Teoría e Ingeniería de Teletráfico

Introducción

TEORÍA E INGENIERÍA DE TELETRÁFICO

¿Qué es la “Teoría de Teletráfico”?

Es la aplicación de las teorías de probabilidades a la solución de problemas de planificación, evaluación de desempeño, operación y mantenimiento de sistemas de telecomunicaciones

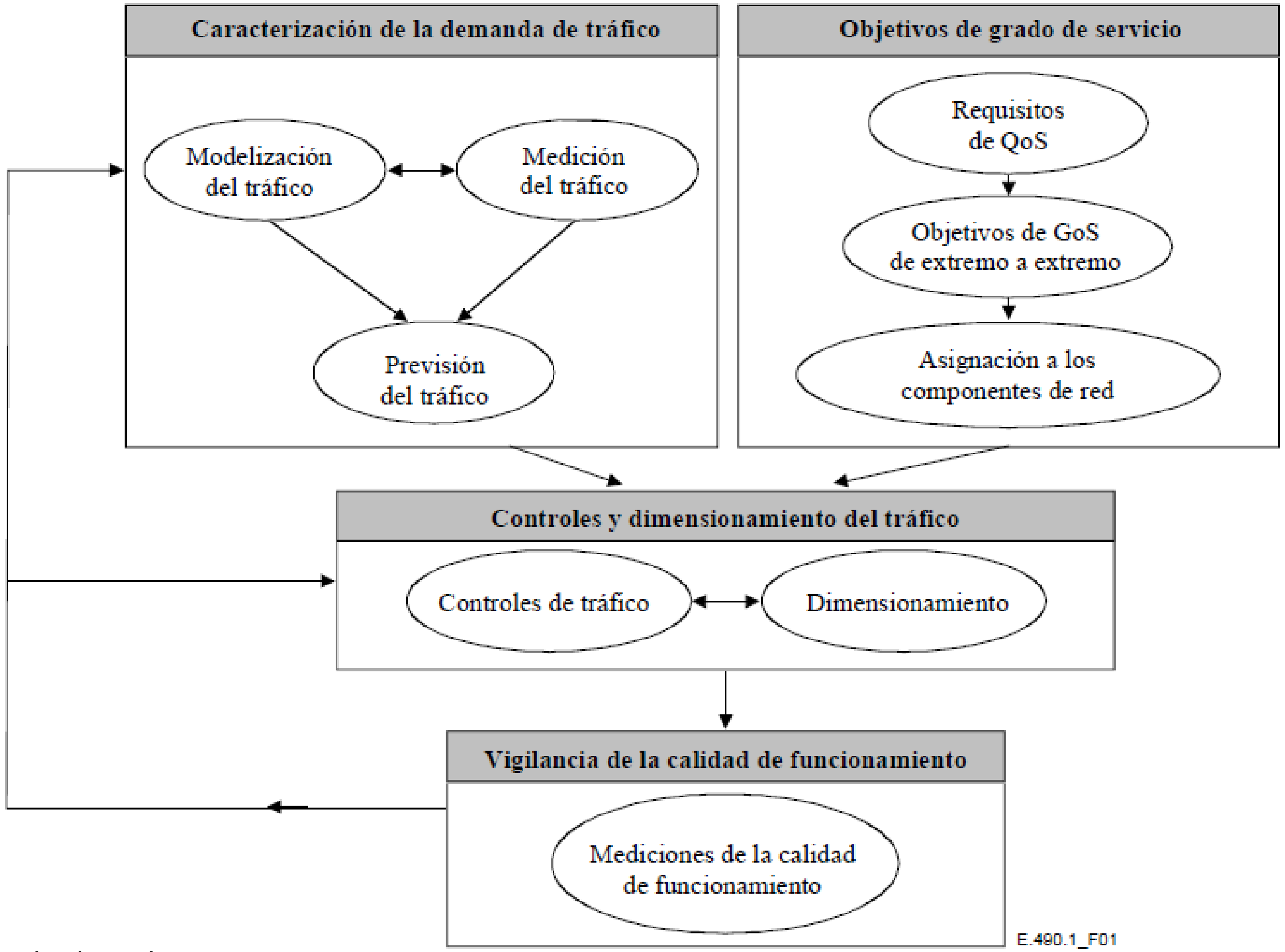
Principales funciones de la Ingeniería de Teletráfico:

- Caracterización de la demanda de tráfico
- Objetivos del grado de servicio (GoS)
- Controles y dimensionamiento del tráfico
- Vigilancia de la calidad de funcionamiento

Según Recomendación ITU-T E.490.1 (01/2003)

Parte de esta presentación se basa en
ITU–D Study Group 2 Question 16/2
Handbook “TELETRAFFIC ENGINEERING”
June 2006

Principales funciones de la Ingeniería de Teletráfico



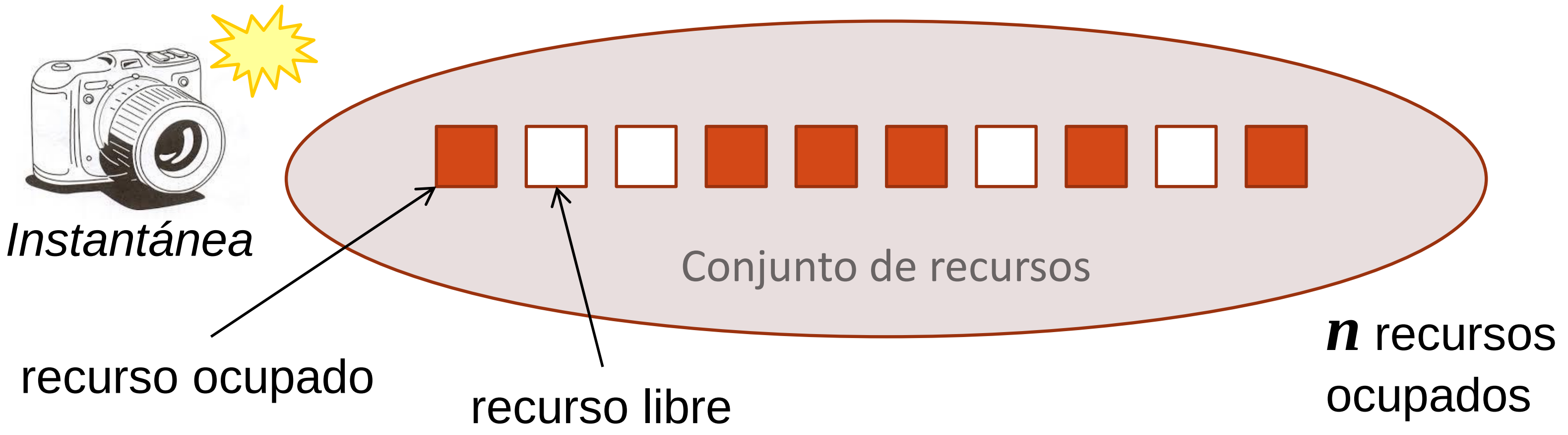
Según Recomendación ITU-T E.490.1 (01/2003)

Intensidad Instantánea de Tráfico

Traffic Intensity

La intensidad *instantánea* de tráfico en un conjunto de recursos es la cantidad de recursos ocupados en un determinado instante de tiempo

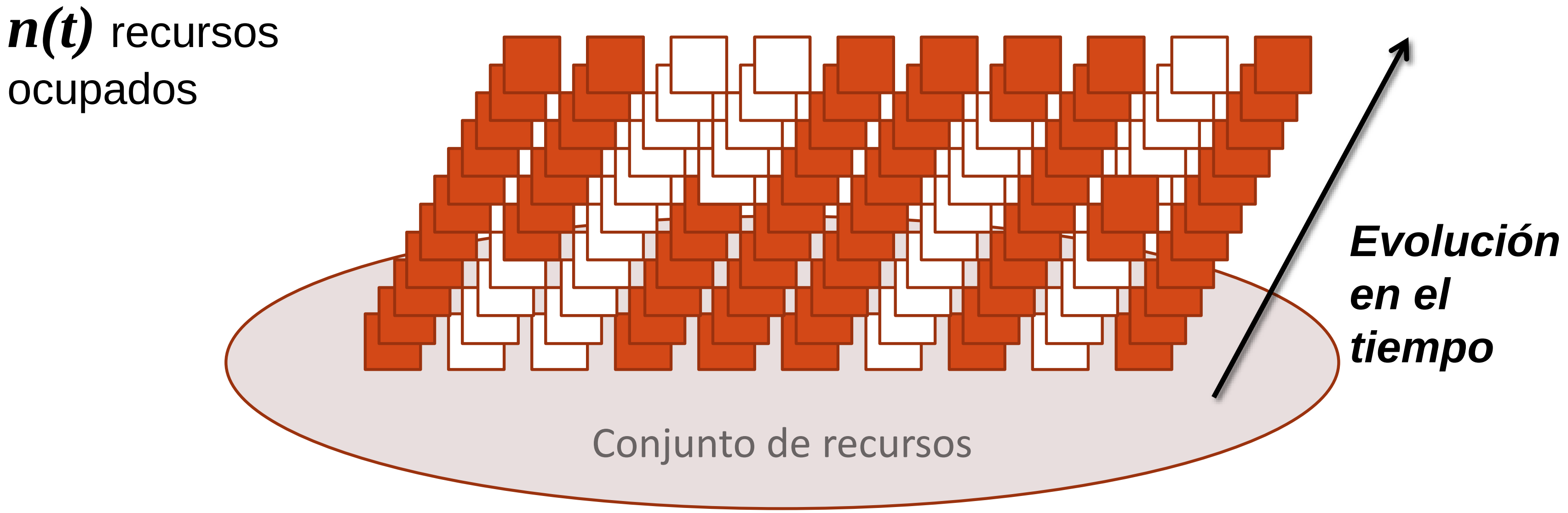
- Los "recursos" pueden ser líneas urbanas, servidores, o cualquier tipo de elemento que puede ser compartido por varios usuarios



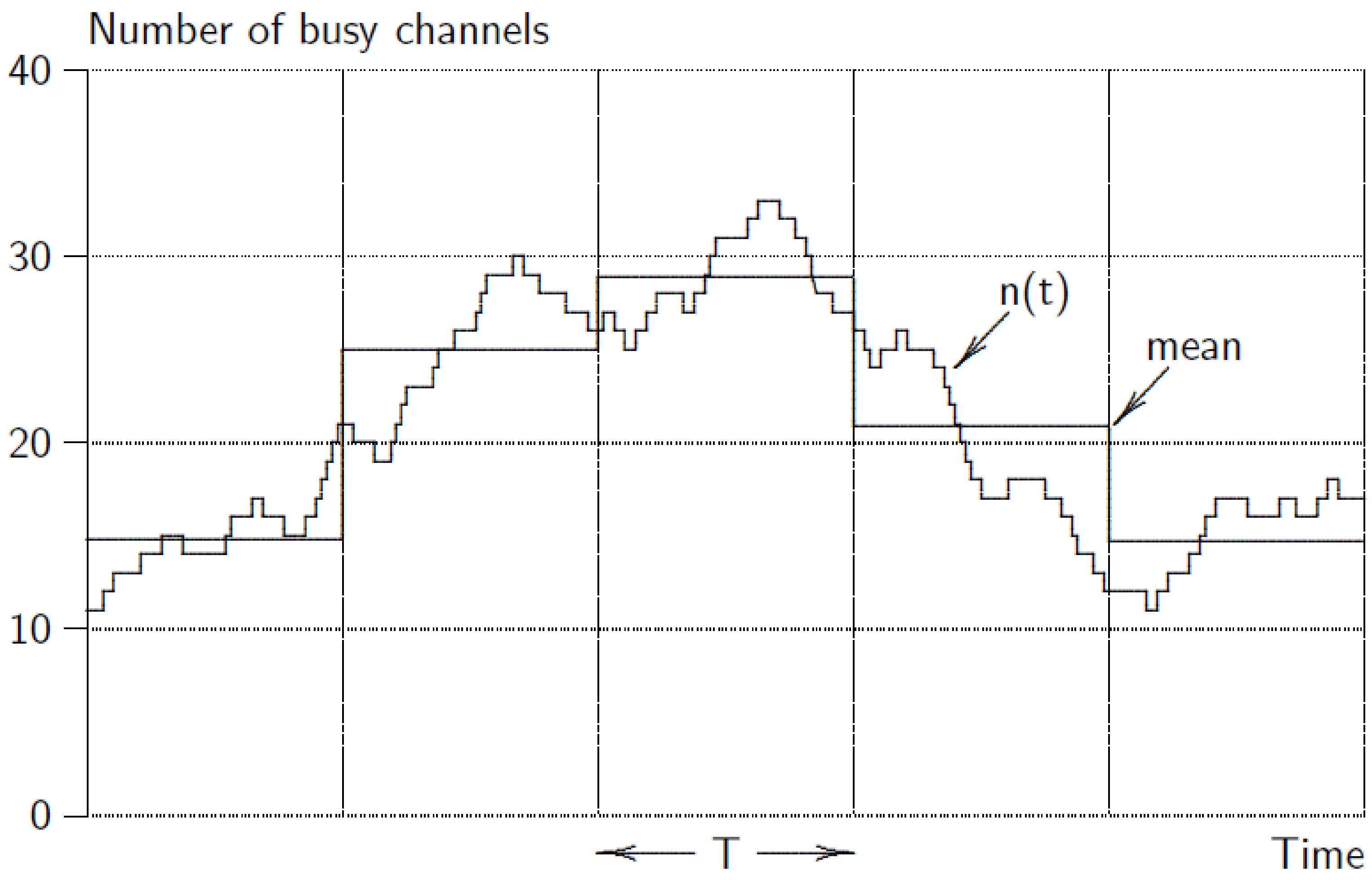
Intensidad Promedio de Tráfico

La ocupación de cada recurso puede variar con el tiempo.

- En cada instante t, hay n recursos ocupados



Intensidad Promedio de Tráfico



Intensidad Promedio de Tráfico

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

Donde ***n(t)*** es la cantidad de recursos ocupados en cada instante ***t*** y ***T*** es un tiempo fijo

Y(T) es adimensionada (“tiempo” / “tiempo”)

Erlang

Unidad de medida de tráfico (definición de la CCIF* del 28 de octubre de 1946):

"For a group of circuits (or connecting devices), the average intensity of traffic during a period T equals the total occupancy divided by T.
"The unit of traffic intensity as defined above is called "*erlang*".

A. K. Erlang
(1878–1929)

* Le Comité Consultatif International des Comunications Telephoniques a grande distance

Unidades de Tráfico

Erlang (***E***):

- Un Erlang corresponde a una intensidad promedio de tráfico de una hora por hora
- Equivalente a un recurso ocupado en forma permanente

Cientos de segundos por hora o *Hundred call seconds per hour* (***CCS***):

- Un CCS corresponde a una intensidad promedio de tráfico de 100 segundos por hora (puede ser asociado a la duración de una llamada típica)
- 36 CCS = 1 E

Volumen de tráfico (*Traffic Volume*)

Es el tráfico total cursado en un periodo de tiempo *T*

- Integral en el tiempo de la intensidad de tráfico

Se mide en *Eh* (Erlang-horas) o *Es* (Erlang-segundos)

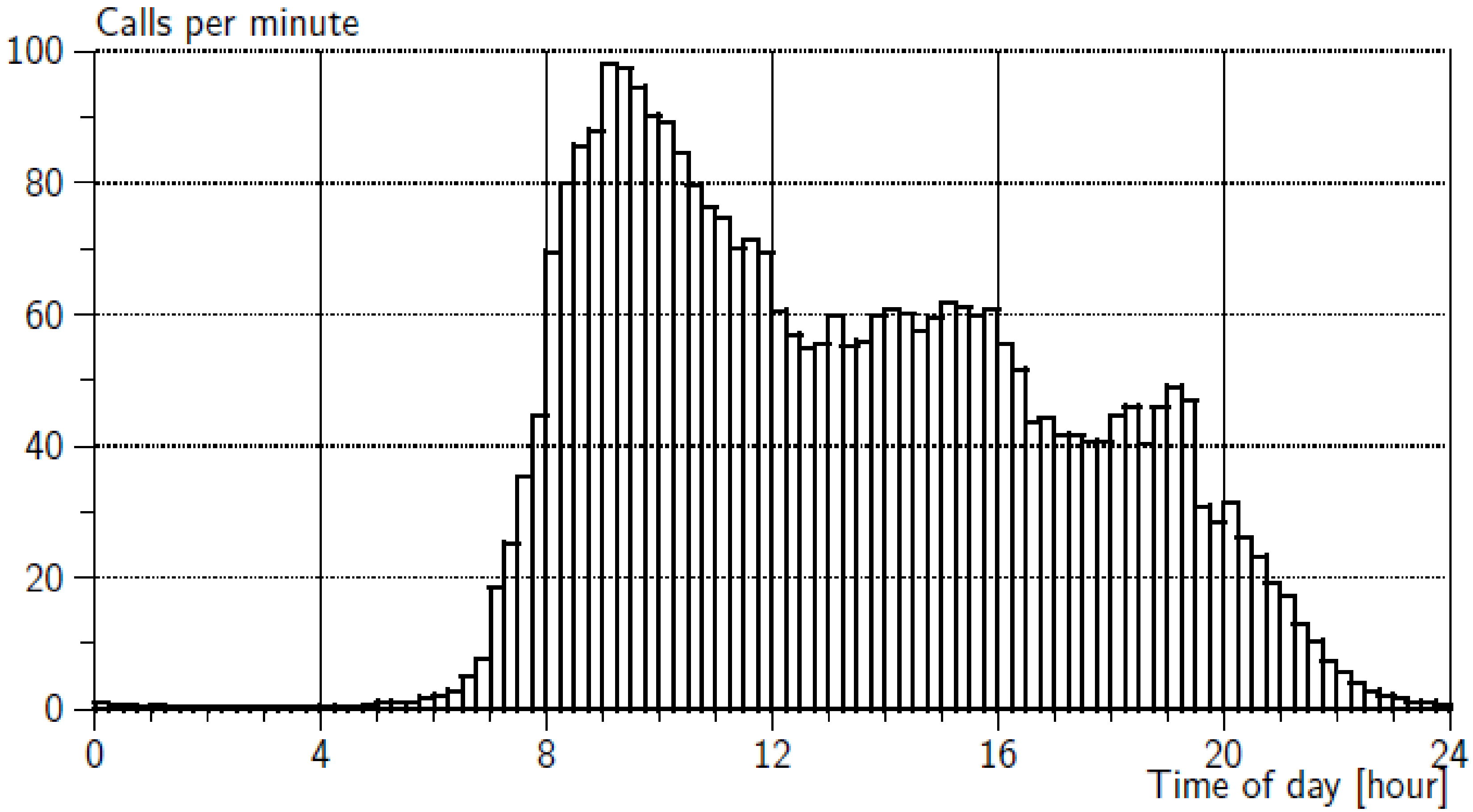
Es la suma de todos los tiempos de ocupación en el periodo medido

Intensidad de llamadas (*Call Intensity*)

Promedio de llamadas por unidad de tiempo

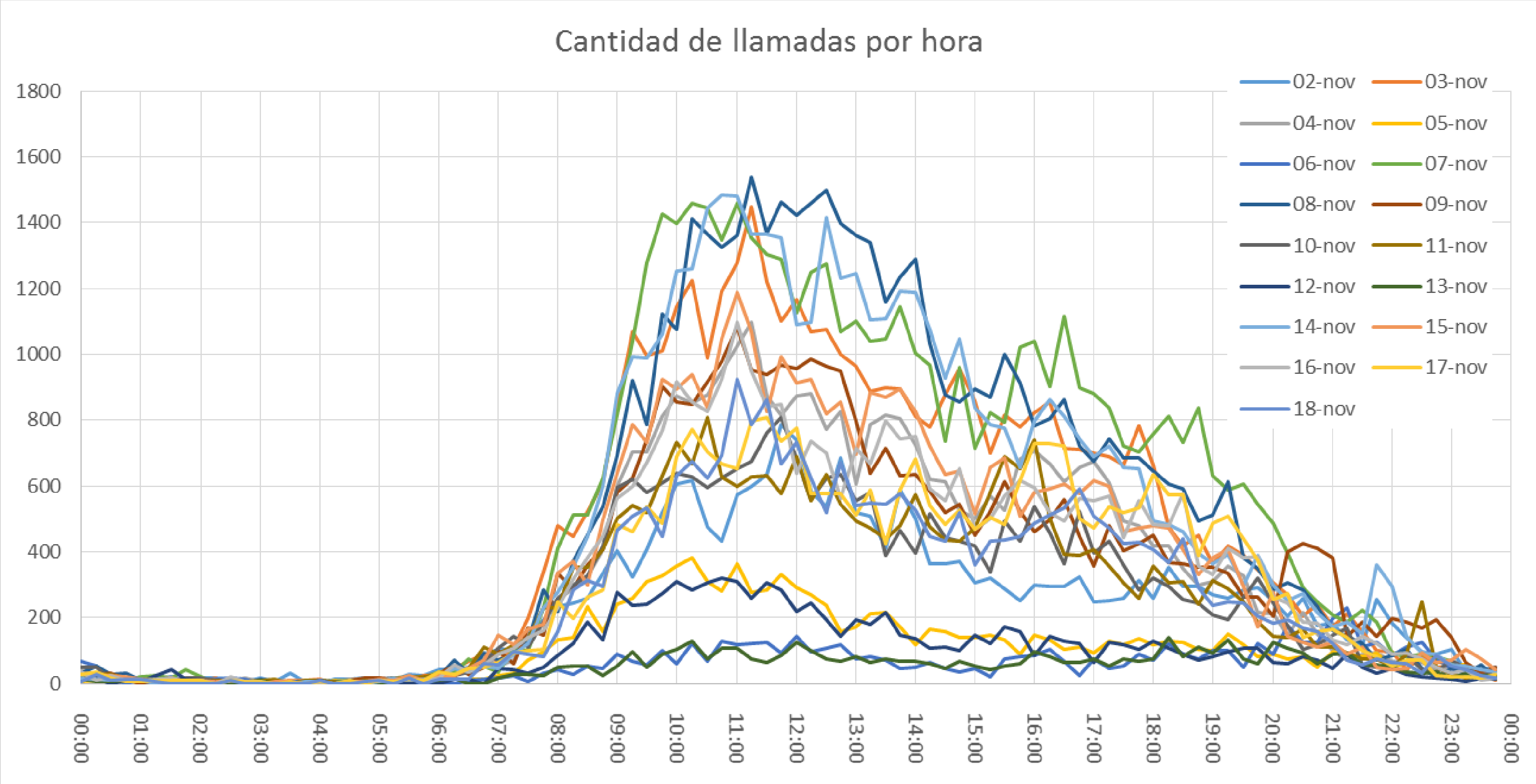
$$\lambda = \frac{llamadas}{unidad\ \ de\ \ tiempo}$$

