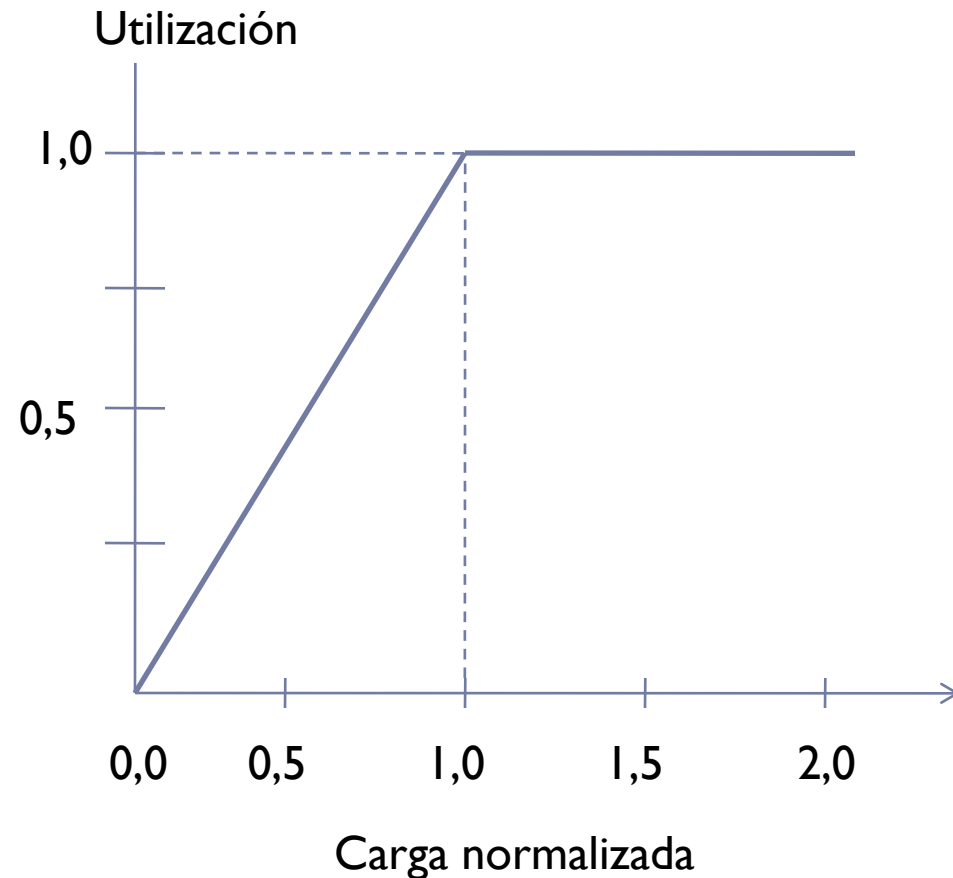
## Sistemas de Espera



# Congestión en sistemas de pérdidas

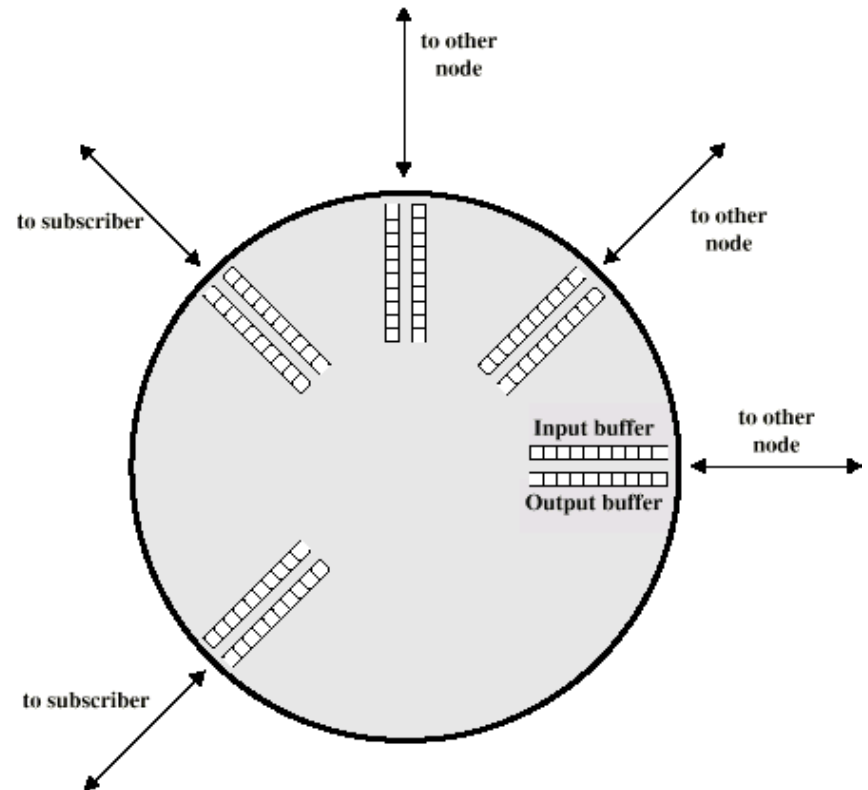
---

- ▶ **Grado de Utilización (Throughput):** Número de comunicaciones atendidas
- ▶ **Antes de la Congestión,** la red tiene un rendimiento lineal (toda comunicación es atendida)
- ▶ **Durante la congestión,** el total de servidores está trabajando al máximo y no puede atender más de su máxima capacidad



