

Conceptos Fundamentales

De TETRAPRO, editado por el Sr. H. Leijon, UIT



UNION INTERNATIONALE DES TELECOMMUNICATIONS
INTERNATIONAL TELECOMMUNICATION UNION
UNION INTERNACIONAL DE TELECOMUNICACIONES



Teoría Básica de Teletráfico (T)

CONCEPTOS FUNDAMENTALES (TFC)

Contenido

1. Unidad de Tráfico

Intensidad de llamadas
Tiempo de ocupación
Número de ocupaciones simultáneas
Unidades de Tráfico
Volumen de Tráfico

2. Generación de Tráfico

Posible resultado de un intento de llamada
Intentos Repetidos
Duración de las diferentes llamadas
Demanda de tráfico, tráfico ofrecido y cursado, tráfico facturado

3. Variaciones de Tráfico

Durante el día, la semana y la temporada
Crecimiento de tráfico
Picos impredecibles
Variaciones dentro de intervalos cortos
Concepto de hora pico
Resultado a largo plazo

4. Principios de Dimensionamiento

Períodos de ampliación
Situación
Requerimientos de grado de servicio
Variaciones de congestión
Relación entre ganancias y gastos
Posible optimización (teórica)
Solución práctica
Selección de períodos de ampliación

5. Casos de Conexión

Estacionaridad (stationarity)
Entrada de tráfico
Distribución de tiempo de ocupación
Agrupación
Cacería (hunting)
Llamadas sin éxito
 Sistemas de pérdida
 Sistemas de espera - disposición en cola (queue discipline)

1. Unidad de Tráfico

Intensidad de llamadas

El tráfico telefónico es generado por un número de abonados. Cuando un abonado tiene intención de hacer una llamada telefónica y levanta el auricular, la central telefónica local recibe un impulso que empieza un número de acciones para hacer posible el recibir información numérica de la persona que llama, de manera que la central telefónica puede conectar al abonado que llama con el abonado deseado. Cuando la central esta lista para recibir el número deseado, un tono de marcar es enviado a la persona que llama. La persona que hace la llamada puede entonces marcar el número deseado y la central telefónica conecta a la persona que llama con el número deseado.

El procedimiento de establecimiento implica generalmente que la llamada es conmutada sobre un número de selectores. Estos selectores pueden estar todos situados en la central local del propio abonado que llama. Sin embargo, cuando la red telefónica es lo suficientemente grande, la conmutación puede pasar por varias etapas en diferentes centrales. Consecuentemente, una unidad de conmutación que está manejando la llamada puede requerir otra unidad de conmutación, que puede ser localizada geográficamente cerca o lejos, para continuar el procedimiento de conmutación. Cuando se hace eso, también debe transferirse la información concerniente al destino deseado. Finalmente, cuando se pasa la última etapa de conmutación, el tono del abonado al que se está llamando indica que alguien quiere hablar con él.

El procedimiento de establecimiento de una llamada telefónica involucra, consecuentemente, una serie de requisitos para que la llamada sea procesada. El número total de dichos pedidos a una unidad de conmutación - o abonado - por unidad de tiempo es la intensidad de llamada.

Consecuentemente, la primera unidad de conmutación en la cadena de selección recibe llamadas de abonados, las siguientes unidades de conmutación reciben llamadas de la anterior, y finalmente, los abonados llamados reciben llamadas del último eslabón en la cadena de conmutación y otro equipo común que toma parte en los procedimientos de establecimiento también recibirá llamadas. Por tanto uno puede definir la intensidad de llamada por cada parte distintiva de la cadena de conmutación, desde la llamada hasta el abonado al que se llama.

Tiempo de ocupación

Cada llamada que es aceptada significa que algunos dispositivos son ocupados por cierto tiempo. La duración de este tiempo de ocupación depende de cuán rápido pueda llevarse a cabo el trabajo de conmutación o de procesamiento y de si el dispositivo se mantiene o no durante los próximos momentos de conmutación y durante la posible siguiente conversación entre los dos abonados. Los tiempos de espera pueden consecuentemente variar de llamada a llamada, dependiendo del tipo de dispositivo de conmutación utilizado y su función en el procedimiento de establecimiento y en el resultado de la llamada.