Intensidad de llamadas (*Call Intensity)*



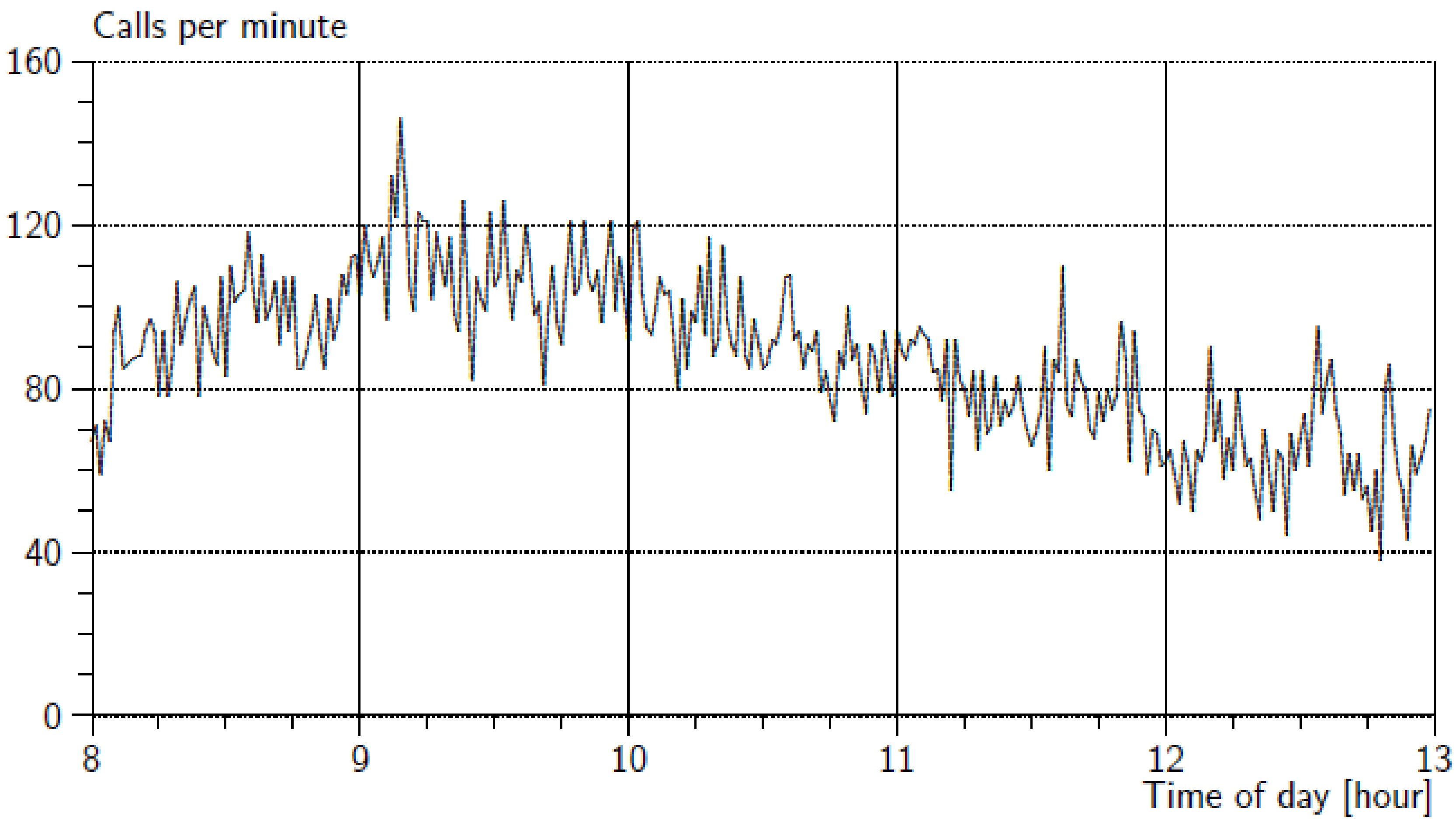
Promedio de llamadas por minuto, tomadas cada 15 minutos,
durante 10 días, de lunes a viernes

Intensidad de llamadas (*Call Intensity)*



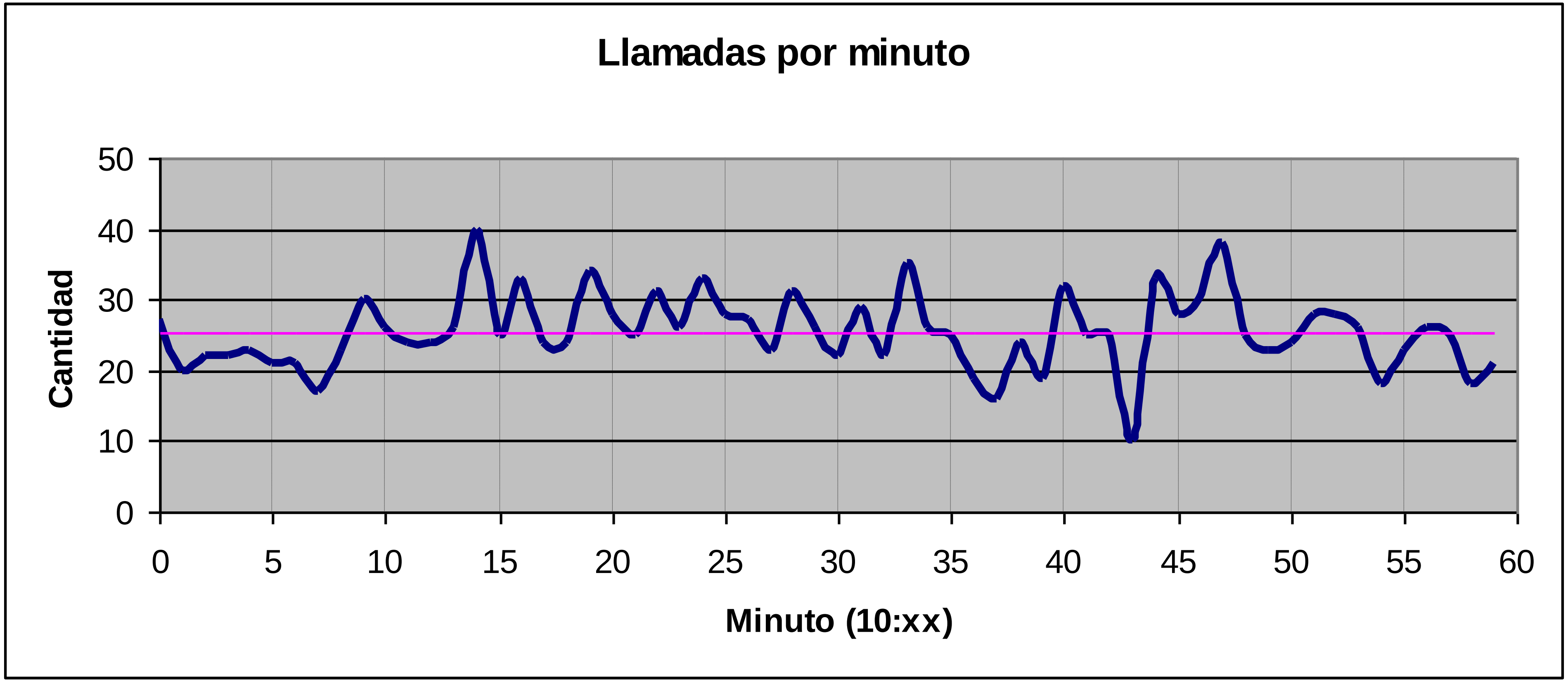
Promedio de llamadas por minuto, tomadas cada 15 minutos, cada día

Intensidad de llamadas (*Call Intensity)*



Llamadas por minuto, entre las 8:00 y las 13:00 de un día particular

Intensidad de llamadas (*Call Intensity*)

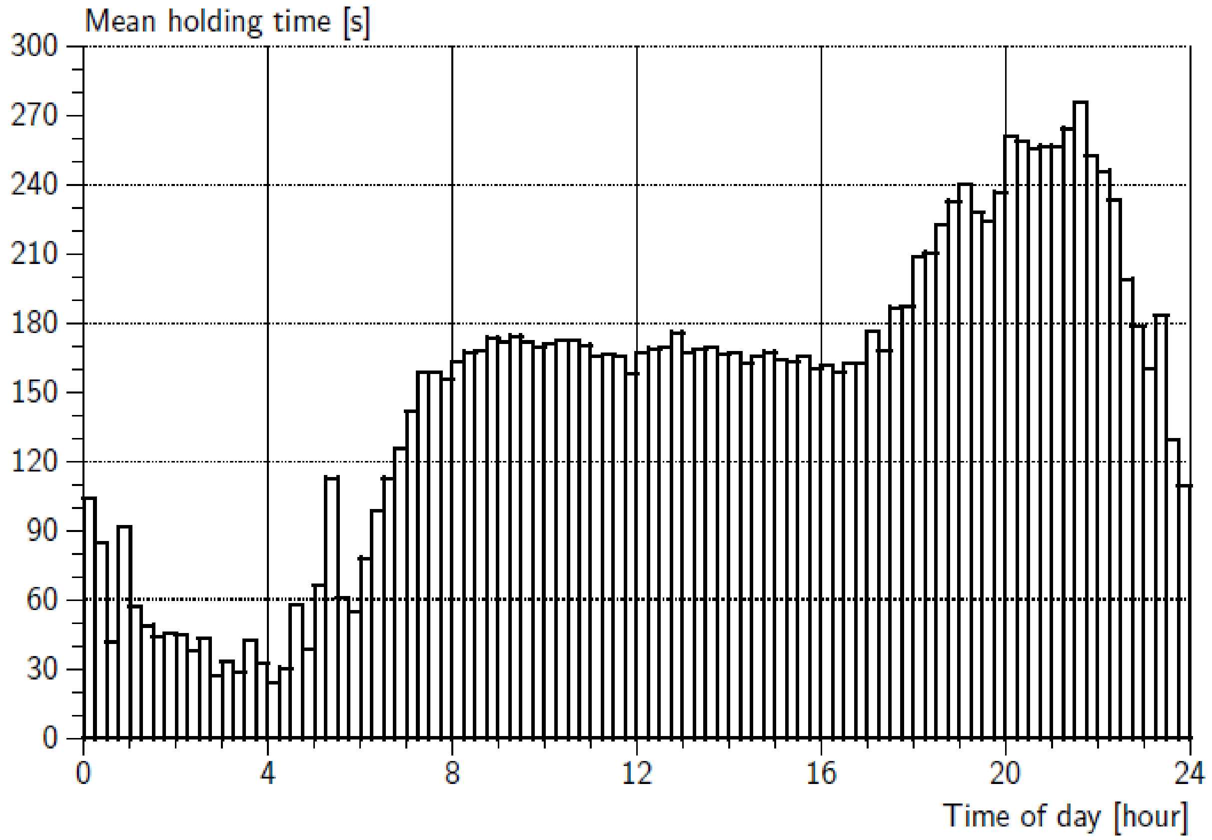


Llamadas por minuto, en una hora (entre las 10 y las 11), en un día particular

Tiempo medio de ocupación (*Mean Holding Time*)

Duración media del tiempo de ocupación

- Por ejemplo: Duración media de las llamadas

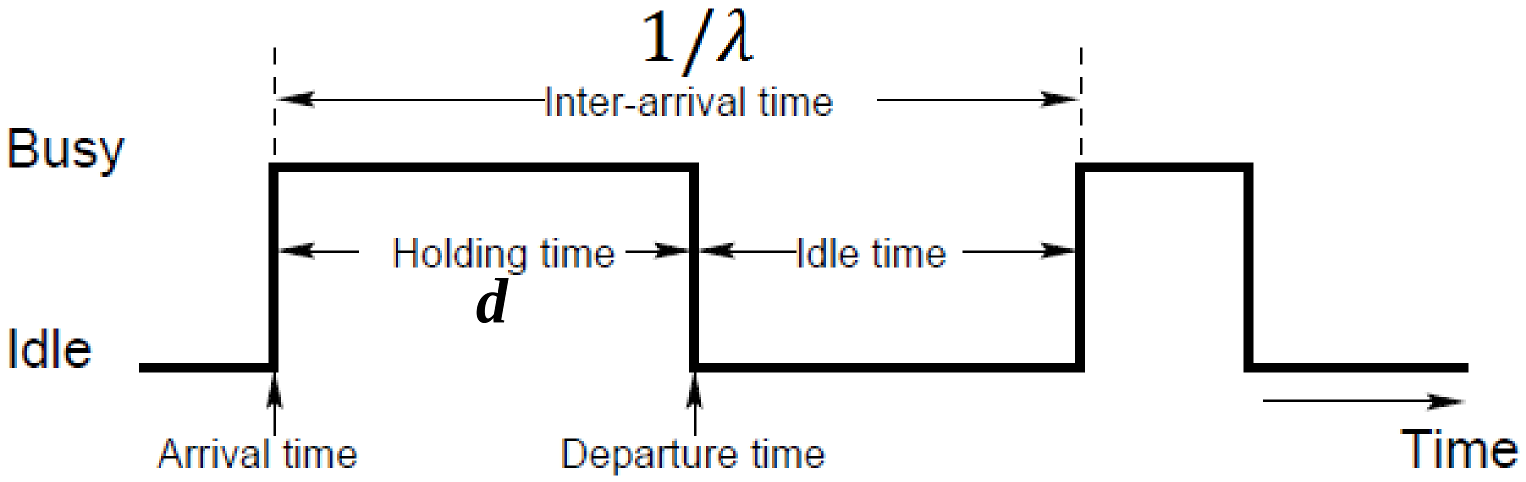


Tráfico en función de intensidad de llamadas y ocupación media

$$\lambda = \frac{llamadas}{segundo}\left(\frac{1}{s}\right)$$

$$d = duración\ media\ \ (s)\ \ o\ \ segundos\ \ por\ \ servicio$$

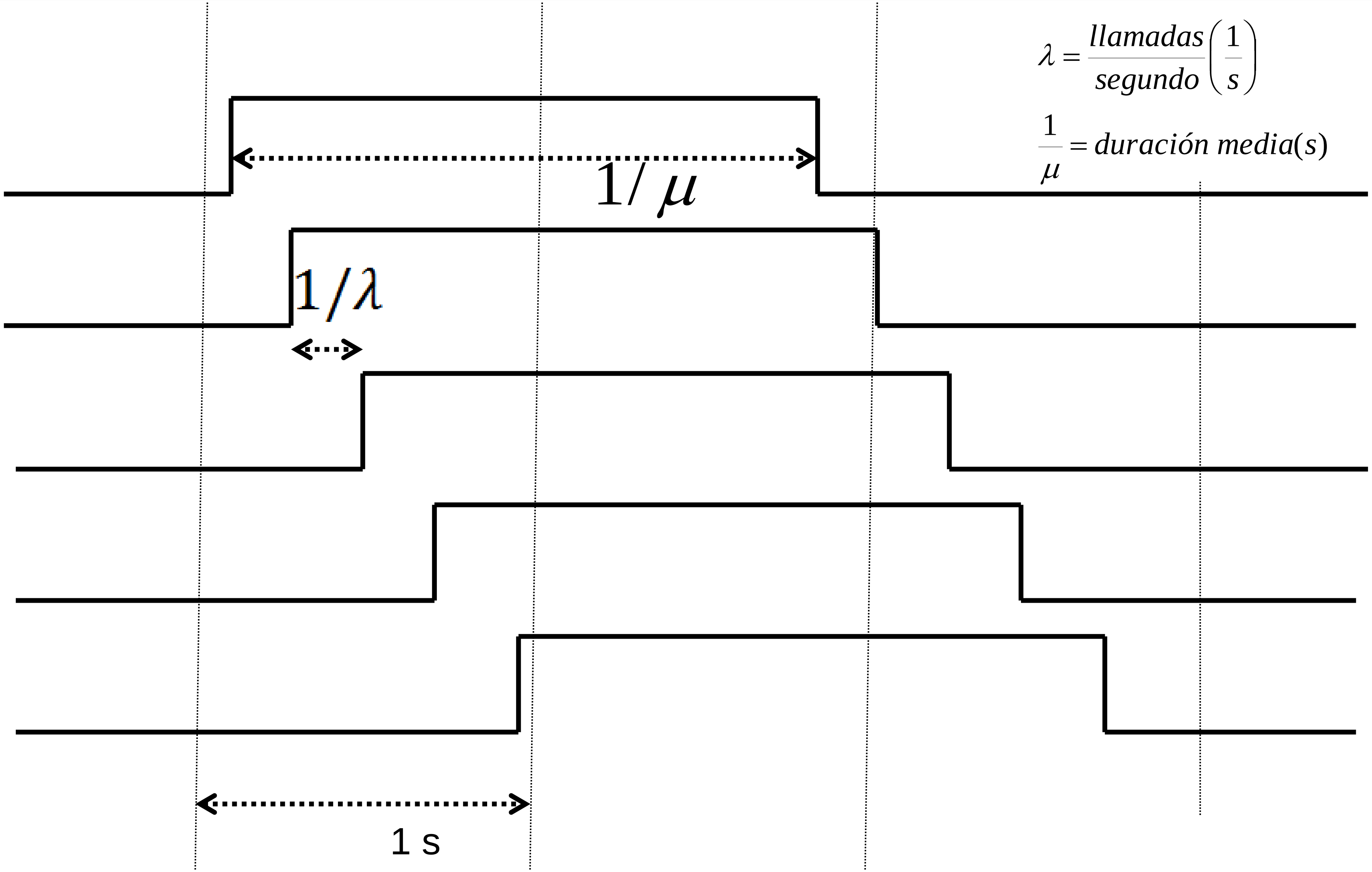
$$\mu = \frac{1}{d}(s^{-1})\ \ tasa\ \ de\ \ servicio$$



Tráfico en función de intensidad de llamadas y ocupación media

$$\lambda = \frac{llamadas}{segundo}\left(\frac{1}{s}\right)$$

$$\frac{1}{\mu} = duración\ media(s)$$



Tráfico en función de intensidad de llamadas y ocupación media

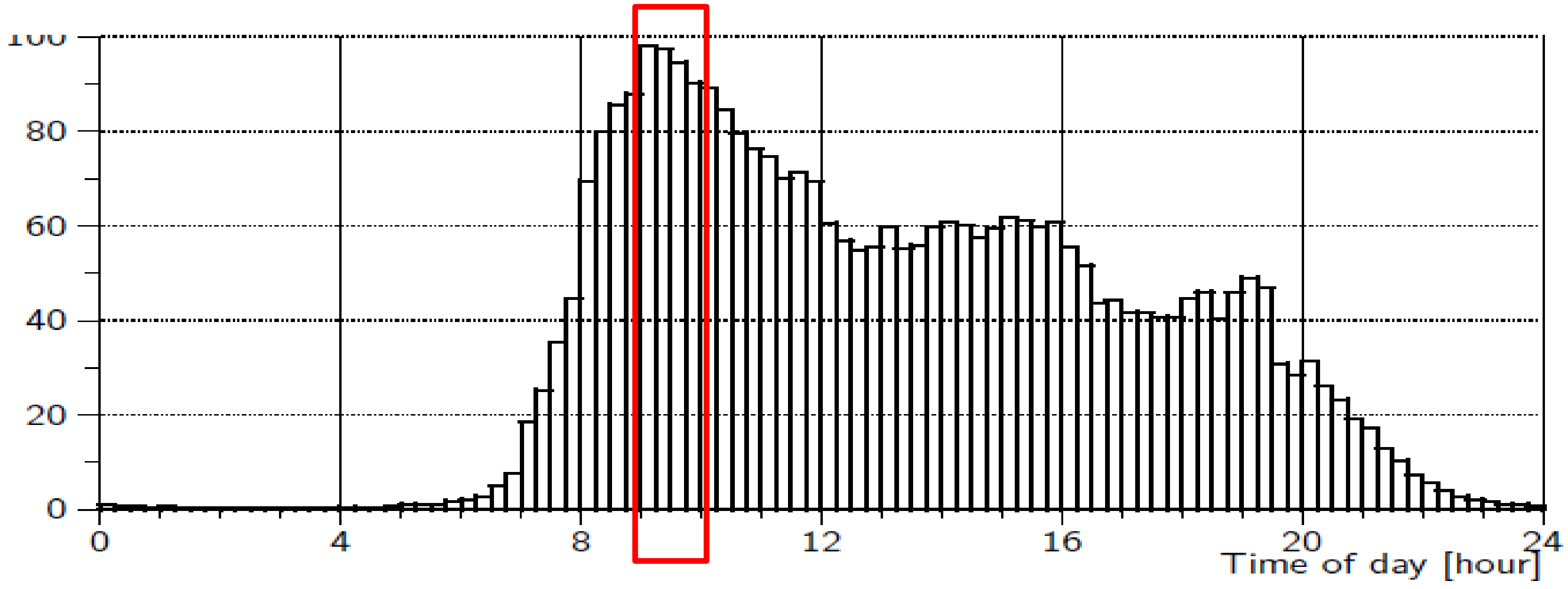
El tráfico $\boldsymbol{A}$ se puede calcular como

$$A = \lambda d = \frac{\lambda}{\mu}$$

Hora pico (*Busy Hour)*

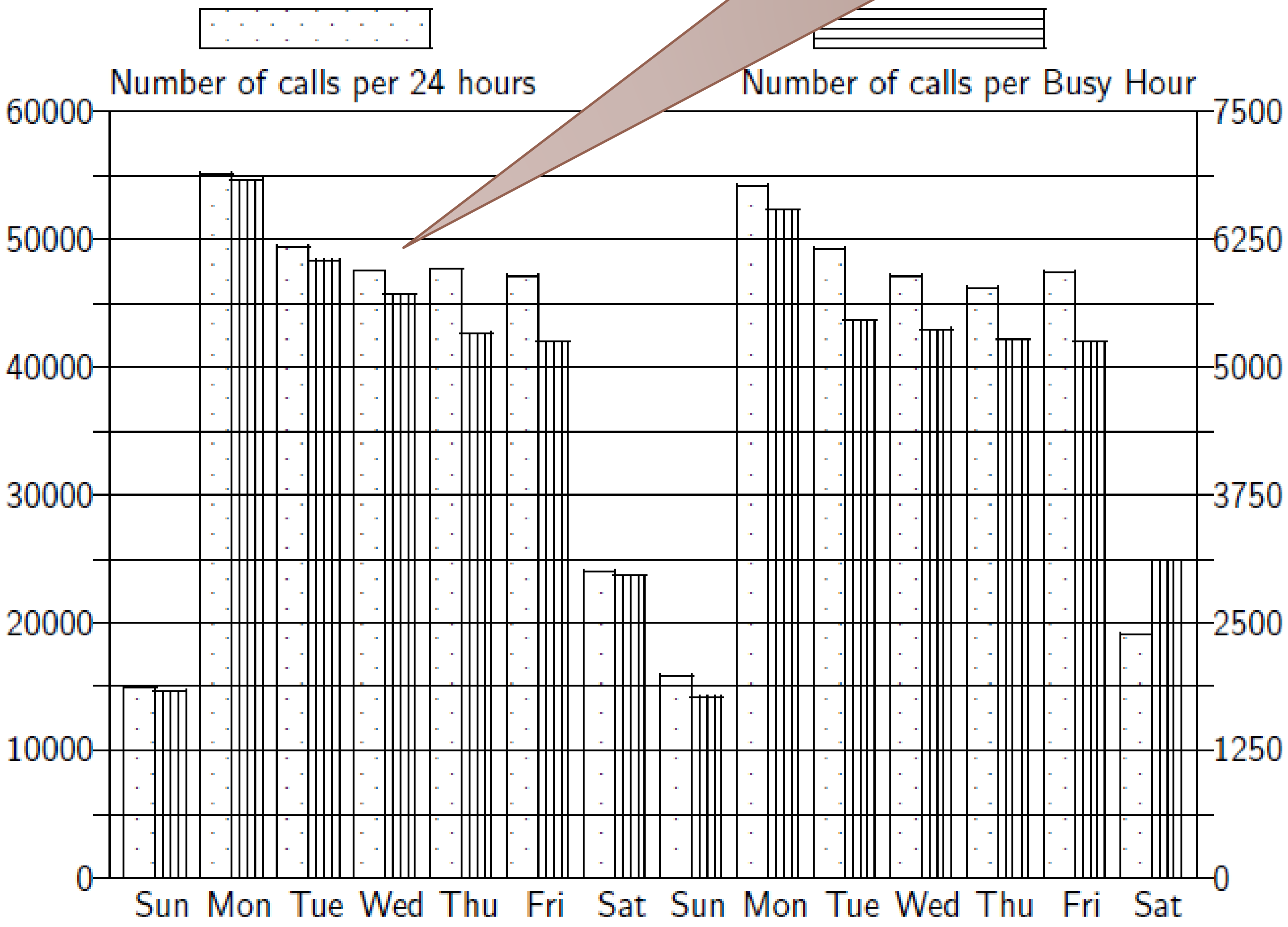
Período de 60 minutos que tiene el máximo de tráfico, tomado en intervalos de 15 minutos