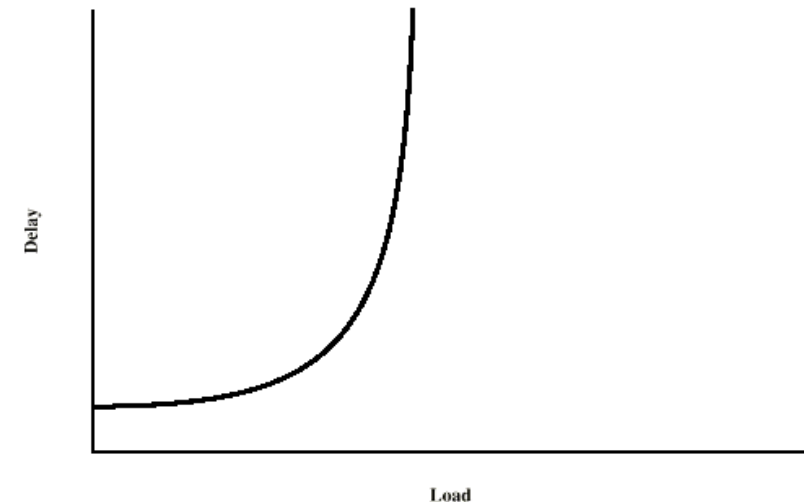
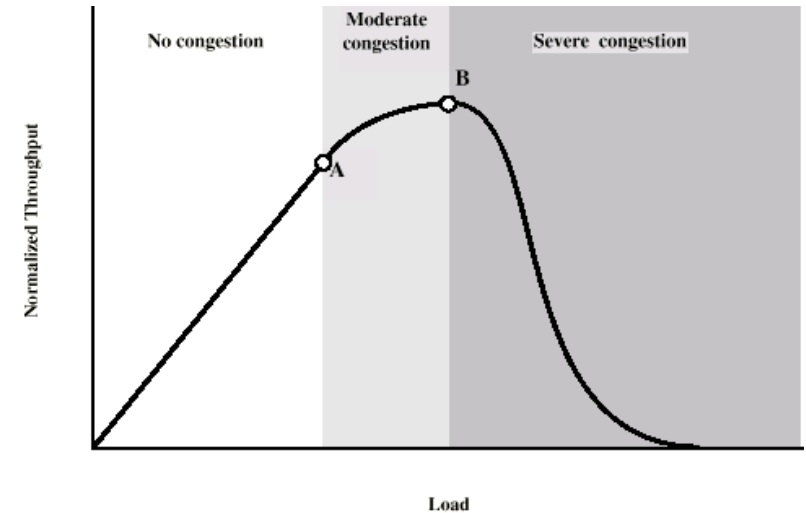
# Congestión en sistemas de espera

- ▶ Una red de datos es una red de colas
- ▶ En cada nodo existe una cola de paquetes asociada a cada enlace de salida
- ▶ Cada componente (Nodo) de la red tiene un límite de congestión
- ▶ En condiciones de baja carga, la congestión puede manejarse independientemente en cada nodo.