Número de ocupaciones simultáneas

Si consideramos un dispositivo de conmutación por separado, se comprende fácilmente que la ocupación de este dispositivo depende de cuántas llamadas recibe y cuánto tiempo ocupa cada llamada este dispositivo.

Si ahora consideramos un grupo de dispositivos donde cada uno sólo puede atender una llamada a la vez, entendemos que el número de dispositivos ocupados puede variar como se muestra en la Figura TFC 1/1.

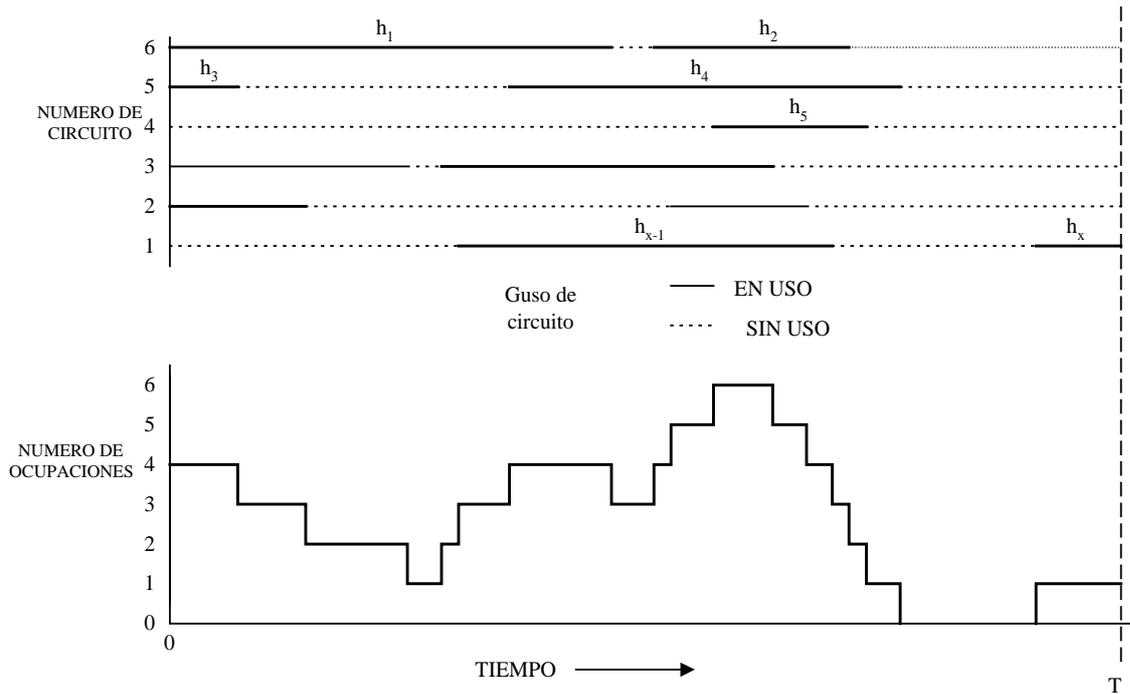


Figura TFC 1/1: Ocupaciones en un grupo de 6 dispositivos

El trabajo total llevado a cabo por el grupo puede, para el intervalo considerado, expresarse como:

$$\bar{x}h = \sum_{i=1}^x h_i$$

si se manejan x llamadas en el intervalo, donde \bar{h} es el promedio de todas las h_i . Sea T la duración del intervalo considerado, entonces

$$\frac{1}{T} \sum_{i=1}^x h_i$$

da el número promedio de ocupaciones

Esto también puede escribirse

$$\frac{x \cdot \bar{h}}{T}$$

y si $y = \frac{x}{T}$ es el número promedio de llamadas por unidad de tiempo, podemos en consecuencia escribir

$$A = y \cdot \bar{h}$$

(TFC 1.1)

que es el número promedio de ocupaciones simultáneas en el intervalo (0, T).

El número promedio de ocupaciones simultáneas es generalmente llamado *intensidad de tráfico* o *flujo de tráfico*.