- Por ejemplo: La hora pico puede ser de 9:15 a 10:15

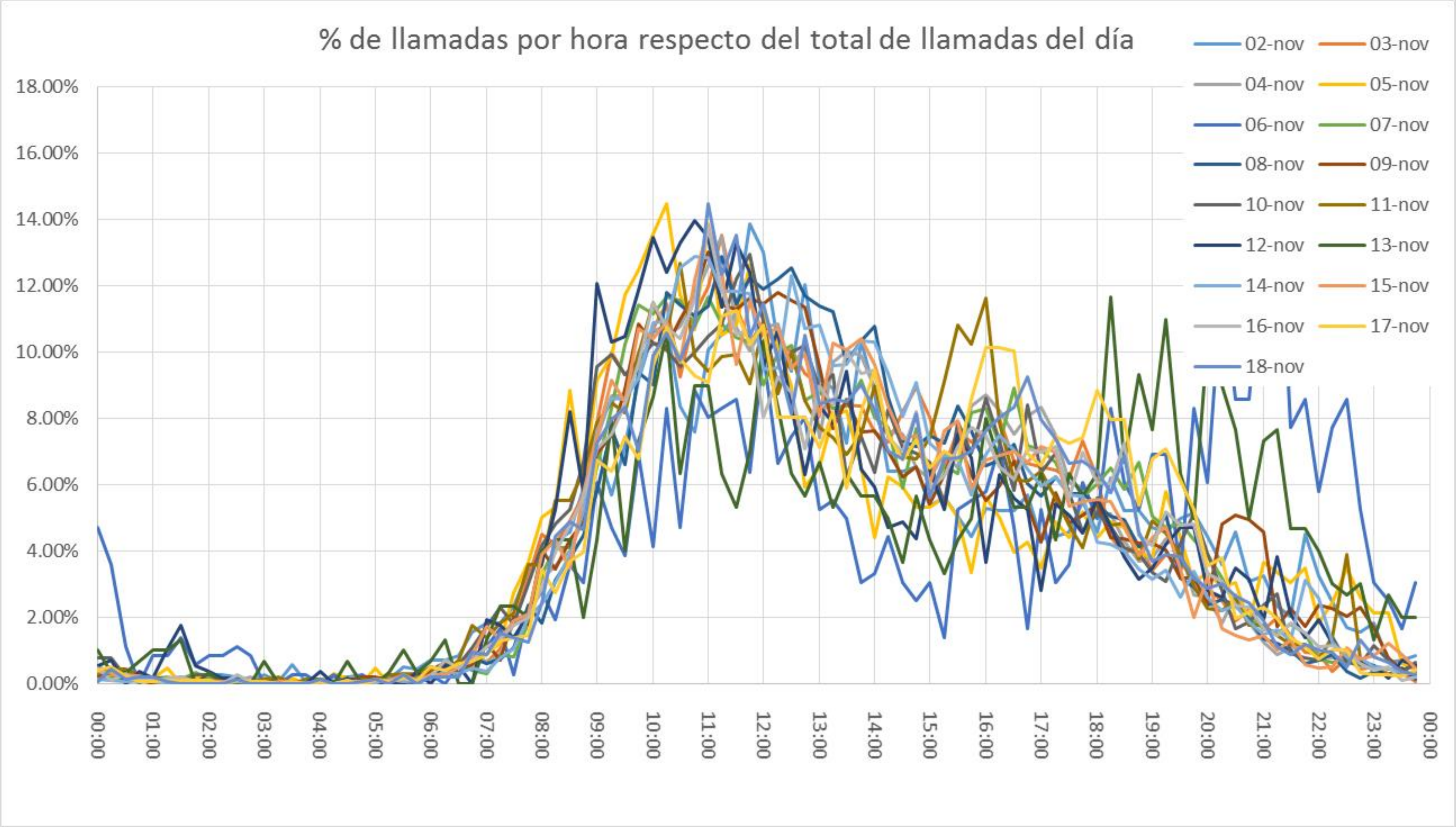


Hora pico (*Busy Hour*)

La relación entre el tráfico en la hora pico y el tráfico diario total es del orden del 12%



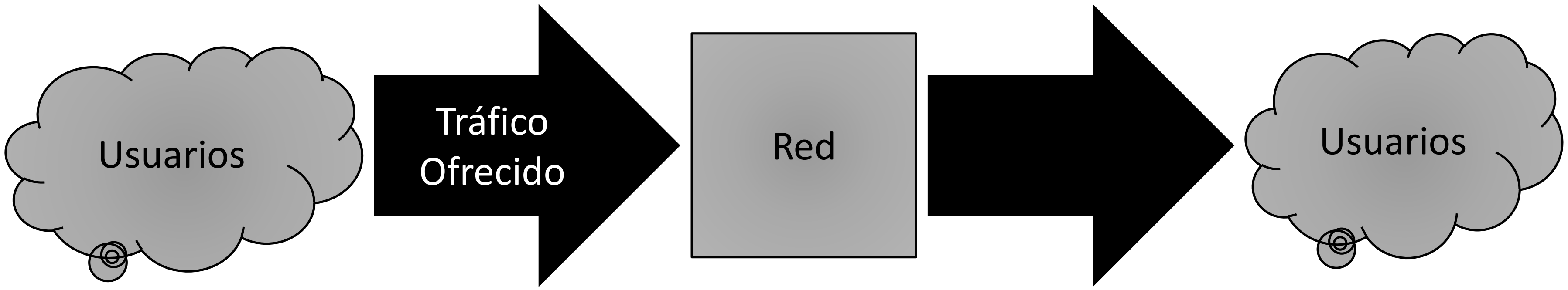
% tráfico respecto al total diario



Tráfico Ofrecido (*Offered Traffic*)

Es el tráfico que se cursaría si no existiera rechazo dentro de la red

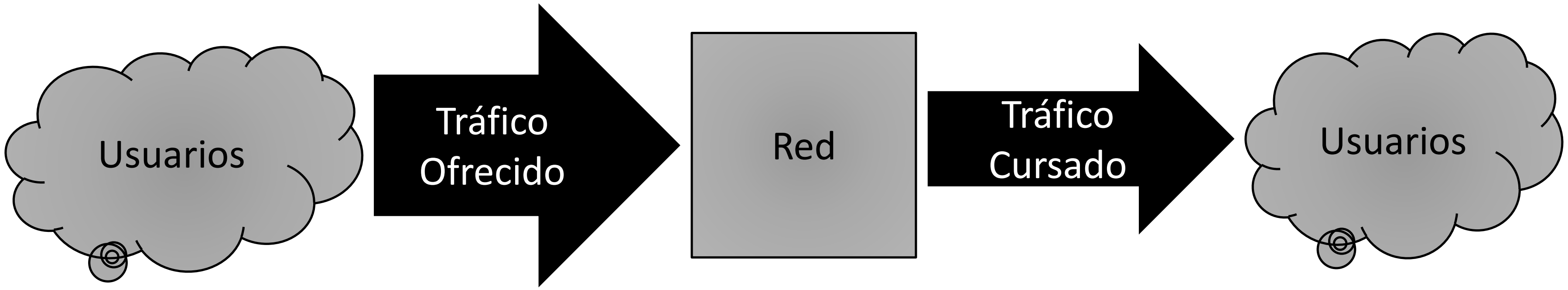
- Se podría cursar con "infinitos" recursos



Tráfico Cursado (*Carried Traffic*)

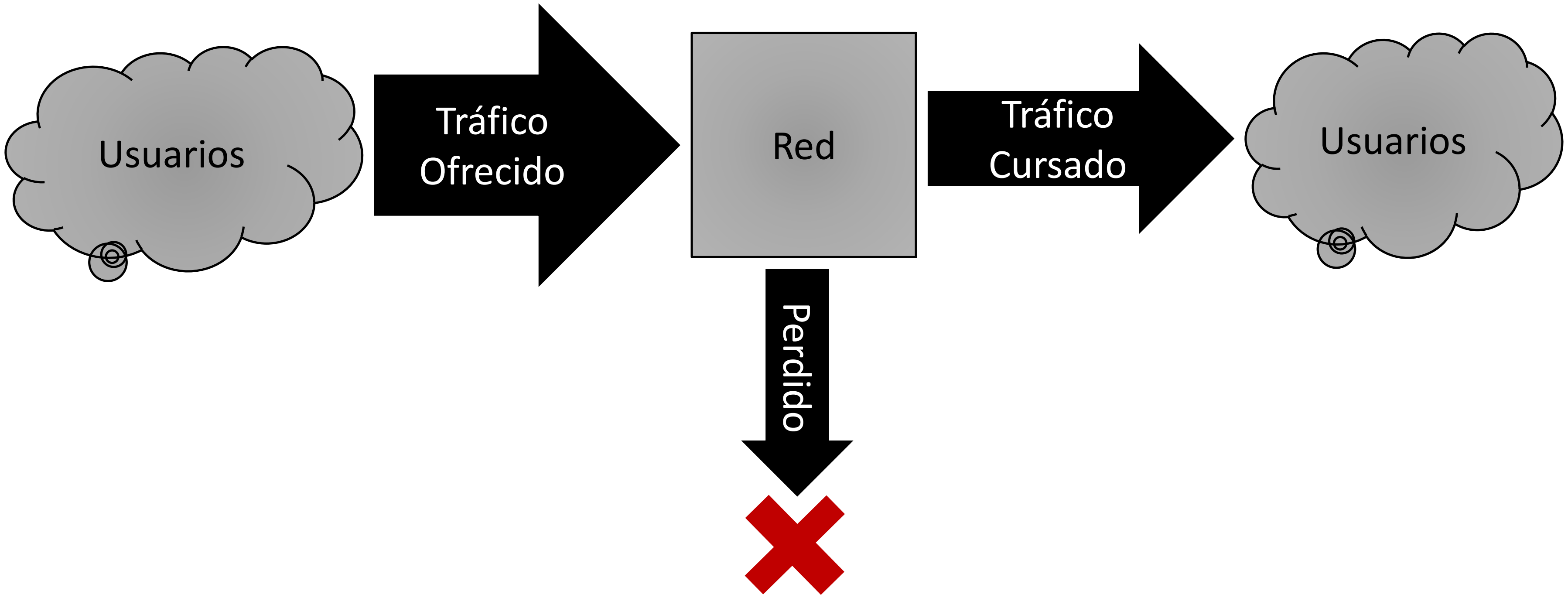
Es el tráfico que efectivamente cursado por la red

- Típicamente “se puede medir”



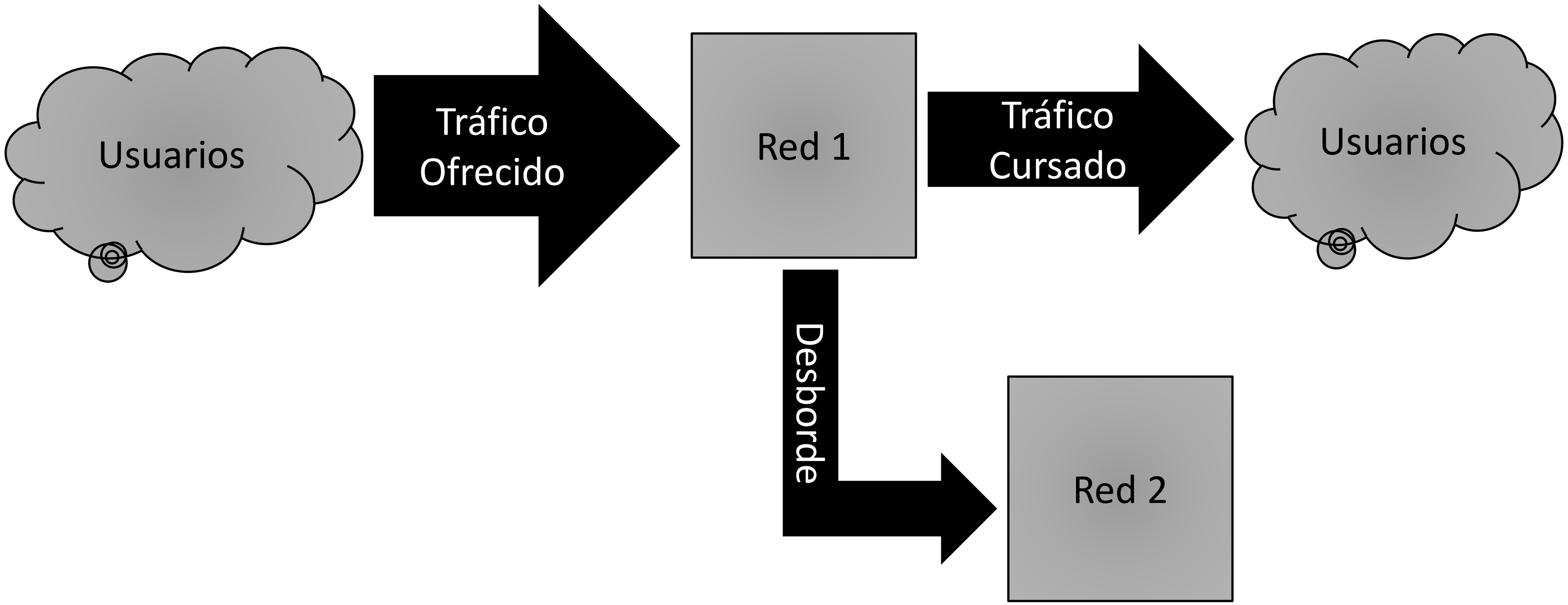
Tráfico Perdido o Rechazado (*Lost Traffic*)

Es el tráfico que NO pudo ser cursado por la red

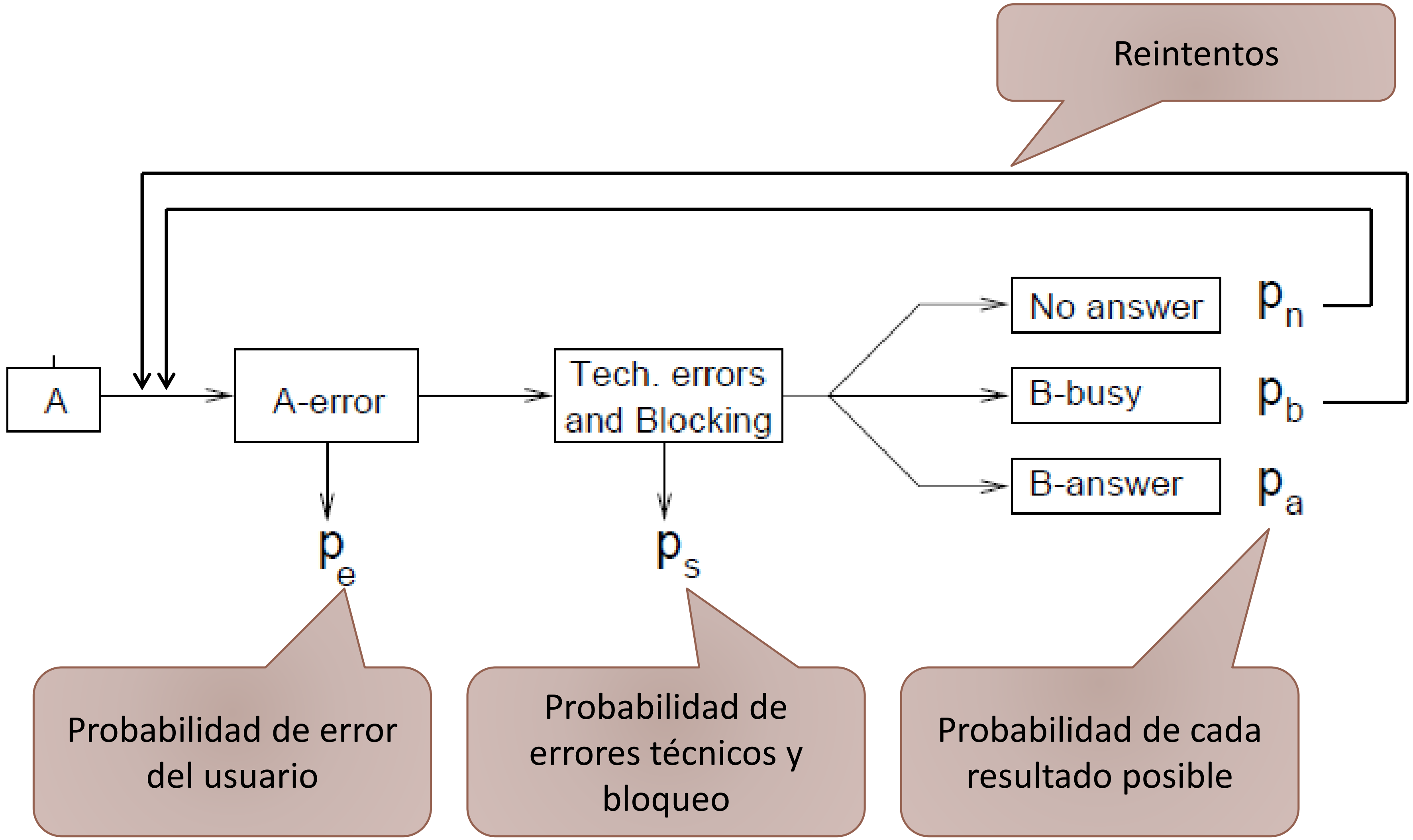


Tráfico de Desborde*Overflow Traffic*

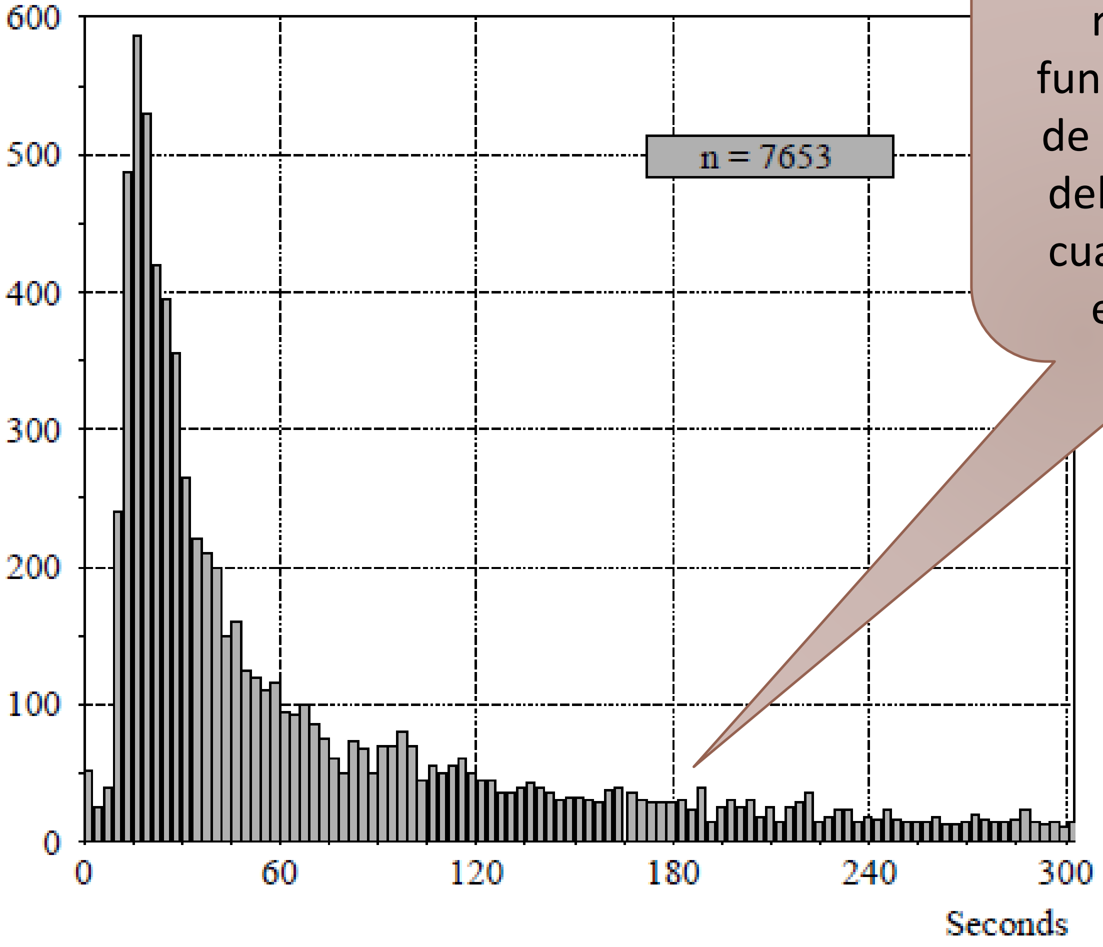
Es el tráfico que no pudo ser cursado por una red y es derivado a otra red



Generación de tráfico



Reintentos cuando el destino está ocupado



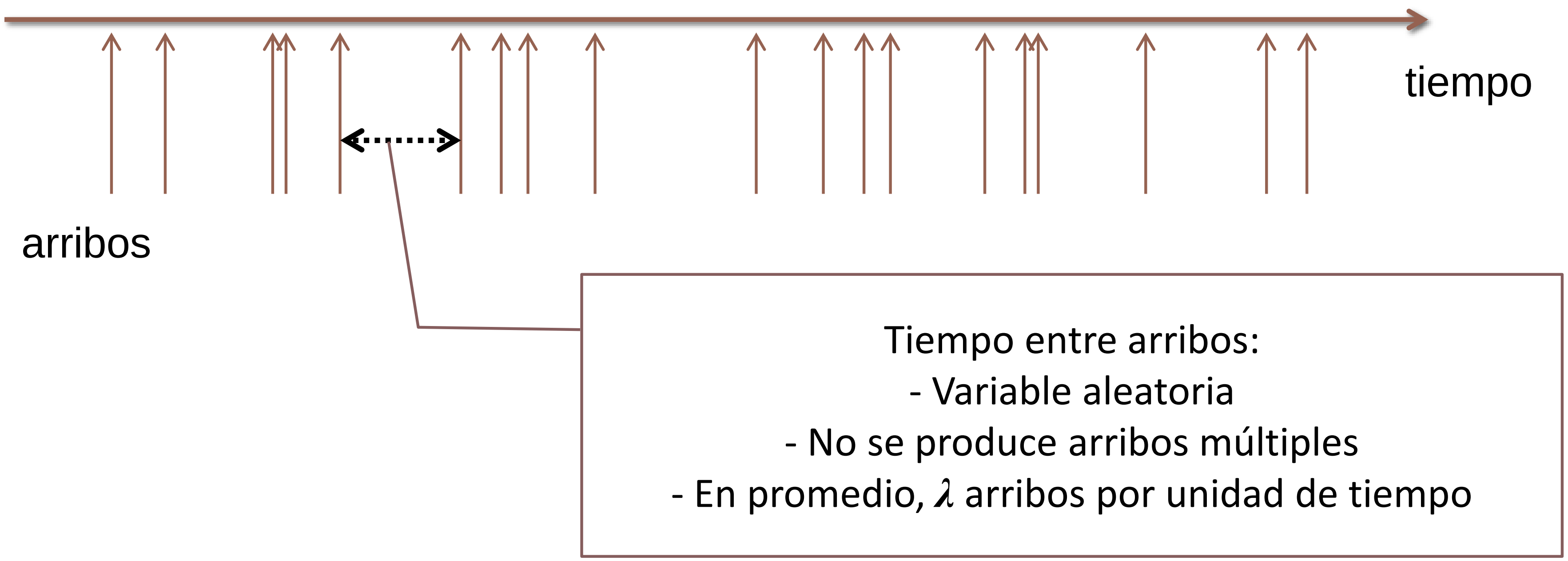
Histograma de los reintentos en función del tiempo de reintento luego del intento inicial, cuando el destino está ocupado

Modelos Matemáticos

TEORÍA E INGENIERÍA DE TELETRÁFICO

Proceso de Arribos

Los procesos de arribo se pueden describir matemáticamente como *procesos estocásticos puntuales*



Proceso de Arribos

Al asumir que cada "tiempo entre arribos" se modela con la misma variable aleatoria, queda implícito que la distribución de arribos es independiente del tiempo

- Puede ser cierto durante períodos cortos de tiempo

Llamaremos **A** a la variable aleatoria que modela el tiempo entre arribos

$$\mathbf{A(T)} = P(\mathbf{A} \leq T) = \int_0^T a(t)dt$$

Proceso de Arribos

El tiempo medio entre arribos (el valor esperado) se puede calcular como

$$\frac{1}{\lambda} = \int_0^\infty t a(t) dt$$

¿Cómo definir **A**?