# Congestión en sistemas de espera

- ▶ El tamaño de las colas crece sin límite si la velocidad de llegada de trabajos es mayor que la velocidad de salida.
- ▶ En esta situación, el retardo de las colas tiende a infinito

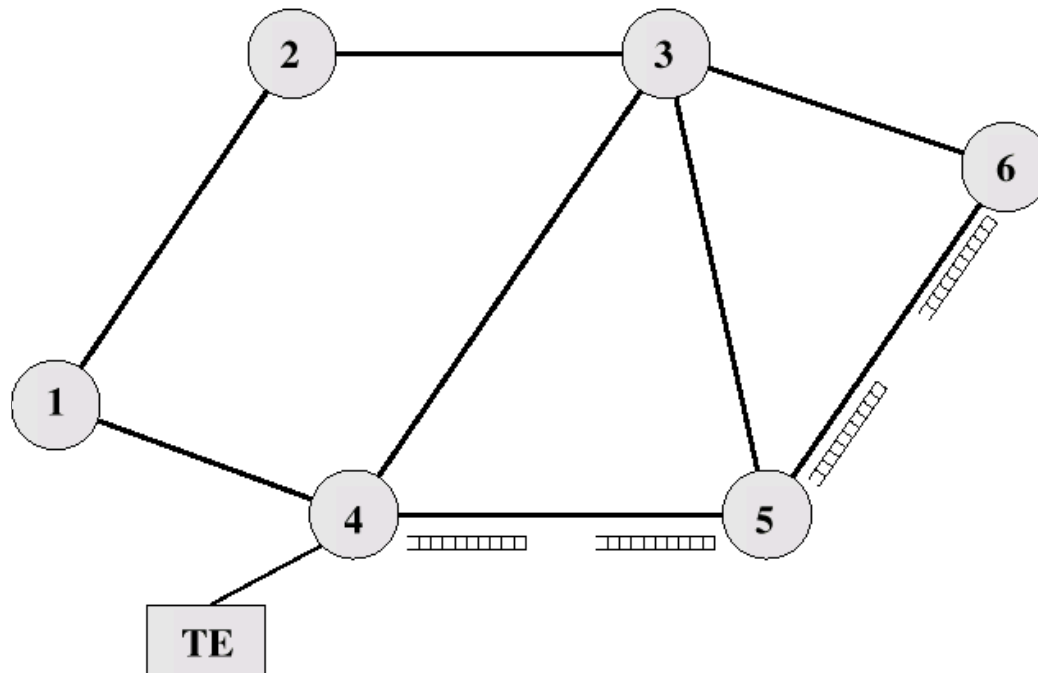


Fuente: Stallings, W. 2005

# Efectos de la congestión en la red

---

- ▶ En condiciones de congestión, la congestión de un enlace altera la congestión en los demás enlaces que conforman la ruta de la comunicación.
- ▶ La congestión en un punto se propaga a toda la red





# Grado de Servicio (GoS) y Calidad de Servicio (QoS)

# Calidad de Servicio (QoS)

---

- ▶ **Definida en la recomendación ITU-T E.800:**
  - ▶ Es el efecto colectivo del rendimiento del servicio, que determina el grado de satisfacción de un usuario del servicio
  - ▶ La QoS posee un conjunto de parámetros pertinentes al rendimiento de tráfico de la red
  - ▶ Adicionalmente, la QoS posee otros conceptos como:
    - ▶ Rendimiento del soporte del servicio
    - ▶ Rendimiento de la operatividad del servicio
    - ▶ Rendimiento de la servicialidad
    - ▶ Rendimiento de la seguridad del servicio
- ▶ **Una mejor QoS trae:**
  - ▶ Un mejor servicio a los usuarios y por tanto los atrae
  - ▶ Requiere de mayores costos de recursos y operación de los mismos
  - ▶ La decisión del nivel de QoS es influenciada también por aspectos políticos que no dependen del operador necesariamente.
- ▶ **La QoS se debe tener en cuenta durante la planeación de la red**



# Grado de Servicio (GoS)

---

- ▶ **Está definido en la recomendación ITU-T E.600:**
  - ▶ Es un número de variables de ingeniería de tráfico que proveen una medida de la adecuación de un grupo de recursos a unas condiciones específicas. Estas variables de GoS podrían ser probabilidad de pérdidas, retardo en el tono de marcación, etc.
  - ▶ Las variables de GoS son llamadas estándares de GoS
  - ▶ Los valores de los parámetros de GoS que se obtienen de la red son llamados Resultados de GoS



# Comparación GoS y QoS

---

## QoS

- ▶ La Calidad de Servicio ve la situación desde el punto de vista del cliente (usuario)

## GoS

- ▶ El Grado de Servicio ve la situación desde el punto de vista de la red

- Los valores que se dan a los parámetros de GoS, se asignan teniendo en cuenta que se alcancen los objetivos extremo a extremo de QoS.
- Los parámetros de GoS se refieren al rendimiento de la red o una porción de ella al ofrecer las comunicaciones entre usuarios.