El "tiempo entre arribos" se puede modelar con una distribución exponencial de parámetro λ

$$A(t) = 1 - e^{-\lambda t}$$

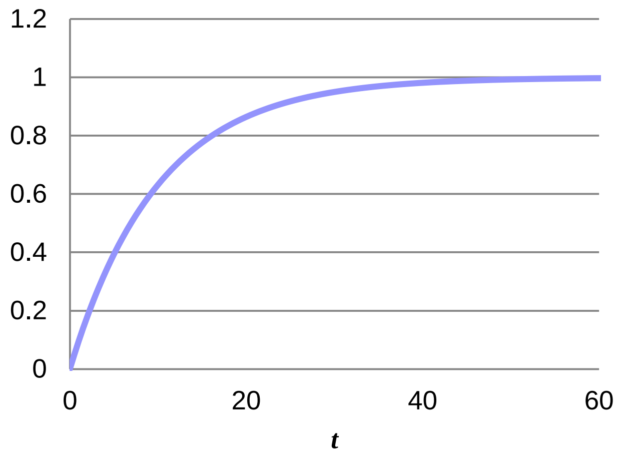
$$a(t) = \lambda e^{-\lambda t}$$

Ejemplo: 360 llamadas por hora

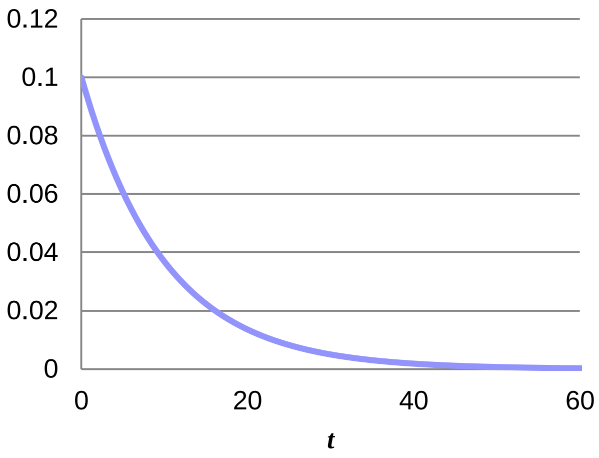
$$\lambda = 360\,\frac{llamadas}{hora} = 0.1\,\frac{llamadas}{segundo}$$

$$\frac{1}{\lambda} = 10\,\frac{segundos}{llamada}(tiempo\ \ promedio\ \ entre\ \ llamadas)$$

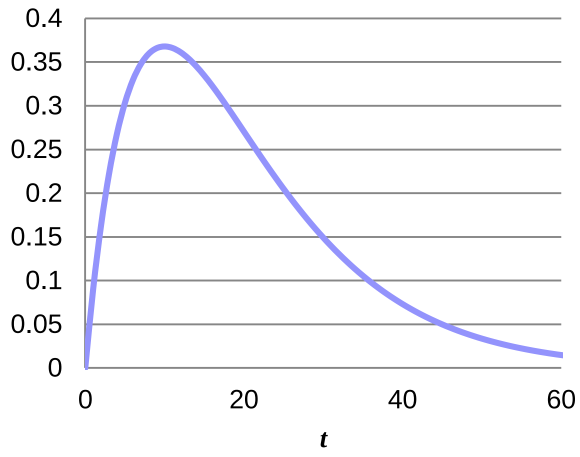
$$A(t) = 1 - e^{-\lambda t}$$



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



$$ta(t) = \lambda t e^{-\lambda t}$$



Propiedades de la distribución Exponencial de parámetro λ

Valor esperado

$$E(\mathbf{A}) = \frac{1}{\lambda}$$

Varianza var **A** = $E(\mathbf{A}^2) - E(\mathbf{A})^2$

$$\mathrm{var}(\mathbf{A}) = \frac{1}{\lambda^2}$$

Propiedades de la distribución exponencial

"No tiene memoria":

$$P(\mathbf{A} > T + h \mid \mathbf{A} \geq T) = P(\mathbf{A} > h)$$

$$P(\mathbf{A} > T) = \int_T^{\infty} a(t)dt = \int_T^{\infty} \lambda e^{-\lambda t} dt$$

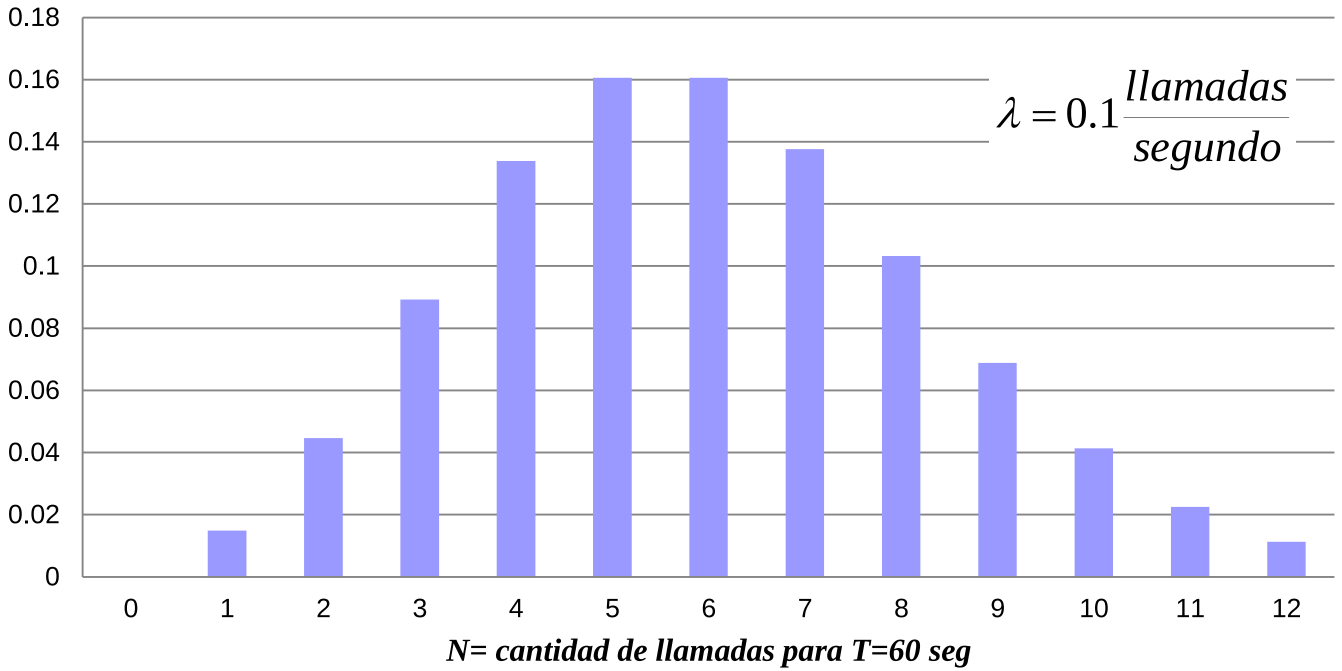
$$P(\mathbf{A} > T + h | \mathbf{A} > T) = \frac{\int_{T+h}^{\infty} \lambda e^{-\lambda t} dt}{\int_T^{\infty} \lambda e^{-\lambda t} dt} = \frac{e^{-\lambda(T+h)}}{e^{-\lambda T}} = e^{-\lambda h} = \int_h^{\infty} \lambda e^{-\lambda t} dt = P(\mathbf{A} > h)$$

La distribución de probabilidad del arribo de una nueva llamada no depende de cuanto tiempo haya pasado desde el arribo de la última llamada

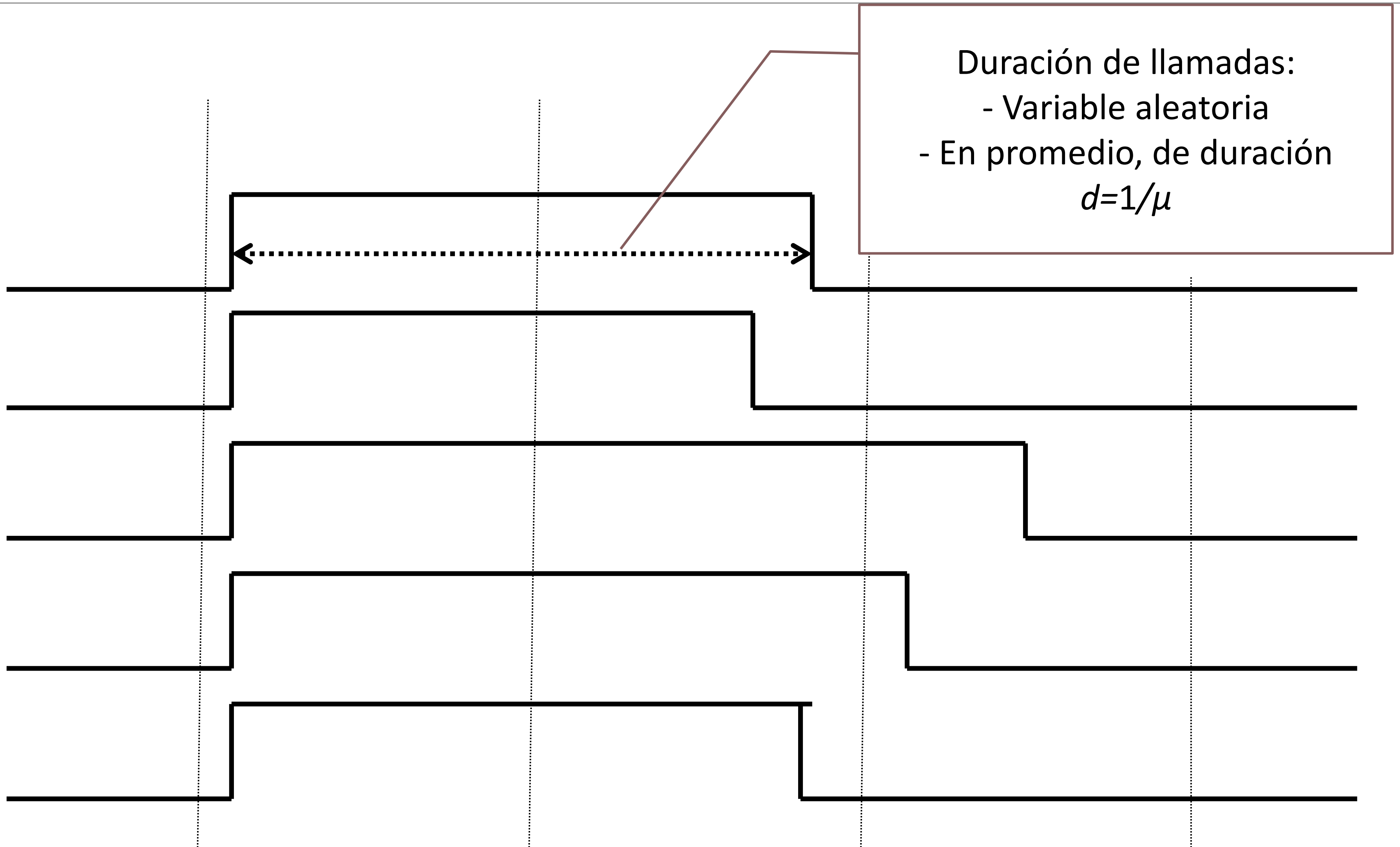
Relación entre la distribución Exponencial y la de Poisson

Si los tiempos entre arribos son exponenciales de parámetro λ , el número de arribos N que ocurren en un lapso de tiempo T tiene un distribución de Poisson de parámetro λT

$$P(N) = \frac{(\lambda T)^N e^{-\lambda T}}{N!}$$



Duración de las llamadas



Duración de las llamadas

Al asumir que la duración de las llamadas se modela con la misma variable aleatoria, queda implícito que la distribución de la duración es independiente del tiempo

- Puede ser cierto durante períodos cortos de tiempo y para un mismo tipo de llamadas

Llamaremos **S** a la variable aleatoria que modela la duración de llamadas

$$S(T) = P(S \leq T) = \int_0^T s(t)dt$$

Duración de las llamadas

La duración media de las llamadas (el valor esperado) se puede calcular como

$$d = \frac{1}{\mu} = \int_0^{\infty} t s(t) dt$$

La duración de las llamadas se puede modelar con una distribución Exponencial de parámetro μ=1/d

$$S(t) = 1 - e^{-\mu t}$$

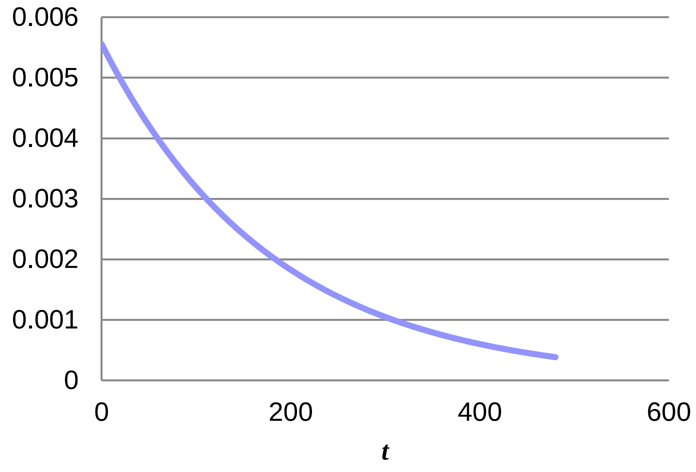
$$s(t) = \mu e^{-\mu t}$$

Ejemplo: duración media de 3 minutos

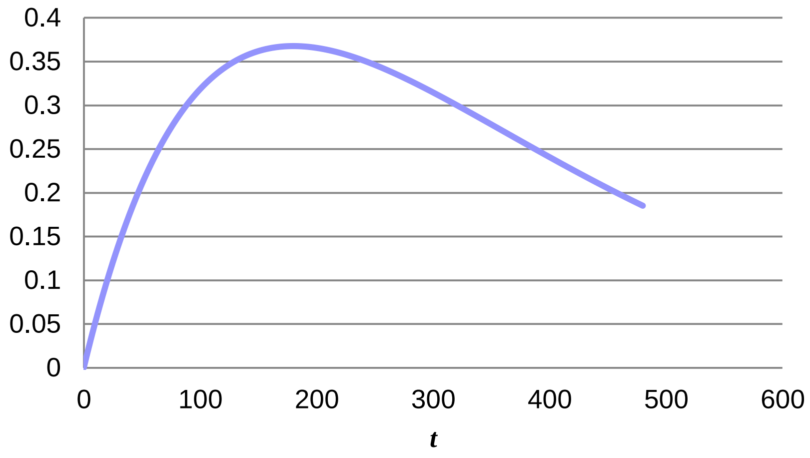
$$d = 180 \quad segundos$$

$$\mu = 1/180 \quad segundos^{-1}$$

$$s(t) = \frac{1}{d} e^{-\frac{t}{d}}$$



$$ts(t) = \frac{1}{d} t e^{-\frac{t}{d}}$$



Modelo con “infinitos recursos”

TEORÍA E INGENIERÍA DE TELETRÁFICO

Consideraciones

Consideramos un sistema con ∞ recursos idénticos, trabajando en paralelo

- "Grupo homogéneo" (*homogeneous group*)

Hay ∞ fuentes que pueden generar tráfico

Una llamada es aceptada en el sistema si existe por lo menos un recurso disponible, y cada llamada ocupa un único recurso

- "Accesibilidad completa" (*full accessibility*)
- Dado que hay ∞ recursos, las llamadas son siempre aceptadas

El arribo de llamadas se puede modelar como un proceso de Poisson de parámetro λ llamadas por segundo

- "Tráfico de Chance Pura"

La duración de las llamadas tiene una distribución exponencial de parámetro μ=1/d segundos^{-1}

El tráfico se puede modelar como un proceso de "nacimiento y muerte"

- Proceso simple de Markov

Diagrama de transición de estados

Definimos el estado del sistema $[i]$ como la cantidad de recursos ocupados i.

En un instante determinado, el sistema se encuentra en el estado $[i]$

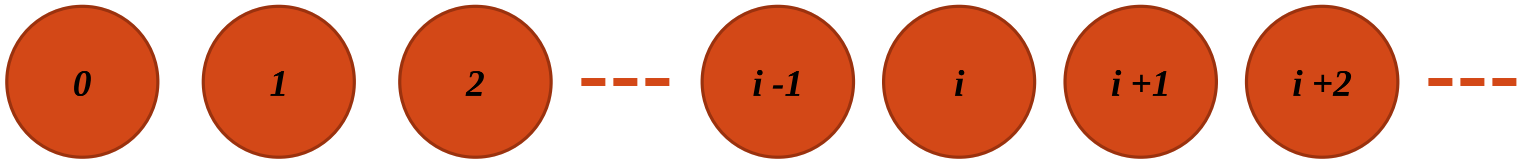


Diagrama de transición de estados

Con el tiempo pueden existir transiciones entre estados

Entre un tiempo ***t*** y ***t +dt***, solo existen transiciones simples

- La probabilidad de arribo o fin de más de 2 llamadas en ***dt*** es despreciable

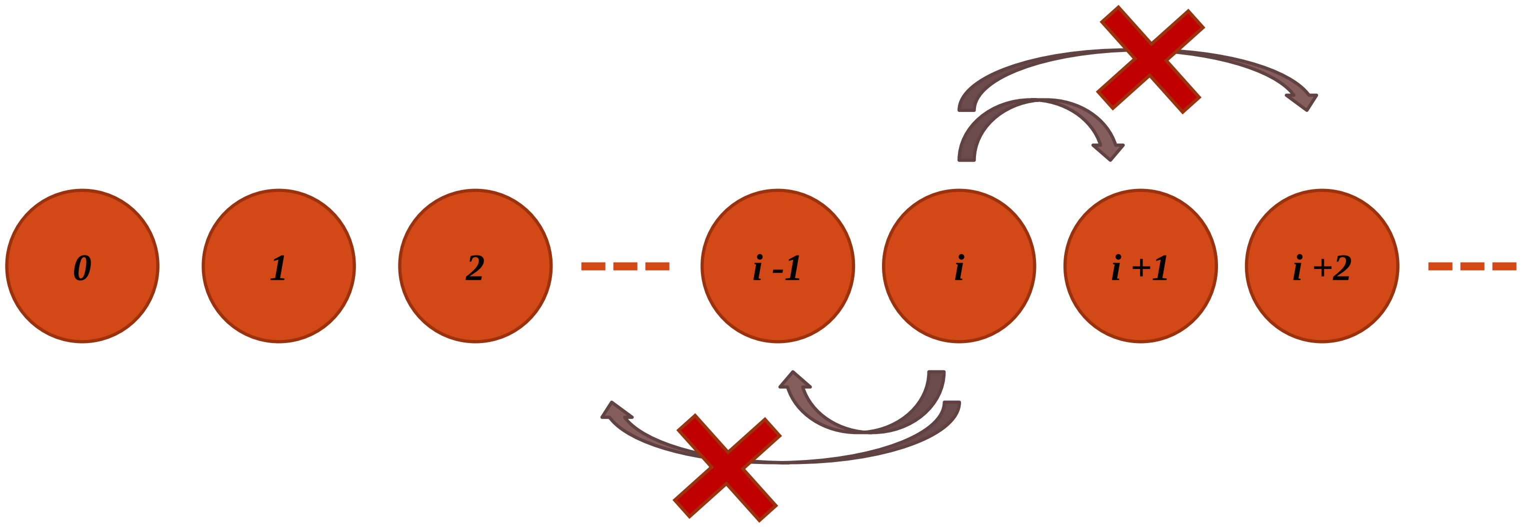


Diagrama de transición de estados

Demostración para el caso de arribos:

Probabilidad de N arribos en un tiempo T:

$$P(N) = \frac{(\lambda T)^N e^{-\lambda T}}{N!}$$

$$T \rightarrow 0$$

$$P(1) = \lambda T e^{-\lambda T} \approx \lambda T$$

$$P(2) = \frac{(\lambda T)^2 e^{-\lambda T}}{2} \approx \frac{(\lambda T)^2}{2} << P(1)$$

Diagrama de transición de estados

En una situación de *equilibrio estadístico*, el sistema se encontrará en el estado $[i]$ una proporción de tiempo $p(i)$

$p(i)$ es la probabilidad de encontrar al sistema en el estado $[i]$

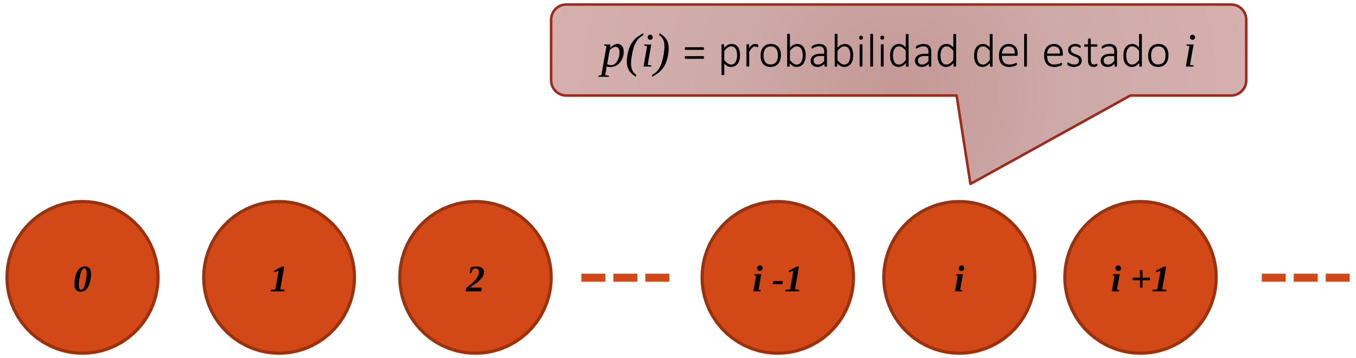


Diagrama de transición de estados

La probabilidad de pasar del estado [i] al [$i+1$] en un intervalo de tiempo T, es la probabilidad de que exista un arribo en ese intervalo de tiempo.