# Parámetros principales del Grado de Servicio

---

- ▶ El grado de servicio se mide mediante una probabilidad porcentual:
  - ▶ Probabilidad de pérdida o bloqueo (sistemas de pérdidas)
  - ▶ Probabilidad de espera superior a un tiempo determinado (sistemas de espera)



# Probabilidad de bloqueo

---

- ▶ Se produce en los sistemas de pérdidas
- ▶ Es la probabilidad de que un usuario intente hacer una llamada y no encuentre recursos.
- ▶ La red reacciona enviando un tono de congestión o un aviso de red ocupada.
- ▶ El usuario debe intentar una nueva llamada posteriormente.



# Probabilidad de espera

---

- ▶ Se utiliza en los sistemas de espera
- ▶ Es la probabilidad de que un trabajo tenga que esperar más de un tiempo de referencia tomado como el tiempo máximo tolerable.
- ▶ Si se supera el tiempo máximo tolerable, en algunas redes se realiza una desconexión forzada de la comunicación.
- ▶ Otros parámetros importantes en sistemas de espera son:
  - ▶ El tiempo medio de espera (delay): La tolerancia de los usuarios no necesariamente es lineal con el tiempo de espera
  - ▶ La variación en el tiempo medio de espera (delay Jitter): Servicios de tiempo real como el vídeo y la voz no permiten variaciones grandes en el tiempo de espera de los paquetes. Esto afecta la *Calidad de la Experiencia (QoE)* del usuario.



# Ejemplo

---

- ▶ Suponga que se quiere tener una Probabilidad de bloqueo extremo a extremo del 1% en una red telefónica
- ▶ Un usuario interpreta esto como que si hiciera 100 intentos de llamada, en media en 99 de ellos tendría éxito.
- ▶ Fijando este objetivo de diseño (**objetivo de QoS**), el operador asigna proporcionalmente una probabilidad de bloqueo (**parámetros de GoS**) para cada uno de los elementos de la red.
- ▶ La red debe ser monitoreada para asegurar que se cumple esta meta
- ▶ El monitoreo sólo asegura que la red en media alcanza el objetivo planteado:
  - ▶ Una sólo línea telefónica podría obtener un GoS mayor, pero la media de todas las redes alcanzaría el objetivo.





# Acuerdo de nivel de Servicio (SLA)

---

- ▶ SLA: Service Level Agreement
- ▶ Es un contrato entre un usuario y el operador de la red.
- ▶ Se definen los valores de los parámetros que tienen un significado importante.
- ▶ Se escribe de manera que sea igualmente entendible tanto por el usuario como por el operador de red.
- ▶ Es útil para poner en práctica unos términos de acuerdo sin importar las sutiles diferencias entre QoS y GoS.
- ▶ El SLA también define qué hacer en caso de que los términos del contrato sean violados.
- ▶ Algunos operadores tienen SLA para todos los usuarios. Otros sólo lo hacen para grandes clientes que saben lo que realmente significan los términos del SLA.



# Monitoreo del rendimiento de la red

---

- ▶ El rendimiento de la red se refiere a la capacidad de la red o una porción de ella de ofrecer las comunicaciones entre usuarios.
- ▶ Se requiere determinar cómo está operando la red, por lo que es necesario hacer mediciones de tráfico.
- ▶ Las mediciones cubren diferentes aspectos del rendimiento de la red:
  - ▶ Capacidad de tráfico
  - ▶ Dependencia de las diferentes partes
  - ▶ Rendimiento en la transmisión
  - ▶ Facturación
- ▶ Al crecer la red, se requiere tener una organización para estar observando la operación y realizar el mantenimiento de la red (Dpto. de operación y mantenimiento: realiza **tareas de gestión de red**).
- ▶ También se requieren herramientas de automatización para dichas tareas.



# Funciones de la Gestión de red

- ▶ OAM&P: Operation, Administration, Maintenance and Provisioning
- ▶ Estas funciones se requieren para Proveer, Monitorear, Interpretar y Controlar la red y los servicios que ella ofrece.