Si T es muy pequeño, esa probabilidad es

$$P(1) = \lambda T e^{-\lambda T} \approx \lambda T$$

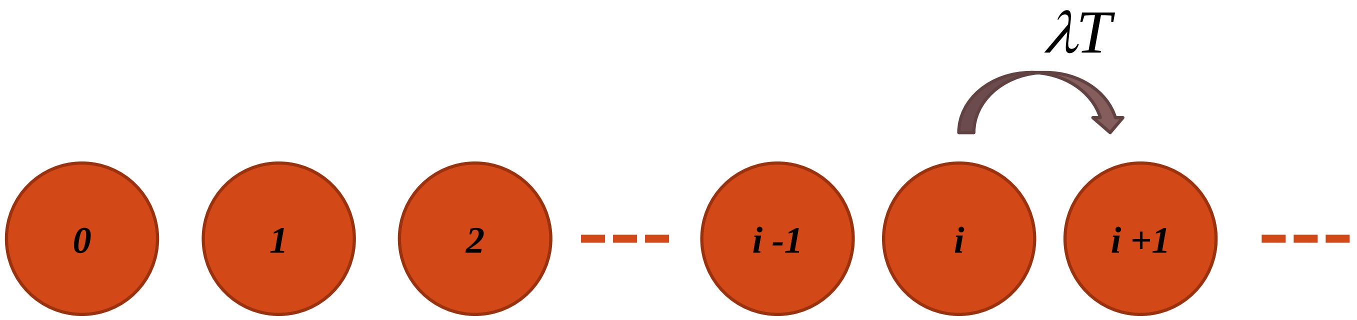


Diagrama de transición de estados

La probabilidad de pasar del estado $[i]$ al $[i+1]$ durante un tiempo corto T

- Es lineal con T
- Solo depende de la tasa de arribos λ, pero no del estado $[i]$ en el que se encuentre el sistema
  - Hay “infinitas fuentes” generadoras de tráfico

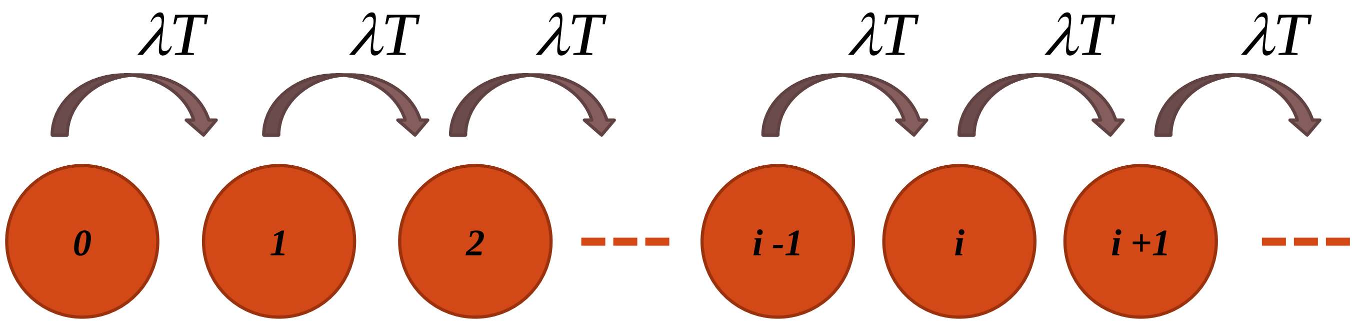


Diagrama de transición de estados

La probabilidad de pasar del estado [i] al [i-1] en un intervalo de tiempo T, es la probabilidad de que termine una llamada en $t < T$

La probabilidad de que termine ***1*** de las i llamadas del estado [i] es

Como hay i llamadas $\mathrm{p}(\ [i] \rightarrow [i\text{-}1]\ ) = i\mu T$

$$S(T) = 1 - e^{-\mu T}$$

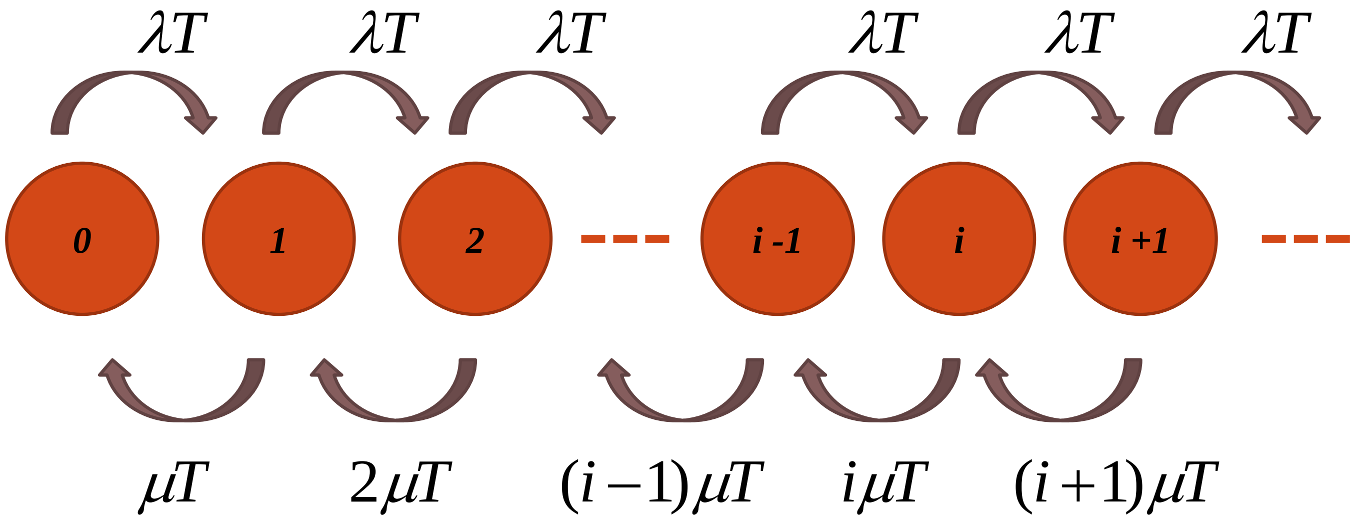
$$S(T) = 1 - (1 - \mu T + \frac{(\mu T)^2}{2} - \ldots) \approx \mu T \quad (para\, T \rightarrow 0)$$

Diagrama de transición de estados

Si T es muy pequeño, la probabilidad de pasar del estado [i] al [i-1] en un intervalo de tiempo T

- Es lineal con T
- Es inversamente proporcional a la duración media d
  - Es proporcional a μ=1/d
- Es proporcional a la cantidad de llamadas (recursos ocupados) en el sistema
  - Es proporcional a i

Diagrama de transición de estados



Válido para T muy pequeño

Ecuaciones de nodo

$p(i)$ es la probabilidad de encontrar al sistema en el estado $[i]$

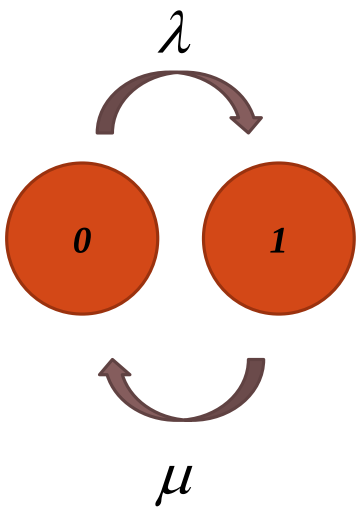
La cantidad media de "saltos" del estado $[0]$ al estado $[1]$ en el intervalo T es $\lambda Tp(0)$

- La cantidad de "saltos" del estado $[0]$ al estado $[1]$ por unidad de tiempo es $\lambda Tp(0)/T=\lambda p(0)$

La cantidad media de "saltos" del estado $[1]$ al estado $[0]$ en el intervalo T es $\mu Tp(1)$

- La cantidad de "saltos" del estado $[1]$ al estado $[0]$ por unidad de tiempo es $\mu Tp(1)/T= \mu p(1)$

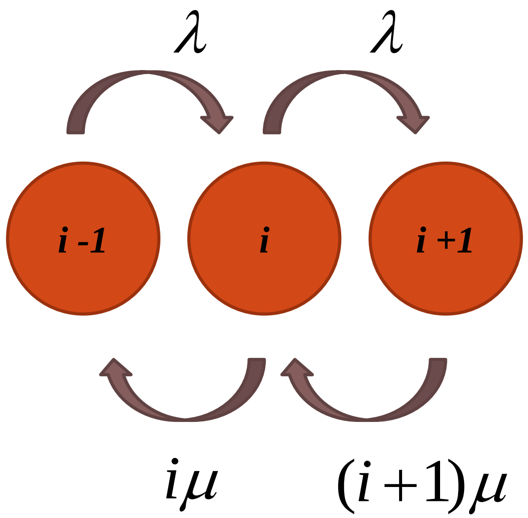
Ecuaciones de nodo



$$\lambda p(0) = \mu p(1)$$

Transiciones por unidad de tiempo

Ecuaciones de nodo



$$i > 0$$

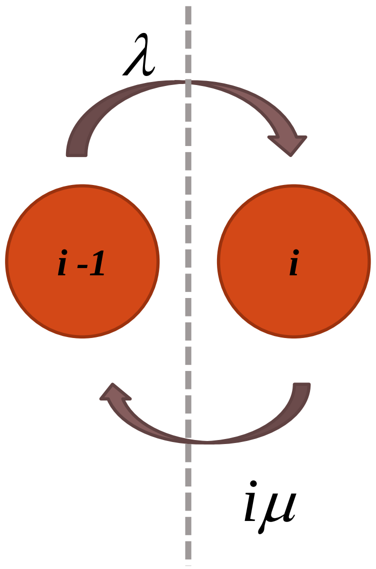
$$\lambda p(i-1) + (i+1)\mu p(i+1) = \lambda p(i) + i\mu p(i)$$

$$\lambda p(i-1) + (i+1)\mu p(i+1) = (\lambda + i\mu)\, p(i)$$

Transiciones por unidad de tiempo

Ecuaciones de corte

En equilibro estadístico, la cantidad de transiciones del estado [i-1] al [i] deben ser iguales a las transiciones del estado [i] al [i-1]



Transiciones por unidad de tiempo

$$i > 0$$

$$\lambda p(i-1) = i\mu p(i)$$

Normalización

El sistema siempre estará en uno de los posibles estados

- La suma de todas las probabilidades de los estados debe ser 1

$$\sum_{i=0}^{\infty} p(i) = 1$$

Deducción de las probabilidades de estados

Aplicamos las ecuaciones de "corte"

$$\lambda p(0) = \mu p(1)$$

$$\lambda p(1) = 2\mu p(2)$$

$$\lambda p(2) = 3\mu p(3)$$

$$\dots\dots$$

$$\lambda p(i-2) = (i-1)\mu p(i-1)$$

$$\lambda p(i-1) = (i)\mu p(i)$$

$$\lambda p(i) = (i+1)\mu p(i+1)$$

$$\dots\dots$$

$$\lambda p(i-1) = i\mu p(i)$$

$$A = \lambda d = \frac{\lambda}{\mu}$$

Deducción de las probabilidades de estados

$$p(1) = Ap(0)$$

$$p(2) = \frac{A}{2} p(1) = \frac{A^2}{2} p(0)$$

$$p(3) = \frac{A}{3} p(2) = \frac{A^3}{3x2} p(0)$$

....

$$p(i-1) = \frac{A}{i-1} p(i-2) = \frac{A^{i-1}}{(i-1)!} p(0)$$

$$p(i) = \frac{A}{i} p(i-1) = \frac{A^i}{i!} p(0)$$

....

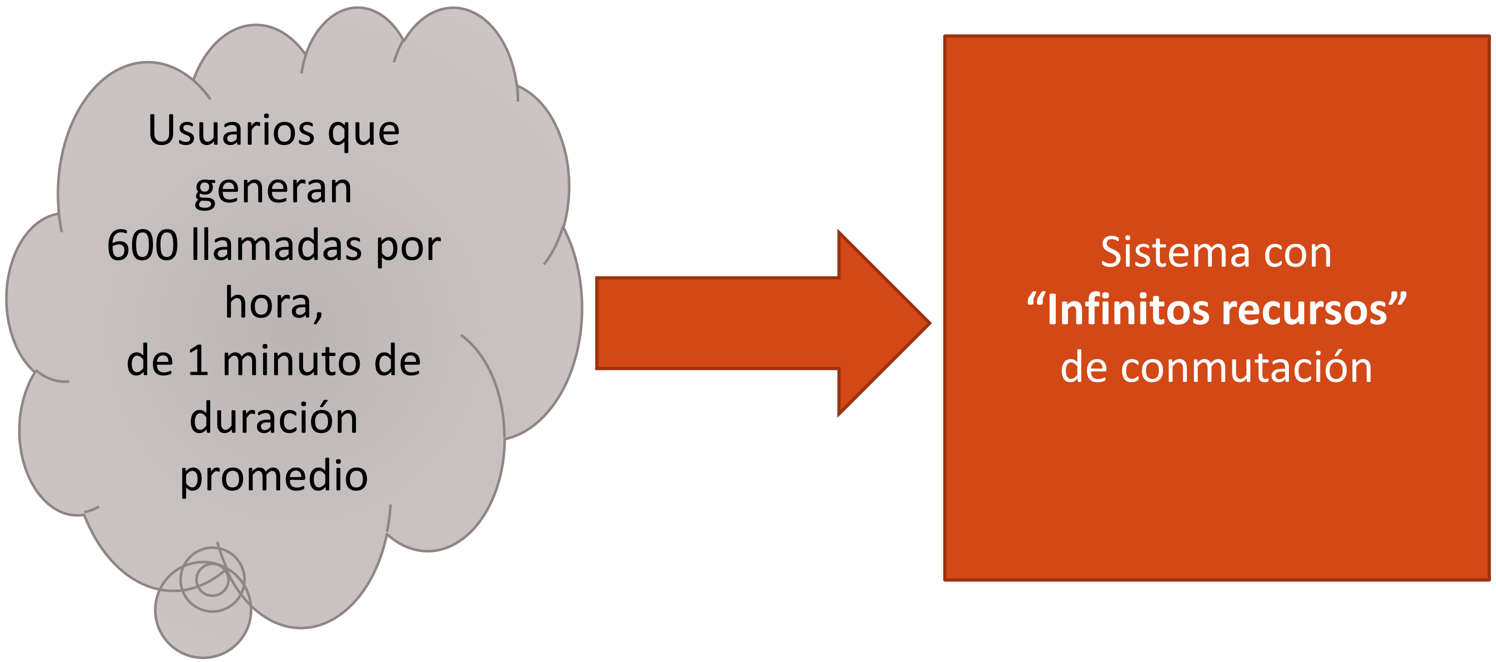
$$\sum_{i=0}^{\infty} p(i) = 1$$

$$\sum_{i=0}^{\infty} \frac{A^i}{i!} p(0) = 1 \Rightarrow$$

$$p(0) = \frac{1}{\sum_{i=0}^{\infty} \frac{A^i}{i!}} = e^{-A}$$

$$\boxed{p(i) = \frac{A^i}{i!} e^{-A}}$$

Ejemplo



¿Cuántos recursos de conmutación serán utilizados?

Ejemplo

600 llamadas por hora = 600/3600 llamadas por segundo

- λ=1/6 seg^{-1}

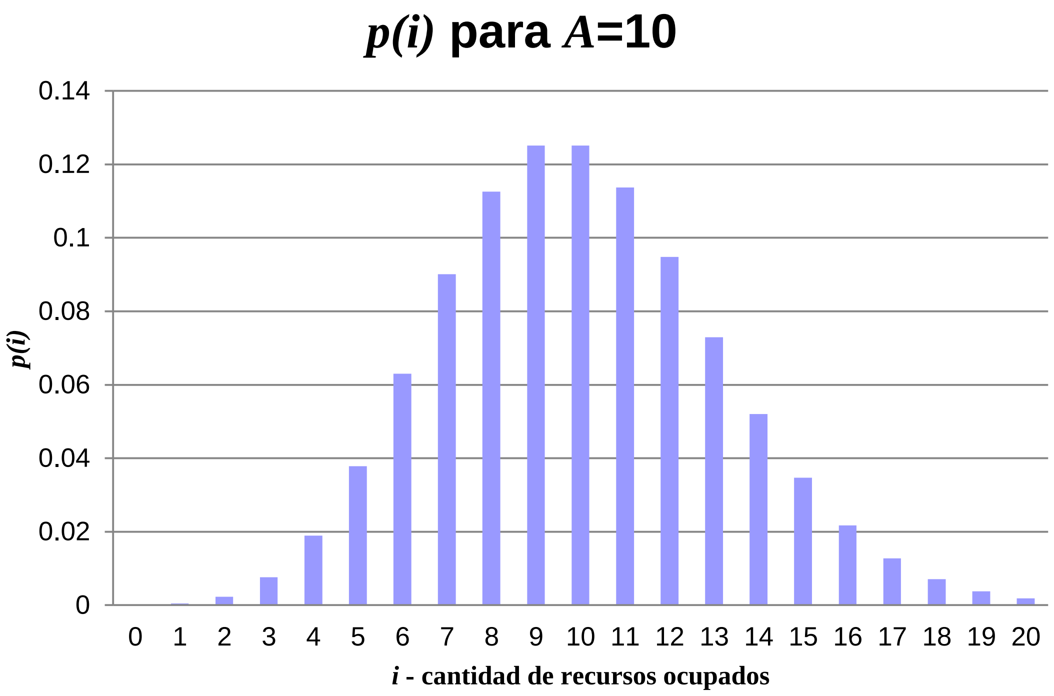
60 segundos
de duración

- d=60 seg
- μ=1/60 seg^{-1}

$A = \lambda/\mu =$ 10 Erlang

¿Cuántos recursos de conmutación serán utilizados?

$$p(i) = \frac{A^i}{i!} e^{-A}$$



Características del tráfico con infinitos recursos

No hay "congestión": Hay infinitos recursos y todos son accesibles

El tráfico cursado Y es igual al tráfico ofrecido

$$Y = \sum_{i=1}^{\infty} i p(i) = \sum_{i=1}^{\infty} i \frac{A^i}{i!} e^{-A} = e^{-A} \sum_{i=1}^{\infty} A \frac{A^{i-1}}{(i-1)!} = A e^{-A} e^{A} = A$$

El tráfico perdido es 0

Recursos finitos: Modelos de pérdida

TEORÍA E INGENIERÍA DE TELETRÁFICO

Consideraciones

Consideramos un sistema con n recursos idénticos, trabajando en paralelo

- "Grupo homogéneo" (*homogeneous group*)

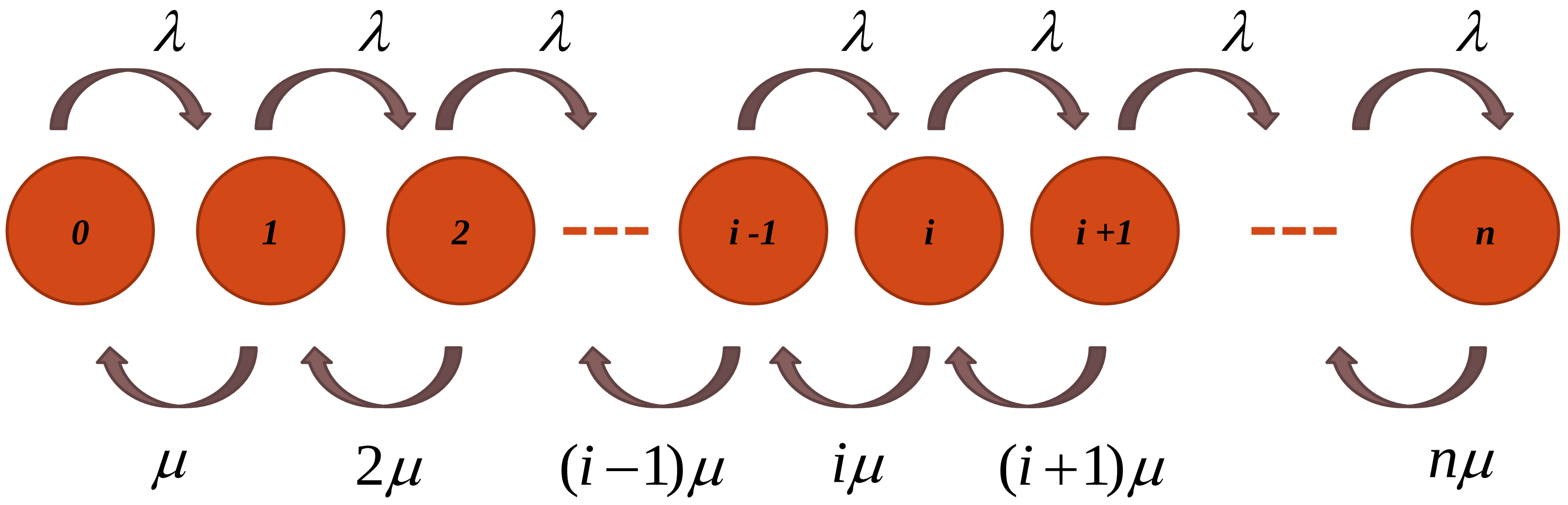
Una llamada es aceptada en el sistema si existe por lo menos un recurso disponible, y cada llamada ocupa un único recurso

- "Accesibilidad completa" (*full accessibility*)

Si todos los recursos están ocupados el sistema está "congestionado" y el intento de llamada es bloqueado

- El intento de llamada en este caso "desaparece", no hay espera ni reintentos
- **Modelo de pérdida (*Lost Calls Cleared*)**

Diagrama de transición de estados



Transiciones por unidad de tiempo

Deducción de las probabilidades de estados

$$p(1) = Ap(0)$$

$$p(2) = \frac{A}{2} p(1) = \frac{A^2}{2} p(0)$$

$$p(3) = \frac{A}{3} p(2) = \frac{A^3}{3x2} p(0)$$

....

$$p(i) = \frac{A}{i} p(i-1) = \frac{A^i}{i!} p(0)$$

....

$$p(n) = \frac{A}{n} p(n-1) = \frac{A^n}{n!} p(0)$$

$$\sum_{i=0}^{n} p(i) = 1$$

$$\sum_{i=0}^{n} \frac{A^i}{i!} p(0) = 1 \Rightarrow$$

$$p(0) = \frac{1}{\sum_{i=0}^{n} \frac{A^i}{i!}}$$

$$\boxed{p(i) = \frac{A^i}{i!} \frac{1}{\sum_{j=0}^{n} \frac{A^j}{j!}}}$$

Congestión

Hay "congestión" cuando todos los recursos están ocupados, o sea cuando el sistema está en el estado $[n]$

La probabilidad de que al llegar un "arribo" encuentre al sistema en el estado $[n]$ es igual a la probabilidad estacionaria de que el sistema se encuentra en el estado $[n]$

- Esto es conocido como propiedad "**Poisson Arrivals See Time Average**" o **PASTA**, demostrada por Ronald W. Wolff en 1982

Congestión

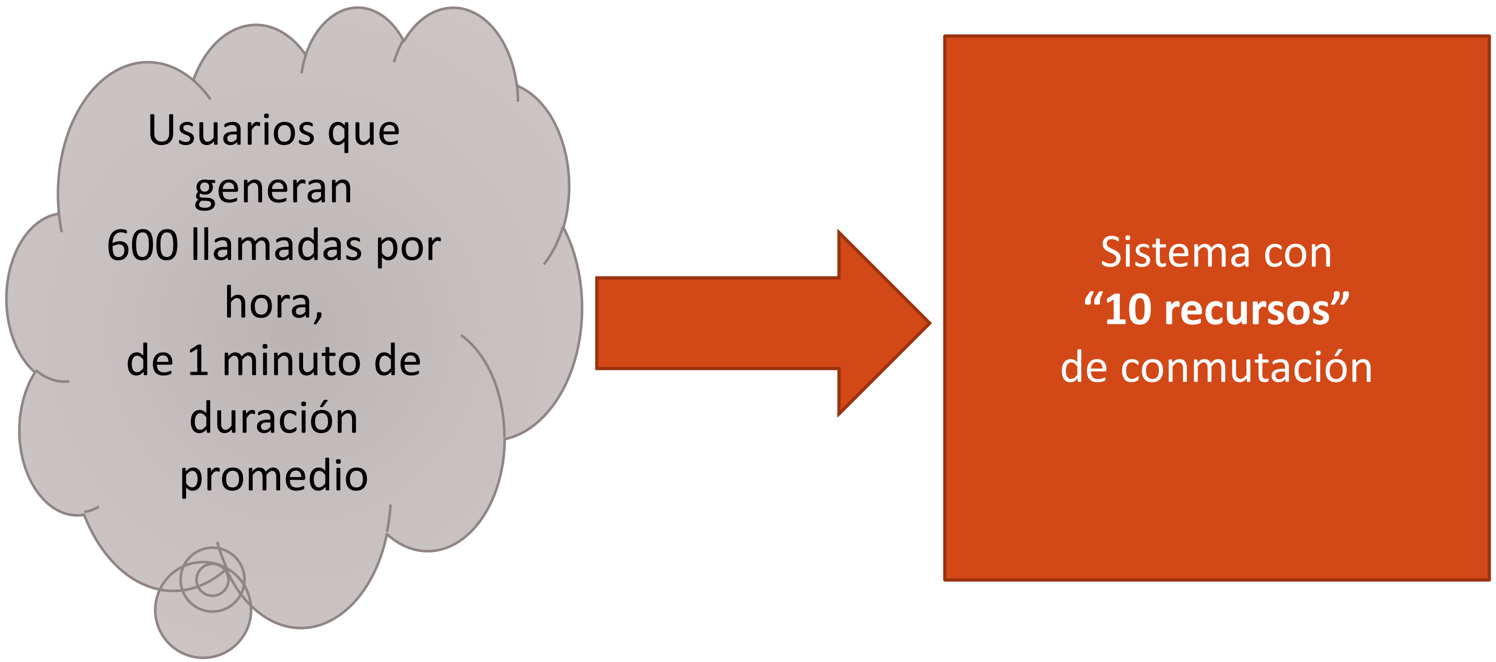
Hay "congestión" cuando todos los recursos están ocupados, o sea cuando el sistema está en el estado $[n]$

Por tanto, la probabilidad de que exista congestión es

$$p(n) = \frac{\dfrac{A^n}{n!}}{\displaystyle\sum_{j=0}^{n} \frac{A^j}{j!}} = E_B(n, A)$$

Conocida como **Fórmula de Erlang-B** (publicada en 1917)

Ejemplo



¿Qué probabilidad hay de que exista congestión?

Ejemplo

λ=1/6 seg^{-1}

μ=1/60 seg^{-1}

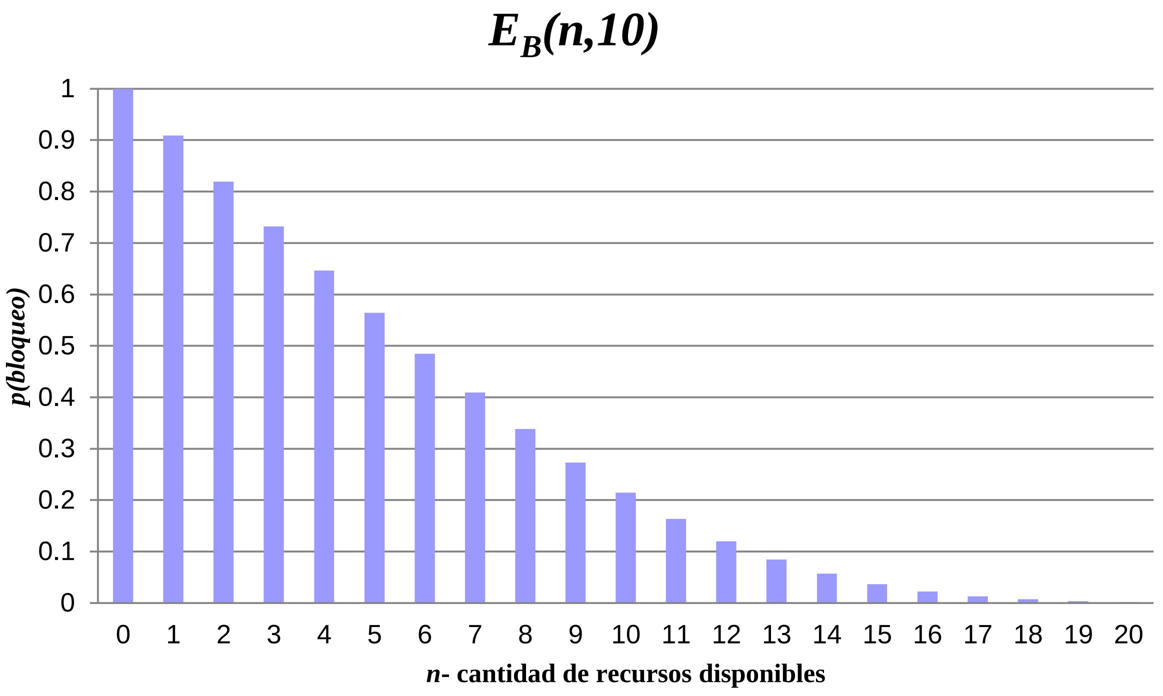
$A = \lambda/\mu =$ 10 Erlang

$$E_B(10,10) = \frac{10^{10}}{10!} \frac{1}{\sum_{j=0}^{10} \frac{10^j}{j!}} = 0.21 = 21\%$$

Ejemplo

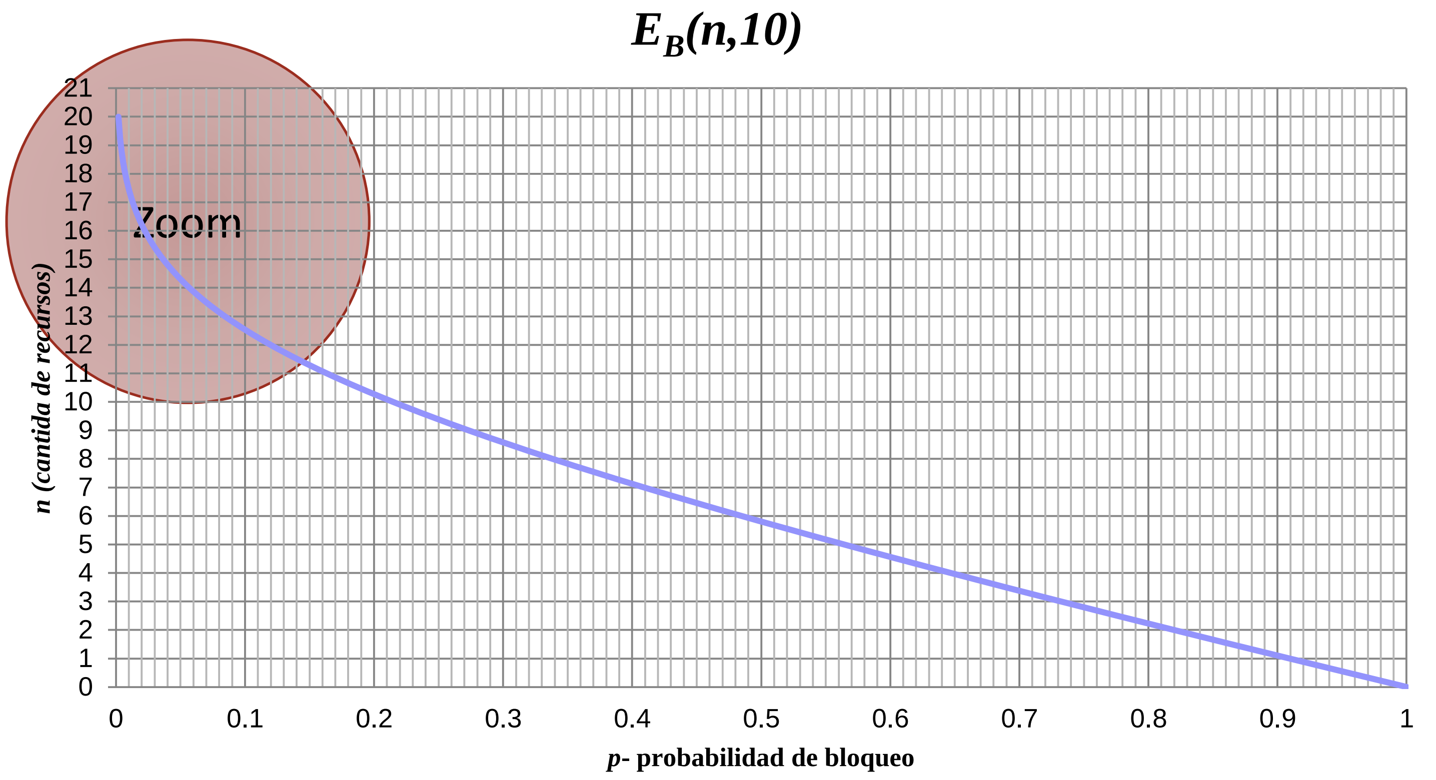
¿Cómo varía la probabilidad de bloqueo según la cantidad de recursos disponibles?

$$E_B(n,A) = \frac{A^n}{n!} \frac{1}{\sum_{j=0}^{n} \frac{A^j}{j!}}$$



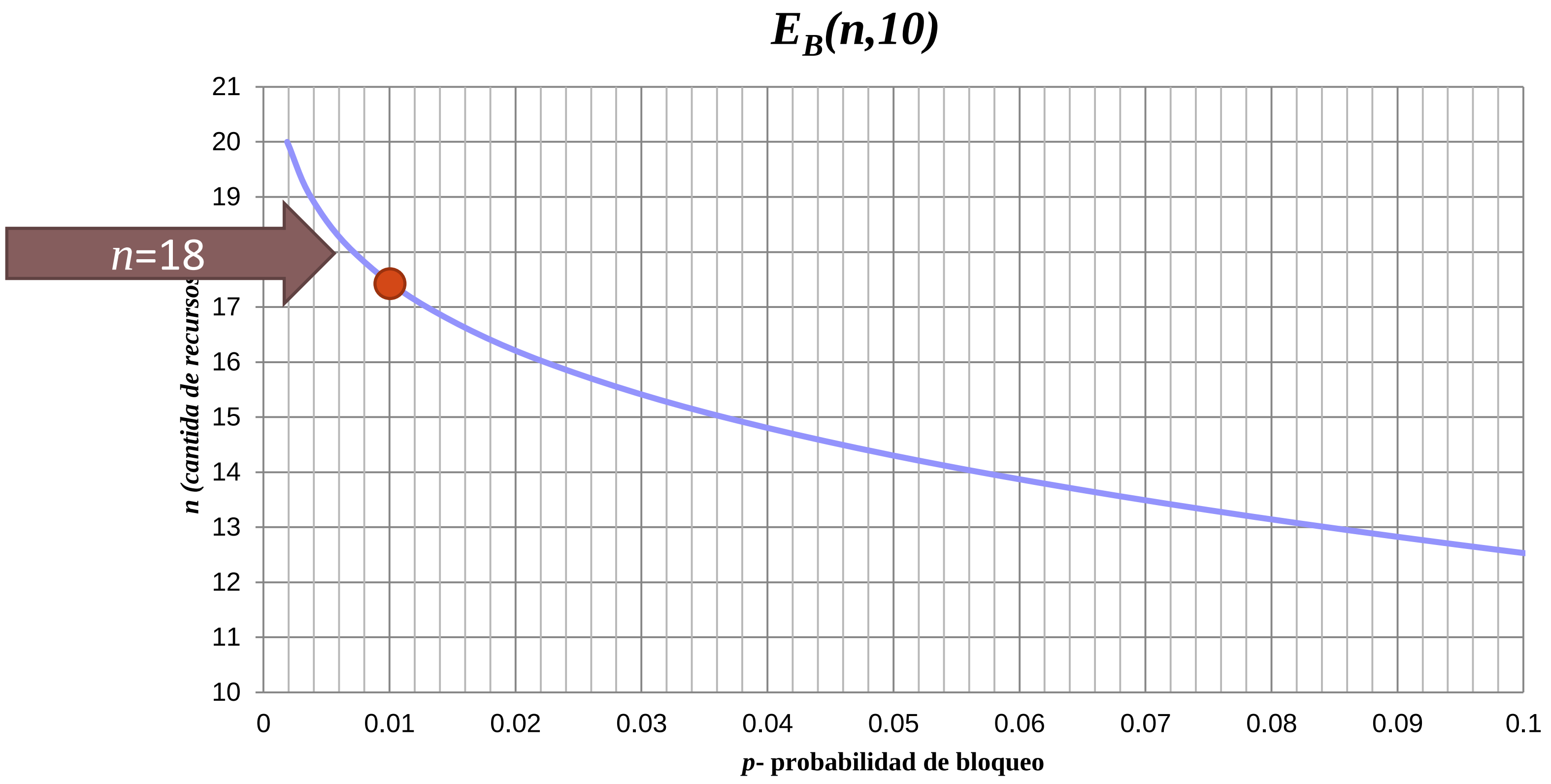
Ejemplo

Si queremos que la probabilidad de bloqueo sea menor al 1%, ¿cuántos recursos necesitamos?



Ejemplo

Si queremos que la probabilidad de bloqueo sea menor al 1%, ¿cuántos recursos necesitamos?



Tablas de Erlang-B

Erlang B Traffic Table

Maximum Offered Load Versus B and N

B is in %

Probabilidad de bloqueo= *En(A)*

N/B	0.01	0.05	0.1	0.5						20
1	.0001	.0005	.0010	.0050	.0101	.0204	.0526	.1111	.1765	.2500
2	.0142	.0321	.0458	.1054	.1526	.2235	.3813	.5954	.7962	1.000
3	.0868	.1517	.1938	.3490	.4555	.6022	.8994	1.271	1.603	1.930
4	.2347	.3624	.4393	.7012	.8694	1.092	1.525	2.045	2.501	2.945
5	.4520	.6486	.7621	1.132	1.361	1.657	2.219	2.881	3.454	4.010
	.7282	.9957	1.146	1.622	1.909	2.276	2.960	3.758	4.445	5.109
	1.054	1.392	1.579	2.158	2.501	2.935	3.738	4.666	5.461	6.230
	1.422	1.830	2.051	2.730	3.128	3.627	4.543	5.597	6.498	7.369
	1.826	2.302	2.558	3.333	3.783	4.345	5.370	6.546	7.551	8.522
10	2.260	2.803	3.092	3.961	4.461	5.084	6.216	7.511	8.616	9.685

n

Tráfico (*A*)

Características del tráfico con modelo de pérdida

Hay "congestión" cuando todos los recursos están ocupados

El tráfico cursado Y es menor al tráfico ofrecido A

$$Y = \sum_{i=1}^{n} ip(i) = \sum_{i=1}^{n} Ap(i-1) = A\big(1 - p(n)\big)$$

$$Y = A\big(1 - E_B(n, A)\big)$$

(usando las ecuaciones de corte)

El tráfico perdido es

$$A_{lost} = A - Y = A - A\big(1 - E_B(n, A)\big) = AE_B(n, A)$$

Resumen de las Hipótesis utilizadas para Erlang-B

Chance pura
- El arribo de llamadas se modela según un proceso de Poisson
- El tiempo entre arribos de llamadas tiene una distribución exponencial
- Existen “infinitas” fuentes generadoras de tráfico

Duración de las llamadas
- La duración de las llamadas tiene una distribución exponencial

Grupo homogéneo de recursos
- Los recursos son idénticos, trabajando en paralelo

Accesibilidad completa
- Una llamada es aceptada en el sistema si existe por lo menos un recurso disponible, y cada llamada ocupa un único recurso

Sistema de pérdida
- Si un intento de llamada no encuentra un recurso libre, se pierde
- No hay “reintentos”

Generalización

Duración de las llamadas

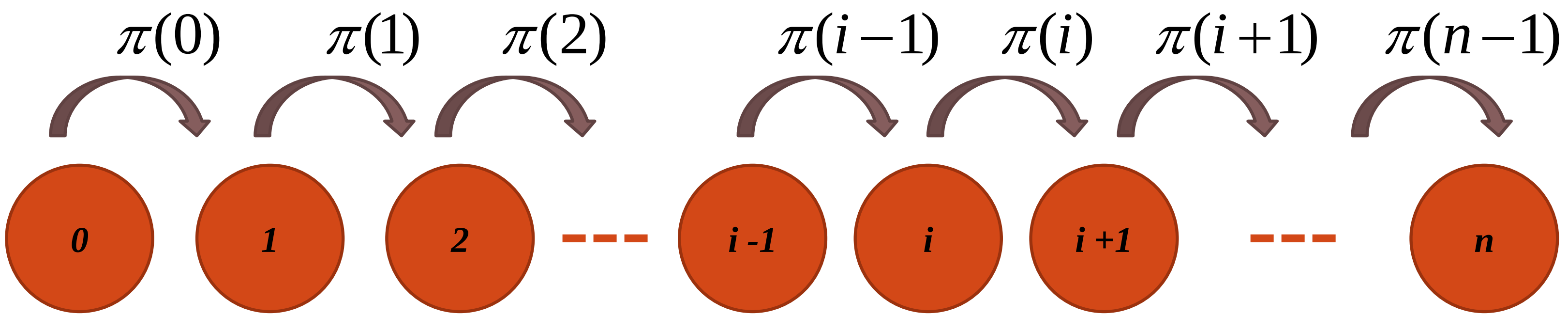
- La fórmula es valida para cualquier distribución de duración de llamadas, y solo depende de la duración media ***d***
- De hecho, solo depende del tráfico

$$A = \lambda d = \frac{\lambda}{\mu}$$

Fuentes finitas

¿Qué sucede si hay un número “finito” de fuentes generadoras de tráfico?

- A medida que las “fuentes” obtienen un recurso, quedan menos “fuentes” para generar tráfico
- Por lo tanto, la probabilidad de transición del estado [i] al [$i+1$] dependerá del estado [i]



Fuentes finitas

Interesa en este caso la tasas de arribos por cada "fuente"

- En promedio γ arribos por unidad de tiempo por cada "fuente"
- Si hay S fuentes, la tasa total de arribos es

$$\lambda = S\gamma$$

- Para el caso de ∞ fuentes aplica

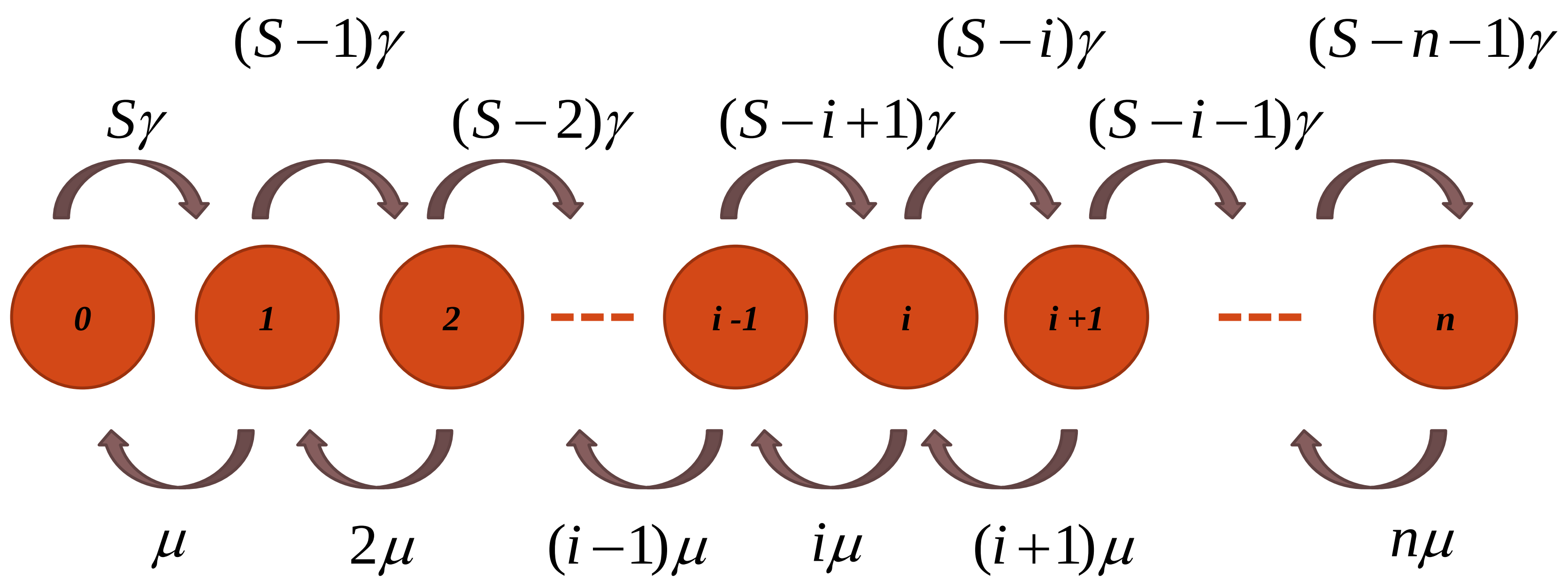
$$\lambda = \lim_{S \to \infty, \gamma \to 0} (S\gamma)$$

Fuentes finitas

El tráfico por cada "fuente" se puede definir como

$$a = \frac{\gamma}{\mu}$$

Diagrama de transición de estados



Transiciones por unidad de tiempo

Deducción de las probabilidades de estados

Aplicamos las ecuaciones de "corte"

$$S\gamma p(0) = \mu p(1)$$

$$(S-1)\gamma p(1) = 2\mu p(2)$$

$$(S-2)\gamma p(2) = 3\mu p(3)$$

....

$$(S-i+2)\gamma p(i-2) = (i-1)\mu p(i-1)$$

$$(S-i+1)\gamma p(i-1) = (i)\mu p(i)$$

$$(S-i)\gamma p(i) = (i+1)\mu p(i+1)$$

....

$$a = \gamma d = \frac{\gamma}{\mu}$$

Deducción de las probabilidades de estados

$$p(1) = Sap(0)$$

$$p(2) = \frac{(S-1)a}{2} p(1) = \frac{S(S-1)a^2}{2} p(0)$$

$$p(3) = \frac{(S-2)a}{3} p(2) = \frac{S(S-1)(S-2)a^3}{3x2} p(0)$$

....

$$p(i-1) = \frac{(S-i+2)a}{i-1} p(i-2) = \frac{S(S-1)(S-2)....(S-i+2)a^{i-1}}{(i-1)!} p(0)$$

$$p(i) = \frac{(S-i+1)a}{i} p(i-1) = \frac{S(S-1)(S-2)....(S-i+1)a^{i}}{i!} p(0)$$

....