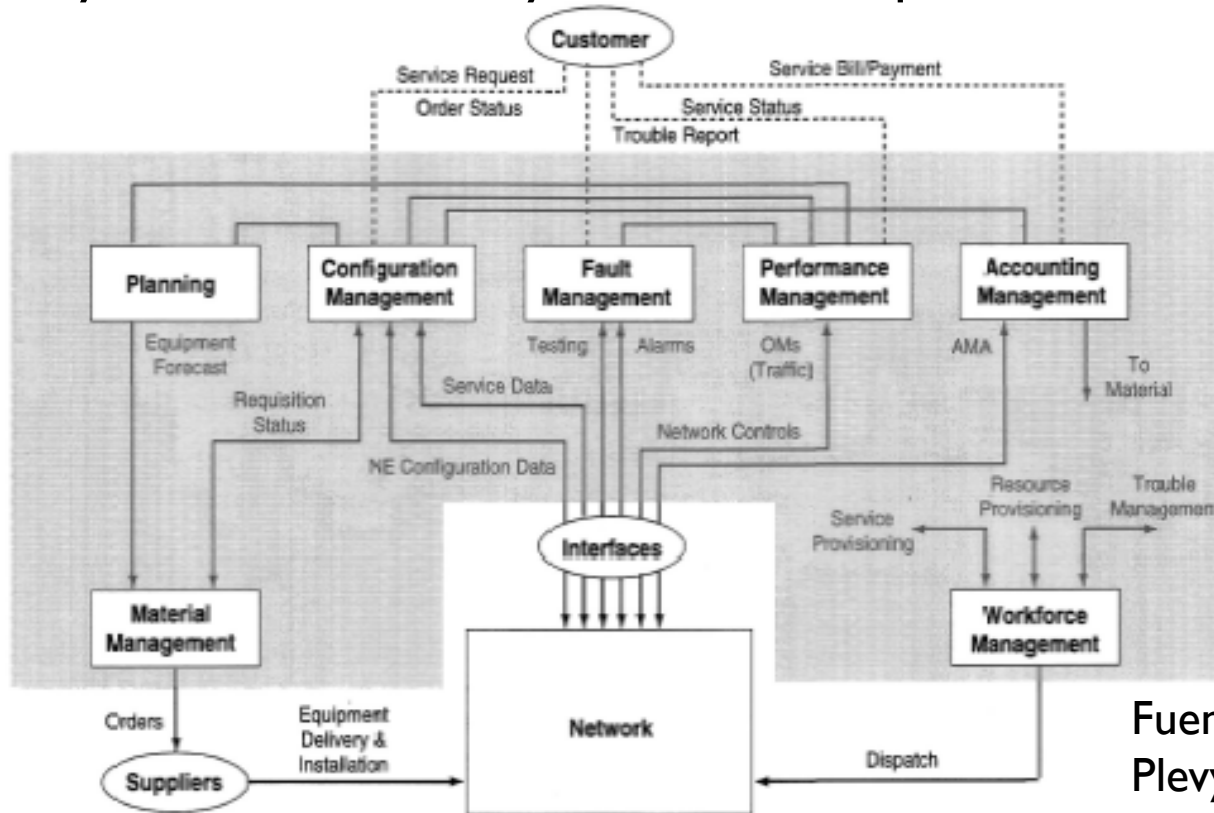


Figure 1-5 Network Management Functions

Fuente: S. Aidarous, T. Plevyak. "Principles of Network Management".

# Configuraciones de referencia

---

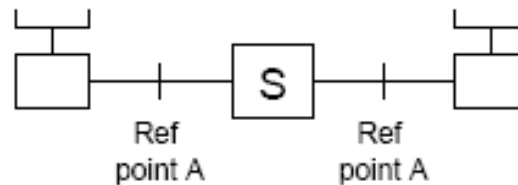
- ▶ Son bosquejos de los caminos que pueden tomar las comunicaciones dentro de la red, junto con unos puntos de referencia que permiten determinar las características del tráfico en enlaces importantes.
- ▶ Los puntos de referencia se ubican estratégicamente en sitios tales como las interfaces entre operadores, puntos de concentración de tráfico, etc.
- ▶ Los puntos de referencia permiten hacer la distribución de los valores de los parámetros de GoS en diferentes porciones de la red.



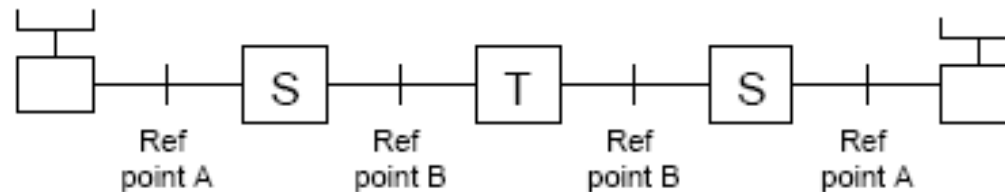
# Configuraciones de referencia típicas en una red telefónica