Deducción de las probabilidades de estados

$$p(i)=\frac{S(S-1)(S-2)....(S-i+1)a^{i}}{i!}\,p(0)$$

$$p(i)=\frac{S!a^{i}}{(S-i)!i!}\,p(0)$$

$$\frac{S!}{(S-i)!i!}=C_{i}^{s}$$

Combinaciones de S tomadas de i

$$\boxed{p(i)=C_{i}^{s}a^{i}p(0)}$$

Deducción de las probabilidades de estados

Normalización

$$\sum_{i=1}^{n} p(i) = 1$$

$$p(i) = C_i^s a^i p(0)$$

$$\sum_{i=1}^{n} C_i^s a^i p(0) = 1$$

$$p(0) = \frac{1}{\sum_{i=1}^{n} C_i^s a^i}$$

Congestión

Hay "congestión" cuando todos los recursos están ocupados, o sea cuando el sistema está en el estado [n]

Por tanto, la probabilidad de que exista congestión es

$$p(n) = \frac{C_n^S a^n}{\sum_{i=0}^{n} C_i^S a^i} = E_{Engset}(a, n, S)$$

Tore Olaus Engset
1865-1943

Conocida como **Fórmula de Engset** (publicada en 1918)

Características del tráfico de pérdida y fuentes finitas

Hay “congestión” cuando todos los recursos están ocupados

El tráfico cursado Y es

$$Y = \sum_{i=1}^{n} ip(i) = \sum_{i=1}^{n} a(S-i+1)\,p(i-1) = \sum_{i=0}^{n-1} a(S-i)\,p(i)$$

$$Y = \sum_{i=0}^{n} a(S-i)\,p(i) - a(S-n)\,p(n) = a\sum_{i=0}^{n} Sp(i) - a\sum_{i=0}^{n} ip(i) - a(S-n)E$$

$$Y = aS - aY - a(S-n)E$$

$$Y = \frac{a}{1+a}\left(S-(S-n)E\right)$$

Resumen de las Hipótesis utilizadas para Engset

Chance pura

- El arribo de llamadas se modela según un proceso de Poisson
- El tiempo entre arribos de llamadas tiene una distribución exponencial
- Existe un número finito S de “fuentes generadoras” de tráfico, que es mayor a la cantidad de recursos n ($S>n$)

Duración de las llamadas

- La duración de las llamadas tiene una distribución exponencial

Grupo homogéneo de recursos

- Los recursos son idénticos, trabajando en paralelo

Accesibilidad completa

- Una llamada es aceptada en el sistema si existe por lo menos un recurso disponible, y cada llamada ocupa un único recurso

Sistema de pérdida

- Si un intento de llamada no encuentra un recurso libre, se pierde
- No hay “reintentos”

Recursos finitos: Modelos de demora

TEORÍA E INGENIERÍA DE TELETRÁFICO

Consideraciones

Consideramos un sistema con n recursos idénticos, trabajando en paralelo

- "Grupo homogéneo" (*homogeneous group*)

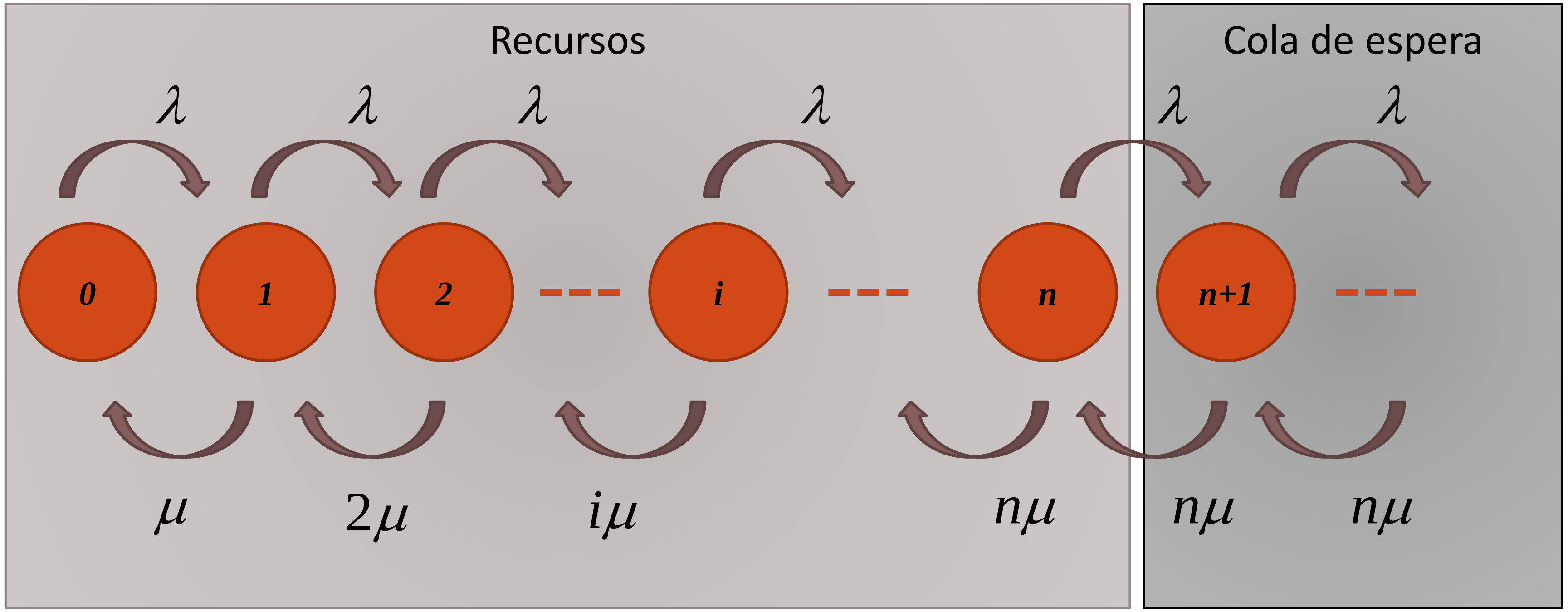
Una llamada es aceptada en el sistema si existe por lo menos un recurso disponible, y cada llamada ocupa un único recurso

- "Accesibilidad completa" (*full accessibility*)

Si todos los recursos están ocupados el sistema está "congestionado" y la llamada se pone en "cola de espera"

- La "cola de espera" no tiene límites
  - Pueden existir ∞ llamadas en espera
- Cuando una llamada ingresa a la "cola de espera", se mantiene hasta que llega su turno
  - NO hay "abandonos"
- **Modelo de demora (*Delay Systems*)**

Diagrama de transición de estados



Transiciones por unidad de tiempo

Deducción de las probabilidades de estados

Aplicamos las ecuaciones de “corte”

$$\lambda p(0) = \mu p(1)$$

$$....$$

$$\lambda p(i) = (i+1)\mu p(i+1)$$

$$....$$

$$\lambda p(n-1) = n\mu p(n)$$

$$\lambda p(n) = n\mu p(n+1)$$

$$....$$

$$\lambda p(n+j) = n\mu p(n+j+1)$$

$$A = \lambda d = \frac{\lambda}{\mu}$$

Deducción de las probabilidades de estados

$i \le n$

$$p(1) = Ap(0)$$

$$p(2) = \frac{A}{2} p(1) = \frac{A^2}{2} p(0)$$

....

$$p(i-1) = \frac{A}{i-1} p(i-2) = \frac{A^{i-1}}{(i-1)!} p(0)$$

$$p(i) = \frac{A}{i} p(i-1) = \frac{A^i}{i!} p(0)$$

....

$i > n$

$$p(i) = \frac{A}{n} p(i-1) = \frac{A^{i-n}}{n^{i-n}} p(n)$$

$$p(i) = \left(\frac{A}{n}\right)^{i-n} \frac{A^n}{n!} p(0)$$

$$p(i) = \frac{A^i}{n^{i-n} n!} p(0)$$

Deducción de las probabilidades de estados

$$\sum_{i=0}^{\infty} p(i) = 1$$

$$\sum_{i=0}^{n-1} \frac{A^i}{i!} p(0) + \sum_{i=n}^{\infty} \frac{A^i}{n^{i-n} n!} p(0) = 1$$

$$p(0)\left(\sum_{i=0}^{n-1} \frac{A^i}{i!} + \frac{A^n}{n!} \sum_{j=0}^{\infty} \frac{A^j}{n^j}\right) = 1$$

$$\sum_{j=0}^{\infty} \frac{A^j}{n^j} = \frac{1}{1-\frac{A}{n}}, A < n$$

$$\boxed{p(0) = \frac{1}{\sum_{i=0}^{n-1} \frac{A^i}{i!} + \frac{A^n}{n!} \frac{n}{n-A}}, A < n}$$

Probabilidad de que exista demora

La probabilidad de que una llamada ingrese a la cola de espera (es decir, que no pueda ser atendida inmediatamente) es la probabilidad de que el sistema se encuentra en cualquiera de los estados [i] mayores o iguales a [n]

$$p(Wait > 0) = \sum_{i=n}^{\infty} p(i)$$

Probabilidad de que exista demora

$$p(Wait > 0) = \sum_{i=n}^{\infty} p(i)$$

$$p(Wait > 0) = \sum_{i=n}^{\infty} \frac{A^i}{n^{i-n} n!} p(0) = \frac{A^n}{n!} \sum_{j=0}^{\infty} \frac{A^j}{n^j} p(0) = \frac{A^n}{n!} \frac{n}{n-A} p(0)$$

$$p(Wait > 0) = \frac{\frac{A^n}{n!} \frac{n}{n-A}}{\sum_{i=0}^{n-1} \frac{A^i}{i!} + \frac{A^n}{n!} \frac{n}{n-A}} = E_C(n, A), \qquad A < n$$

Conocida como **Fórmula de Erlang-C**

Probabilidad de que existan llamadas en espera

La probabilidad de que existan llamadas en espera es la probabilidad de que el sistema se encuentra en cualquiera de los estados [i] mayores estrictos a [n]

$$p(L>0)=\sum_{i=n+1}^{\infty} p(i)=\frac{A}{n}E_C(n,A)$$

Número promedio de llamadas en espera

$$\overline{L} = 1p(n+1) + 2p(n+2) + 3p(n+3) + ... = \sum_{k=1}^{\infty} kp(n+k)$$

$$p(i) = \frac{A^i}{n^{i-n} n!} p(0)$$

$$\overline{L} = \sum_{k=1}^{\infty} k \frac{A^n A^k}{n! n^k} p(0) = p(0) \frac{A^n}{n!} \sum_{k=1}^{\infty} k \left(\frac{A}{n} \right)^k = p(n) \sum_{k=1}^{\infty} k \left(\frac{A}{n} \right)^k$$

$$k(A/n)^k = (A/n) \frac{\partial}{\partial (A/n)} (A/n)^k$$

$$\overline{L} = p(n) \frac{A}{n} \sum_{k=1}^{\infty} \frac{\partial}{\partial (A/n)} (A/n)^k = p(n) \frac{A}{n} \frac{\partial}{\partial (A/n)} \sum_{k=1}^{\infty} (A/n)^k$$

$$\overline{L} = p(n) \frac{A}{n} \frac{\partial}{\partial (A/n)} \left\{ \frac{(A/n)}{1-(A/n)} \right\} = p(n) \frac{(A/n)}{(1-(A/n))^2}$$

$$\overline{L} = p(n) \frac{n}{n-A} \frac{A}{n-A} = \frac{A}{n-A} E_C(n, A)$$

Número promedio de llamadas en espera cuando hay cola

¿Cuántas llamadas en promedio habrá en espera, si sabemos que existe cola de espera?

$$\bar{L}\Big|_{L>0} = \frac{\sum_{k=1}^{\infty} kp(n+k)}{\sum_{k=n+1}^{\infty} p(k)} = \frac{p(n)\dfrac{A/n}{(1-A/n)^2}}{p(n)\dfrac{A}{n-A}}$$

$$\bar{L}\Big|_{L>0} = \frac{n}{n-A}$$

Tiempo promedio de espera

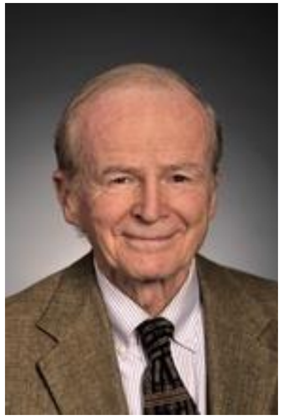
¿Cuánto tiempo en promedio deberá esperar una llamada hasta ser atendida?

Teorema de Little (John Little, 1961):

- La cantidad promedio de llamadas en espera $\boldsymbol{L}$ es igual a la tasa de arribos $\boldsymbol{\lambda}$ multiplicada por la demora media $\boldsymbol{W}$

$$L = \lambda W$$

- Es válido para cualquier sistema de encolamiento, sin importar la distribución de arribos ni la distribución de la duración del servicios o llamadas



John Dutton Little
1928-

Tiempo promedio de espera

$$\overline{W} = \frac{\overline{L}}{\lambda} = \frac{1}{\lambda}\frac{A}{n-A}E_C(n,A)$$

$$A = \frac{\lambda}{\mu} = \lambda d$$

$$\overline{W} = \frac{d}{n-A}E_C(n,A)$$

Esta es la demora promedio para TODAS las llamadas

- Algunas tuvieron demora, otras fueron atendidas sin demora

Tiempo promedio de espera para las llamadas en cola

Si una llamada es encolada, ¿cuánto tiempo en promedio deberá esperar para ser atendida?

$$\overline{W}\Big|_{W>0} = \frac{\overline{W}}{p(Wait > 0)} = \frac{d}{n - A}$$

Generalizaciones

TEORÍA E INGENIERÍA DE TELETRÁFICO

Notación de Kendall

David George Kendall fue un matemático especializado en estadística

En 1953 Propuso una notación para describir modelos de encolamiento generales, según

- La distribución del proceso de arribos
- La distribución de la duración del servicio
- El número de "recursos"

David George Kendall
1918-2007

Notación de Kendall

Notación: A/B/n

A = Proceso de Arribos

B = Distribución del tiempo de servicio

n = número de recursos

Los valores de A y B pueden ser:

- M = Proceso “Markoviano” (Poisson, distribución exponencial)
- D = Determinística
- G = General (Distribución arbitraria)
- Otros…

Notación de Kendall - Ejemplo

M/M/n

- Sistema de “chance pura” con proceso de arribo de Poisson, tiempos de servicios con distribución exponencial y un número n finito de recursos

M/M/∞

- Sistema de “chance pura” con proceso de arribo de Poisson, tiempos de servicios con distribución exponencial y un número infinito de recursos

Notación de Kendall – Extensión

A/B/n/K/S/X

K = Capacidad total del sistema (K – n = número de posiciones para la cola de espera)

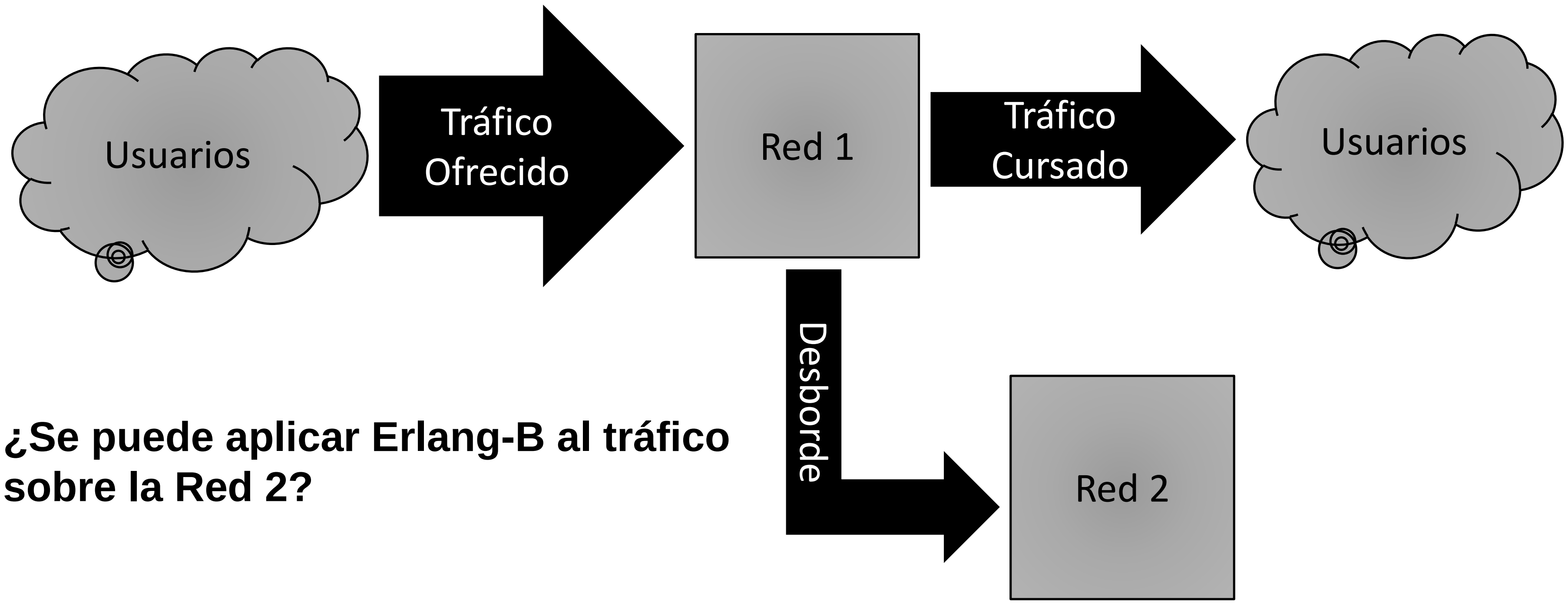
S = Número de “fuentes” generadoras de tráfico

X = Comportamiento de la cola

- FIFO,LIFO,…

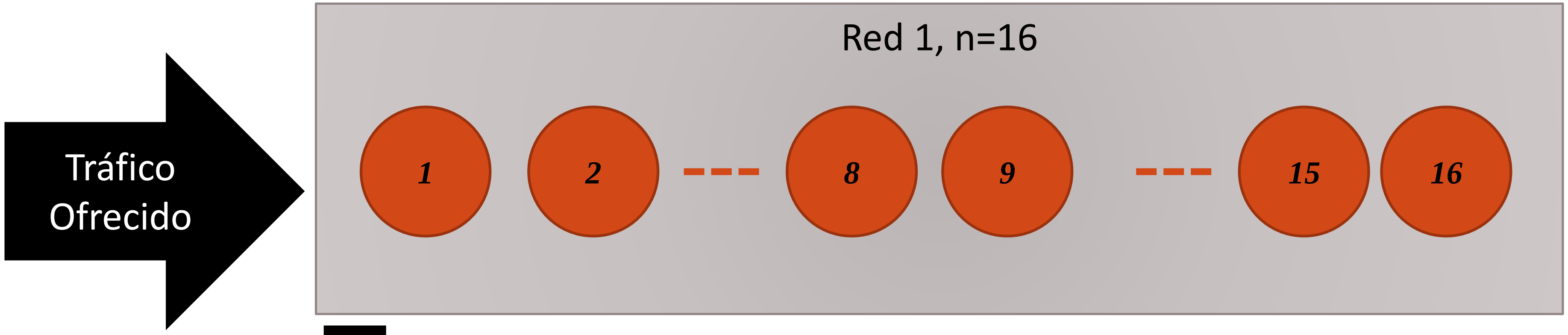
Tráfico de desborde

Es el tráfico que no pudo ser cursado por una red y es derivado a otra red

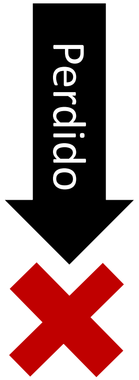


¿Se puede aplicar Erlang-B al tráfico sobre la Red 2?