---

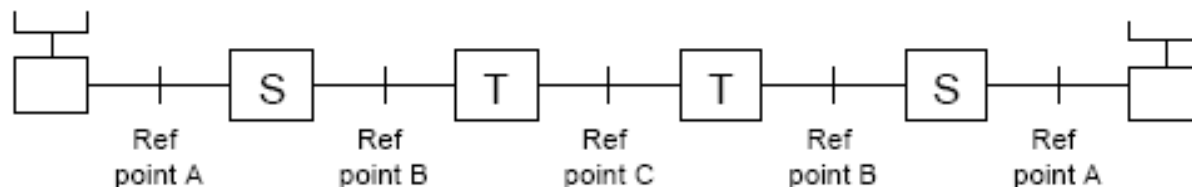
terminal → subscriber switch → terminal



terminal → subscriber switch → transit switch → subscriber switch → terminal



terminal → subscriber switch → transit switch → transit switch → subscriber switch → terminal



# Conexiones

---

## ▶ **Conexión:**

- ▶ Asociación de recursos que proveen medios para la comunicación entre dos o más dispositivos conectados a una red de comunicaciones.
- ▶ Los recursos asociados a una conexión pueden variar en número y tipo.
- ▶ Se requieren unas conexiones de referencia que muestren casos típicos dentro de la red.

## ▶ **Tipos de conexiones según los dominios de red que interconectan:**

- ▶ **Conexión Local**
  - ▶ **Conexión Nacional**
  - ▶ **Conexión Internacional**
- 



# Conexiones en redes de paquetes

---

- ▶ Los paquetes atraviesan diferentes dominios de red llamados “nubes”.
- ▶ Nube:
  - ▶ Conjunto de routers y enlaces que pertenecen a la jurisdicción de un mismo operador de red (ej: un mismo ISP)
  - ▶ Las nubes se interconectan entre sí a través de Gateways
- ▶ Es importante determinar el comportamiento del tráfico en las conexiones entre dos nubes
- ▶ El camino entre dos usuarios puede estar compuesto por varias nubes



# Recomendaciones ITU-T para redes IP

---

Recomendación	Objetivo
I.380/Y.1540	Define parámetros de rendimiento para la transferencia de paquetes IP
Y.1541	Asignación de valores y objetivos de rendimiento
E.651	Conexiones de referencia para redes de acceso IP





# Comportamiento de los usuarios

# Por qué estudiarlo?

---

- ▶ El comportamiento de los usuarios es de vital importancia para la ingeniería de tráfico
- ▶ Dependiendo de las horas de más tráfico en la red, de la cantidad de tráfico generada por los usuarios, del comportamiento estadístico del tráfico generado por los usuarios, se podrán obtener modelos matemáticos más adecuados, permitiendo una mejor planificación de la red.



# Comportamiento de los usuarios en redes telefónicas

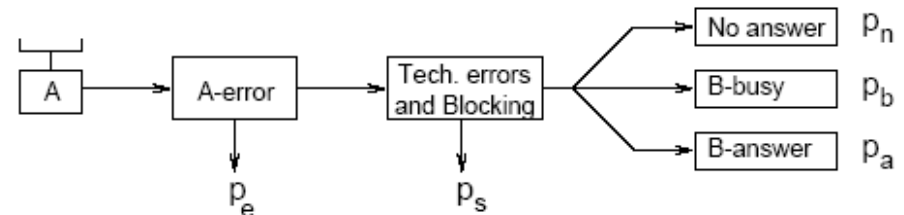
- ▶ Cuando un usuario A intenta hacer una llamada a un usuario B, el resultado tendrá dos posibilidades:
  - ▶ *Llamada exitosa*: La red acepta la llamada.
  - ▶ *Llamada fallida*: La llamada es rechazada por la red. Esto se debe a que la red está ocupada (no tiene recursos).
- ▶ Cuando un usuario hace un intento de llamada fallida, lo normal es que vuelva a intentar una nueva llamada ó incluso puede realizar varios intentos de llamada fallida.

Resultado durante la hora Pico	País Industrializado	País en desarrollo
Error de A	15%	20%
Bloqueo y errores técnicos (Red)	5%	35%
B no responde antes de que A cuelgue	10%	5%
B ocupado	10%	20%
<b>B contesta= conversación</b>	<b>60%</b>	<b>20%</b>
<b>No conversación</b>	<b>40%</b>	<b>80%</b>

Fuente: Datos de la ITU-T.

# Cálculo de probabilidades

- ▶ Se pueden calcular las probabilidades suponiendo un una secuencia de eventos como la de la figura.
- ▶ Se supone también independendia entre los procesos.
- ▶ Tomando los valores de la tabla anterior se despejan los valores de probabilidad para cada resultado posible.



$$p\{\text{A-error}\} = p_e$$

$$p\{\text{Congestion \& tech. errors}\} = (1 - p_e) \cdot p_s$$

$$p\{\text{B-no answer}\} = (1 - p_e) \cdot (1 - p_s) \cdot p_n$$

$$p\{\text{B-busy}\} = (1 - p_e) \cdot (1 - p_s) \cdot p_b$$

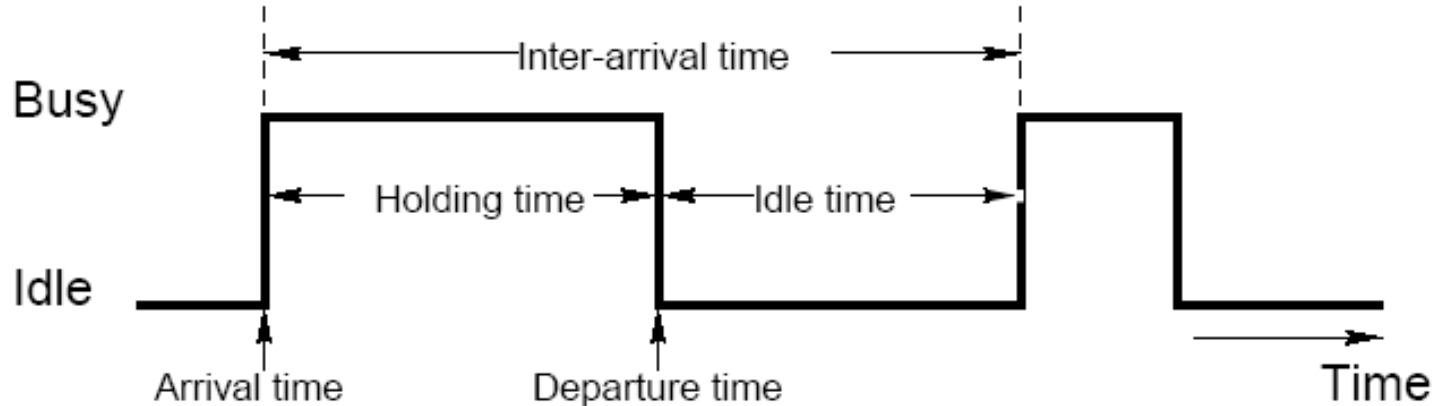
$$p\{\text{B-answer}\} = (1 - p_e) \cdot (1 - p_s) \cdot p_a$$

I - country			D - country		
$p_e$	$= \frac{15}{100}$	$= 15\%$	$p_e$	$= \frac{20}{100}$	$= 20\%$
$p_s$	$= \frac{5}{85}$	$= 6\%$	$p_s$	$= \frac{35}{80}$	$= 44\%$
$p_n$	$= \frac{10}{80}$	$= 13\%$	$p_n$	$= \frac{5}{45}$	$= 11\%$
$p_b$	$= \frac{10}{80}$	$= 13\%$	$p_b$	$= \frac{20}{45}$	$= 44\%$
$p_a$	$= \frac{60}{80}$	$= 75\%$	$p_a$	$= \frac{20}{45}$	$= 44\%$



# Procesos del tráfico en una red

---



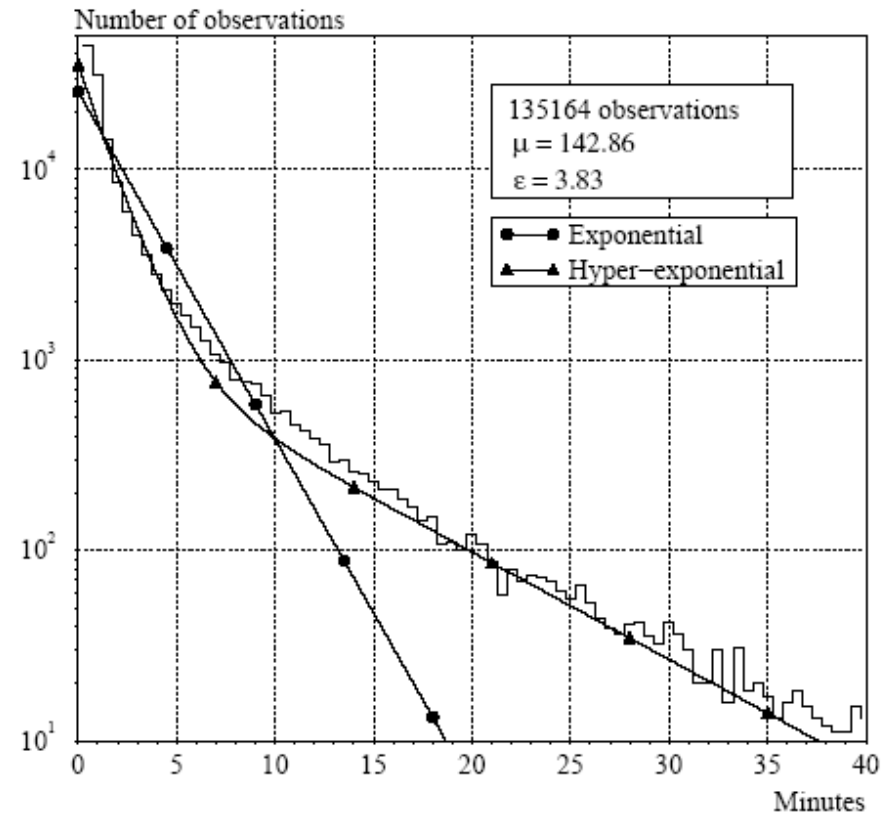
**-Proceso de llegadas:** Es la forma en que llegan las llamadas, paquetes, flujos, etc., a una red. Estas llegadas son producidas por los usuarios de la red y son aleatorias. Parámetro principal: Tiempo entre llegadas (Inter-Arrival Time).