Ejemplo



A=10E



2.2%

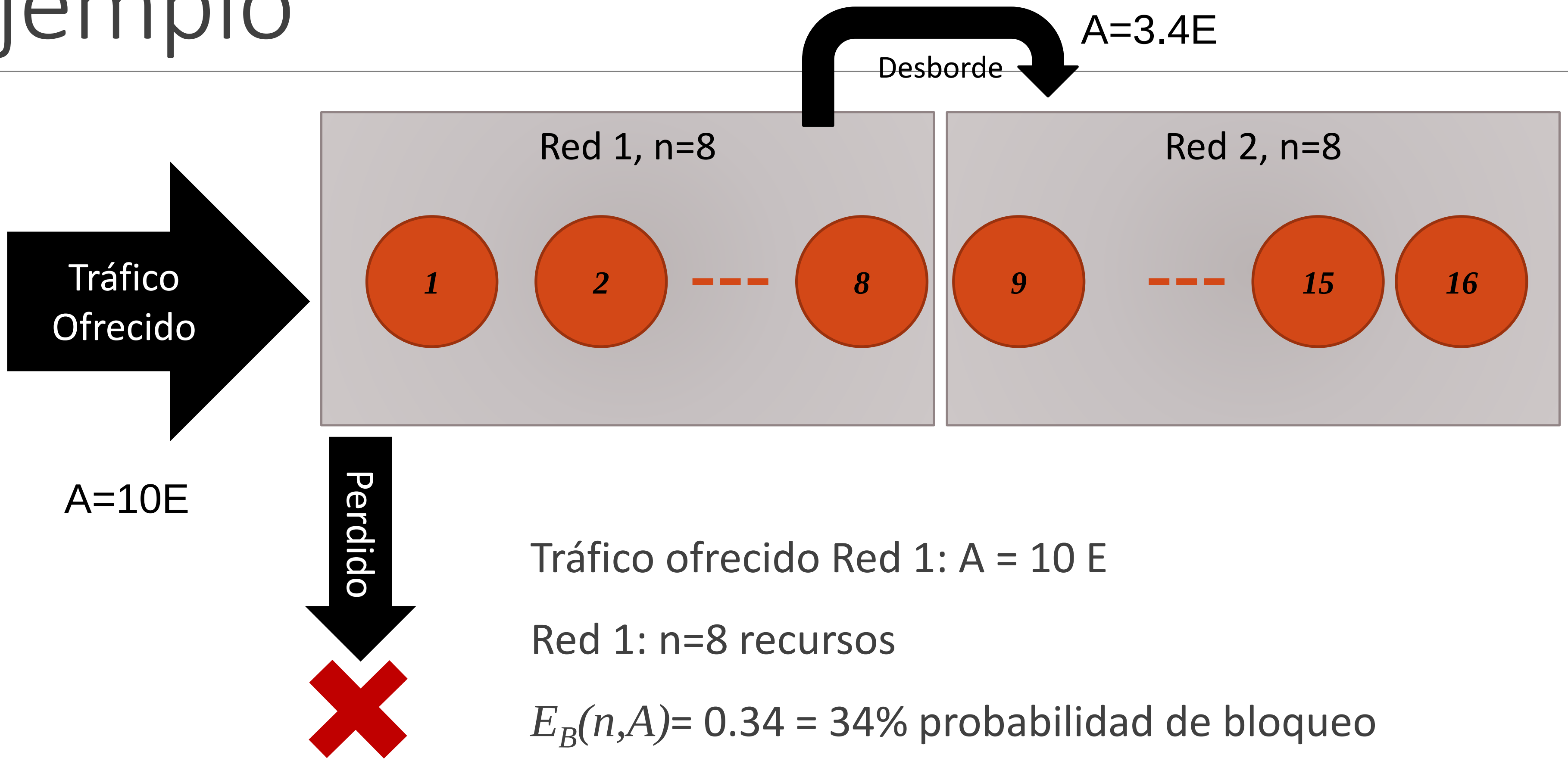
Tráfico ofrecido: A = 10 Erlang

Red 1: n=16 recursos

$E_B(n,A)$=0.022 = 2.2% de probabilidad de bloqueo

Tráfico total perdido=10 E x 0.022 = 0.22 E = 2.2% del tráfico total ofrecido

Ejemplo



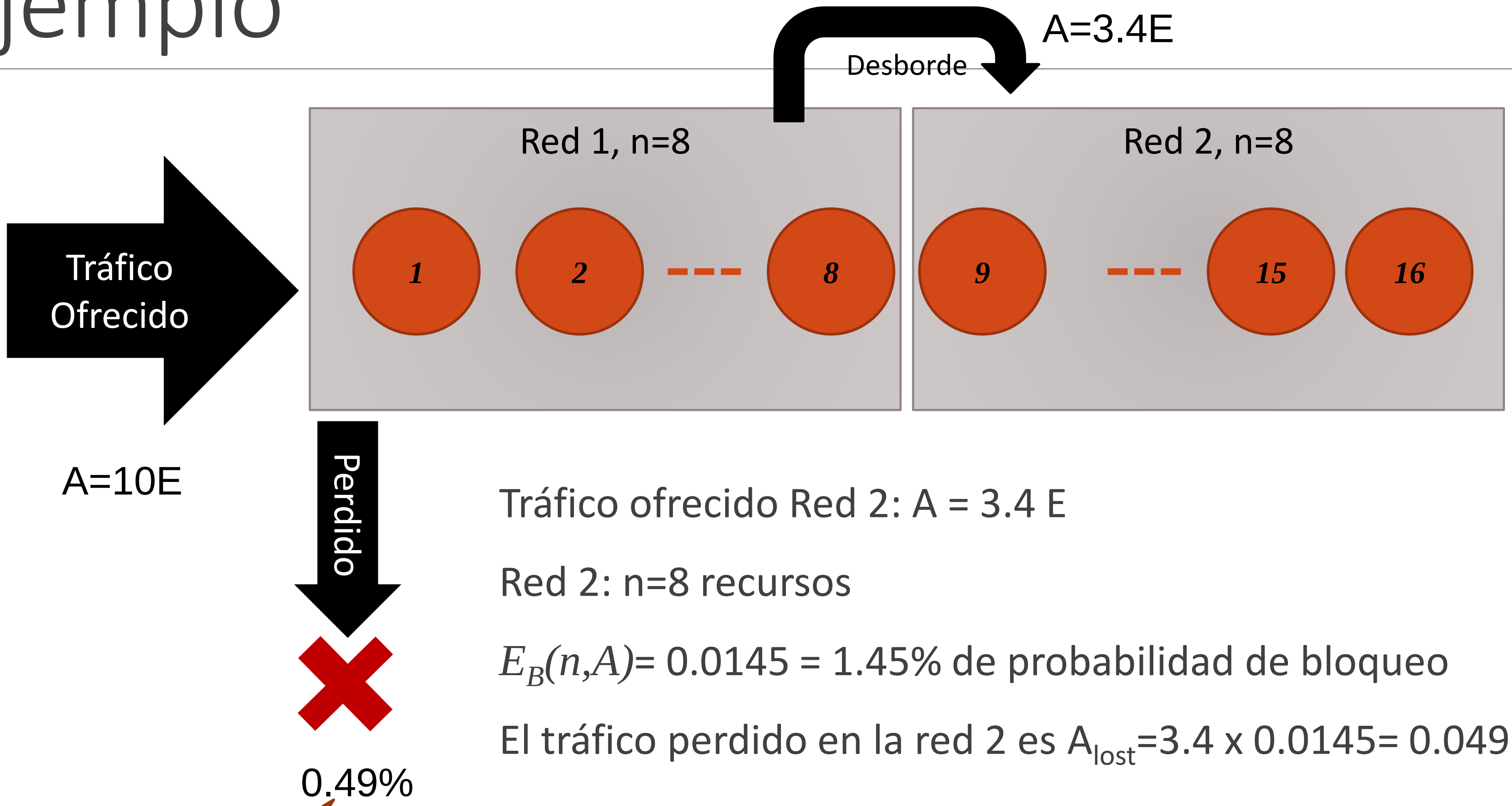
Tráfico ofrecido Red 1: A = 10 E

Red 1: n=8 recursos

$E_B(n,A)$= 0.34 = 34% probabilidad de bloqueo

Tráfico perdido Red 1=10 x 0.34 = 3.4 E = Trafico ofrecido a Red 2

Ejemplo



Tráfico ofrecido Red 2: A = 3.4 E

Red 2: n=8 recursos

$E_B(n,A)$= 0.0145 = 1.45% de probabilidad de bloqueo

El tráfico perdido en la red 2 es A_{lost}=3.4 x 0.0145= 0.049 E

= 0.49% del tráfico total ofrecido

Pero el valor real es 2.2%!!

Tráfico de desborde

El tráfico de desborde no cumple las hipótesis de Erlang-B

Es un tráfico que presenta características de “ráfagas”, en los momentos en que la Red anterior está completa.

Fue estudiado por Roger I. Wilkinson (en 1956) y por G. Bretschneider (en el mismo año)

Ejemplos

TEORÍA E INGENIERÍA DE TELETRÁFICO

Ejemplo 1

Una Empresa desea incorporar un sistema de “correo de voz” en su red, para todos sus usuarios

Se sabe que

- Hay 1000 usuarios que utilizarán el correo de voz
- Cada comunicación con el correo de voz tiene una duración media de 2 minutos
- Por cada usuario, se espera que el correo de voz atienda en promedio 1 llamada en la hora pico
- Se acepta que de 100 intentos, 1 no consiga conectarse

¿Cuántos “canales” se requieren en el “correo de voz”?

Ejemplo 1

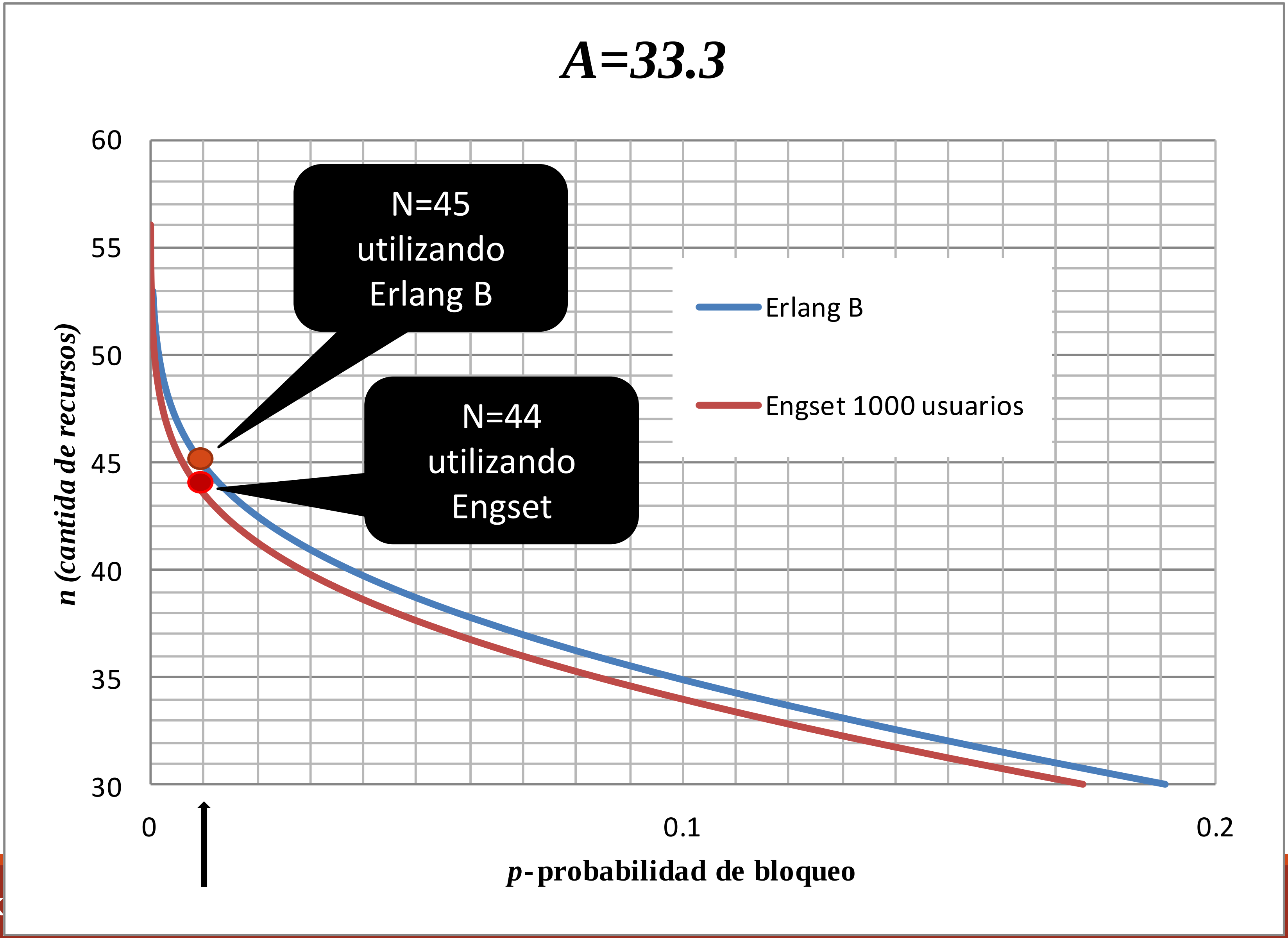
Habrán 1.000 llamadas por hora, de 2 minutos de duración

$A = \lambda d$ = 1000 . 120/3600 = 33.3 E

¿Qué modelo aplicamos?

- Erlang-B
- Engset
- Erlang-C
- …

Ejemplo 1



A=33.3
60
55
50
45
40
35
30
n (cantida de recursos)
0
0.1
0.2
p-probabilidad de bloqueo
N=45 utilizando Erlang B
N=44 utilizando Engset
Erlang B
Engset 1000 usuarios

Ejemplo 1

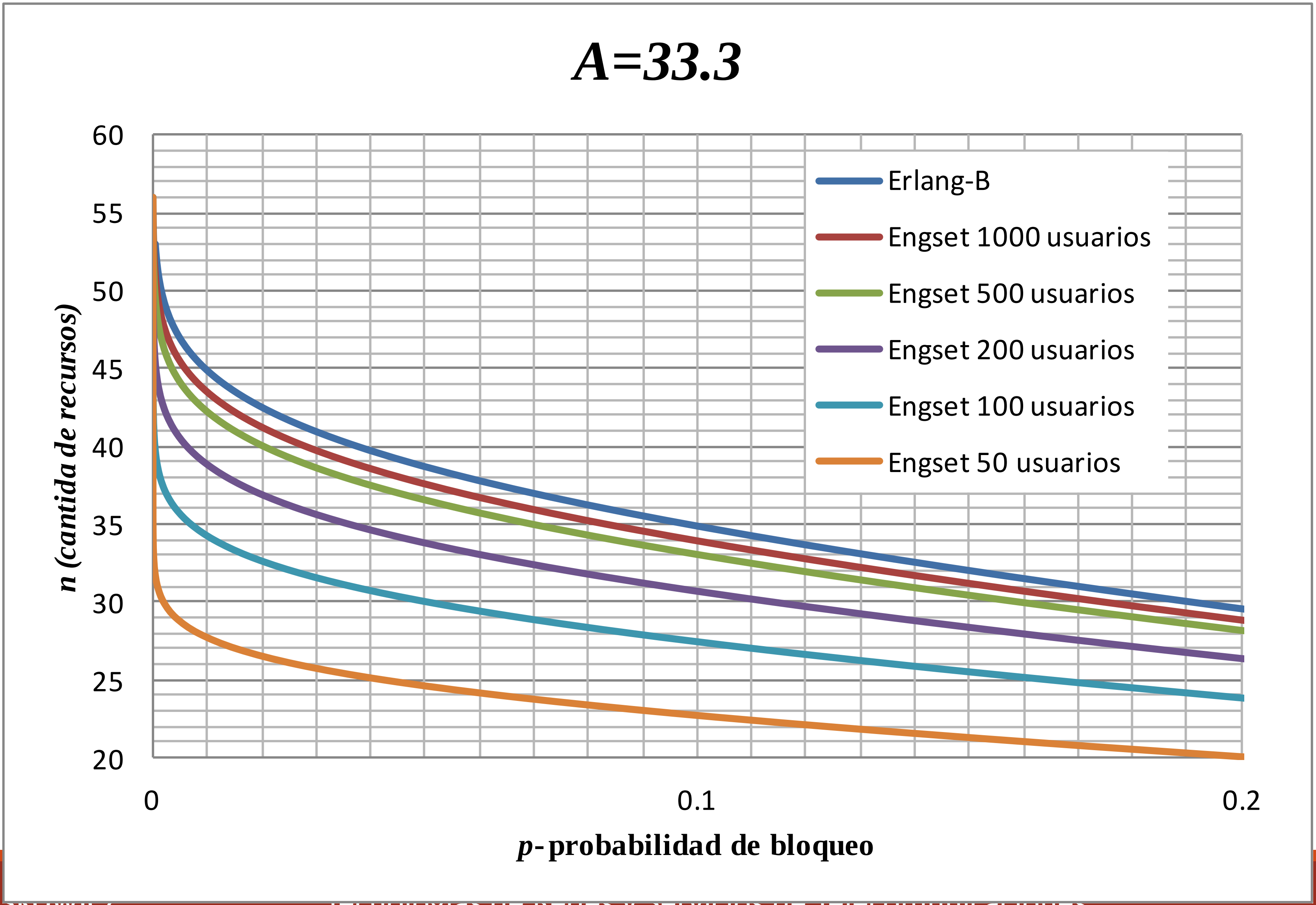
Utilizando las tablas de Erlang-B

Erlang B Traffic Table

Maximum Offered Load Versus B and N
B is in %

N/B	0.01	0.05	0.1	0.5	1.0	2	5	10	15	20
1	.0001	.0005	.0010	.0050	.0101	.0204	.0526	.1111	.1765	.2500
2	.0142	.0321	.0458	.1054	.1526	.2235	.3813	.5954	.7962	1.000
3	.0868	.1517	.1938	.3490	.4555	.6022	.8994	1.271	1.603	1.930
4	.2347	.3624	.4393	.7012	.8694	1.092	1.525	2.045	2.501	2.945
5	.4520	.6486	.7621	1.132	1.361	1.657	2.219	2.881	3.454	4.010
44	24.33	26.53	27.64	30.80	32.54	34.68	38.56	43.09	47.09	51.09
45	[illegible]	[illegible]	[illegible]	[illegible]	33.43	35.61	39.55	44.17	48.25	52.32

Engset vs Erlang-B



Ejemplo 2

Una Empresa desea incorporar un sistema de “correo de voz” en su red, para todos sus usuarios

Se sabe que

- Hay 1000 usuarios que utilizarán el correo de voz
- Cada comunicación con el correo de voz tiene una duración media de 2 minutos
- Por cada usuario, se espera que el correo de voz atienda en promedio 1 llamada en la hora pico
- Se acepta que de 100 intentos, 1 se vea demorada hasta ser atendida

¿Cuántos “canales” se requieren en el “correo de voz”?

Ejemplo 2

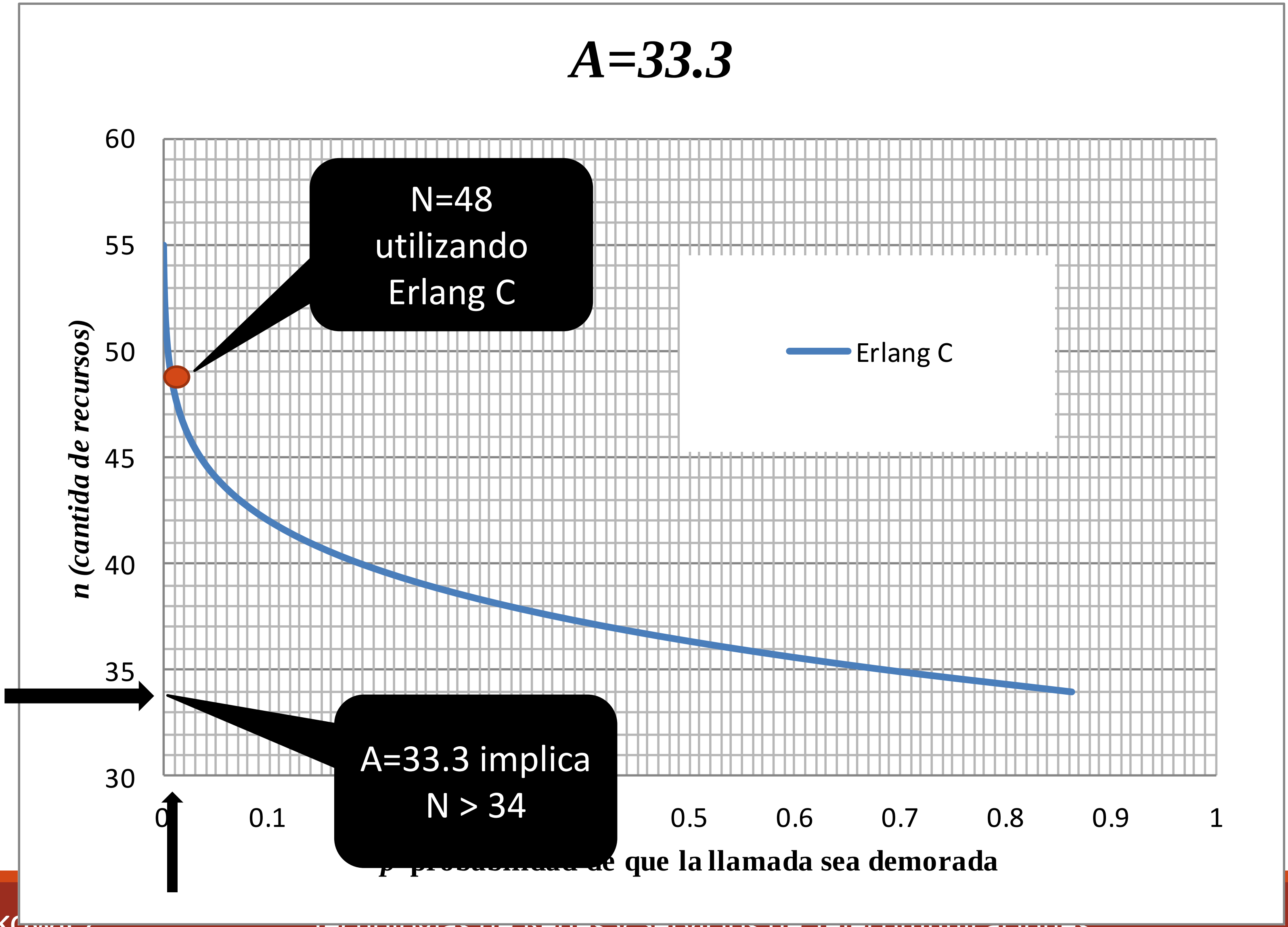
Habrán 1.000 llamadas por hora, de 2 minutos de duración

$A=\lambda d = 1000 \cdot 120/3600 = 33.3$ E

¿Qué modelo aplicamos?

- Erlang-B
- Engset
- Erlang-C
- …

Ejemplo 2


A=33.3
60
55
50
45
40
35
30
n (cantida de recursos)
N=48
utilizando
Erlang C
Erlang C
A=33.3 implica
N > 34
0
0.1
0.5
0.6
0.7
0.8
0.9
1
que la llamada sea demorada

Ejemplo 2

Utilizando las tablas de Erlang-C

Erlang C Traffic Table

Maximum Offered Load Versus B and N

B is in %

N/B	0.01	0.05	0.1	0.5	1.0	2	5	10	15	20
1	.0001	.0005	.0010	.0050	.0100	.0200	.0500	.1000	.1500	.2000
2	.0142	.0319	.0452	.1025	.1465	.2103	.3422	.5000	.6278	.7403
3	.0860	.1490	.1894	.3339	.4291	.5545	.7876	1.040	1.231	1.393
4	.2310	.3533	.4257	.6641	.8100	.9939	1.319	1.653	1.899	2.102
5	.4428	.6289	.7342	1.065	1.259	1.497	1.905	2.313	2.607	2.847
46	24.88	26.82	27.75	30.24	31.49	32.87	34.97	36.83	38.05	39.00
47	25.60	27.57	28.52	31.05	32.32	33.72	35.84	37.72	38.96	39.92
48	26.34	28.33	29.30	31.86	33.14	34.56	36.72	38.62	39.87	40.84
49	27.07	29.10	30.08	32.68	33.97	35.41	37.59	39.52	40.79	41.76
50	27.80	29.86	30.86	33.49	34.80	36.26	38.47	40.42	41.70	4?,69

Ejemplo 2

¿Cuál será la demora promedio?

$$\overline{W} = \frac{d}{n-A} E_C(n,A) = \frac{120s}{48-33.33} 0.01 = 0.08s$$

Si una llamada es demorada, ¿Cuál será su demora esperada?

$$\overline{W}\Big|_{W>0} = \frac{\overline{W}}{p(Wait>0)} = \frac{d}{n-A} = \frac{120}{48-33.3} = 8.2s$$

Muchas Gracias!