**-Proceso de Servicio:** Se refiere a la forma en que el sistema da servicio (atiende) a las llamadas (duración de la llamada) o paquetes (tiempo que tarda el sistema en tomar decisiones y re-enviarlo). Parámetro principal: Tiempo de Servicio (Holding Time).

**-Tiempo de servicio de una llamada**= tiempo de establecimiento + tiempo conversación + tiempo finalización

# Distribución estadística de los procesos

- ▶ Es importante conocer las distribuciones de los procesos de llegadas y de servicio
- ▶ En la figura se observa la distribución del proceso de servicio de las troncales para una central telefónica local.
- ▶ La distribución estadística permite hacer cálculos de probabilidades y tomar decisiones de diseño más acertadas.



# Ejemplo: Tiempo medio de servicio

---

- ▶ Supongamos el tiempo medio de servicio de las llamadas interrumpidas antes de que B conteste como 20 seg.
- ▶ Supongamos el tiempo medio de servicio de las llamadas que llegan a B (no contestadas, B-ocupado, Contestadas por B) como 180 seg.
- ▶ Por tanto, el tiempo medio de servicio para el suscriptor A es:

$$\text{I - country:} \quad m_a = \frac{20}{100} \cdot 20 + \frac{80}{100} \cdot 180 = 148 \text{ seconds}$$

$$\text{D - country:} \quad m_a = \frac{55}{100} \cdot 20 + \frac{45}{100} \cdot 180 = 92 \text{ seconds}$$

- ▶ Estos cálculos salen de la tabla de probabilidades deducida para los datos de la ITU-T.
- 



# Concluyendo....

---

- ▶ Si conocemos los tiempos medios de servicio para cada fase de un intento de llamada, podríamos calcular:
  - ▶ La proporción de llamadas que se pierden durante las fases individuales. Esto se usa para dimensionamiento de los diferentes módulos de control de las llamadas.
- ▶ En una red telefónica:
  - ▶ La carga de tráfico de cada intento de llamada en los equipos de control de la llamada es casi constante.
  - ▶ La carga de tráfico de los equipos que dan servicio a la voz es proporcional a la duración de las llamadas.
- ▶ Por tanto, podría haber congestión en los equipos de control de llamadas (los usuarios insisten mucho), mientras que la red puede tener capacidad disponible para la voz.





# Influencia del comportamiento de los usuarios en la carga de la red

---

Attempt no.	Number of observations				Persistence
	Success	Continue	Give up	$p\{\text{success}\}$	
		75.389			
1	56.935	7.512	10.942	0.76	0.41
2	3.252	2.378	1.882	0.43	0.56
3	925	951	502	0.39	0.66
4	293	476	182	0.31	0.72
5	139	248	89	0.29	0.74
> 5	134		114		
Total	61.678		13.711		

- ▶ La tabla representa una secuencia observada de intentos de llamada repetidos.
  - ▶ La probabilidad de éxito decrece con el aumento en el número de intentos de llamadas, mientras que la persistencia (prob. de re-intento de llamada) se incrementa.
  - ▶ En este caso se considera un intento de llamada repetido como un intento de llamada hacia el mismo usuario dentro de la hora siguiente.
  - ▶ **IMPORTANTE: El comportamiento del usuario es un área de gran interés en la ingeniería de tráfico**
- 