Unidades de tráfico

La intensidad de tráfico con frecuencia es llamada simplemente tráfico.

La unidad de tráfico es el erlang. Consecuentemente, el número de erlangs es simplemente el número promedio de ocupaciones simultáneas para un tiempo interno definido.

El número de erlangs (A) usualmente se calcula como:

$$A = y \cdot h$$

donde

y = al número de llamadas por unidad de tiempo o al número de nuevas ocupaciones por unidad de tiempo

h = al promedio de duración de estas ocupaciones, expresado en la misma unidad de tiempo.

Ejemplo:

$$y = 3600 \text{ llamadas/hora} = 60 \text{ llamadas/min} = 1 \text{ llamadas/seg}$$

$$h = \frac{1}{60} \text{ hora} = 1 \text{ min} = 60 \text{ segs}$$

$$A = 3600 \times \frac{1}{60} = 60 \times 1 = 1 \times 60 = 60 \text{ erl.}$$

El tráfico también puede calcularse como la suma de todos los tiempos de ocupación dividida entre el período de tiempo concerniente, tal como se muestra anteriormente.

Ejemplo:

La suma de todos los tiempos de espera en un grupo era de 225 minutos durante un intervalo de tiempo de 15 minutos, consecuentemente:

$$A = \frac{225}{15} = 15 \text{ erl.}$$

Un tercer modo algo artificial para calcular el tráfico, consiste en definir cuánto tiempo están una, dos, tres, etc. ocupaciones en un grupo. Asumamos que el grupo tiene n dispositivos y t_p es parte del intervalo de tiempo total T, que hay exactamente p ocupaciones, entonces el tráfico se puede calcular como:

$$A = \sum_{p=1}^n p \frac{t_p}{T}$$

donde:

$$\sum_{p=0}^n t_p = T$$

Un método práctico para medir el tráfico es explorar el grupo a intervalos regulares. El promedio de estas muestras provee un estimado del tráfico observado, esto es,

$$A = \frac{1}{N} \sum_{v=1}^N f_v$$

donde N registros son hechos en el intervalo de exploración y el número de dispositivos ocupados encontrados en la v : la exploración es fv.

Este método tiene cierta incertidumbre estadística que depende del intervalo de exploración, T/N, el promedio de tiempo de ocupación y el tráfico cursado en el grupo observado.

La unidad de tráfico erlang es una unidad aceptada internacionalmente. En Estados Unidos de América, sin embargo, la unidad de señalización por canal común (Cien Segundos de Llamadas) se usa generalmente en el trabajo de tráfico práctico.(Señalización por canal común, SCC; Common channel signalling, CCS). De acuerdo a esta definición tenemos que:

$$\begin{aligned} 1 \text{ erlang (erl)} &= 36 \text{ CCS or} \\ 100 \text{ CCS} &= 2.778 \text{ erl.} \end{aligned}$$

Volumen de tráfico

Para expresar la suma de todos los tiempos de espera cursado por un grupo durante un período dado, se utiliza la unidad de erlang- horas. Estos tiempos de espera son entonces expresados en horas.

Ejemplo:

Un grupo cursa 1200 llamadas en un período de 24 horas. El promedio de espera fue de 2 minutos = 1/30 hora

$$1200 \times \frac{1}{30} = 40 \text{ erlang horas}$$

Una unidad frecuentemente utilizada para el volúmenes de tráfico es el número de minutos de llamadas pagadas (pcm) por mes o por año. Es evidente que pcm dividido entre 60 nos da el número de erlang horas pagadas.

Los volúmenes de tráfico son especialmente de interés en razón de las ganancias.

Cuando se trata con el concepto de tráfico, debe hacerse una clara distinción entre

DEMANDA DE TRAFICO
TRAFICO OFRECIDO
TRAFICO CURSADO

todos ellos expresados en erlangs, pero sólo el tráfico cursado puede ser medido

2. Generación de tráfico

Resultado posible de un intento de llamada, intentos repetidos

Cuando un abonado A desea hablar con un abonado B, el resultado es el establecimiento de una conversación o que A desiste. Ver Figura RFC 2/1:

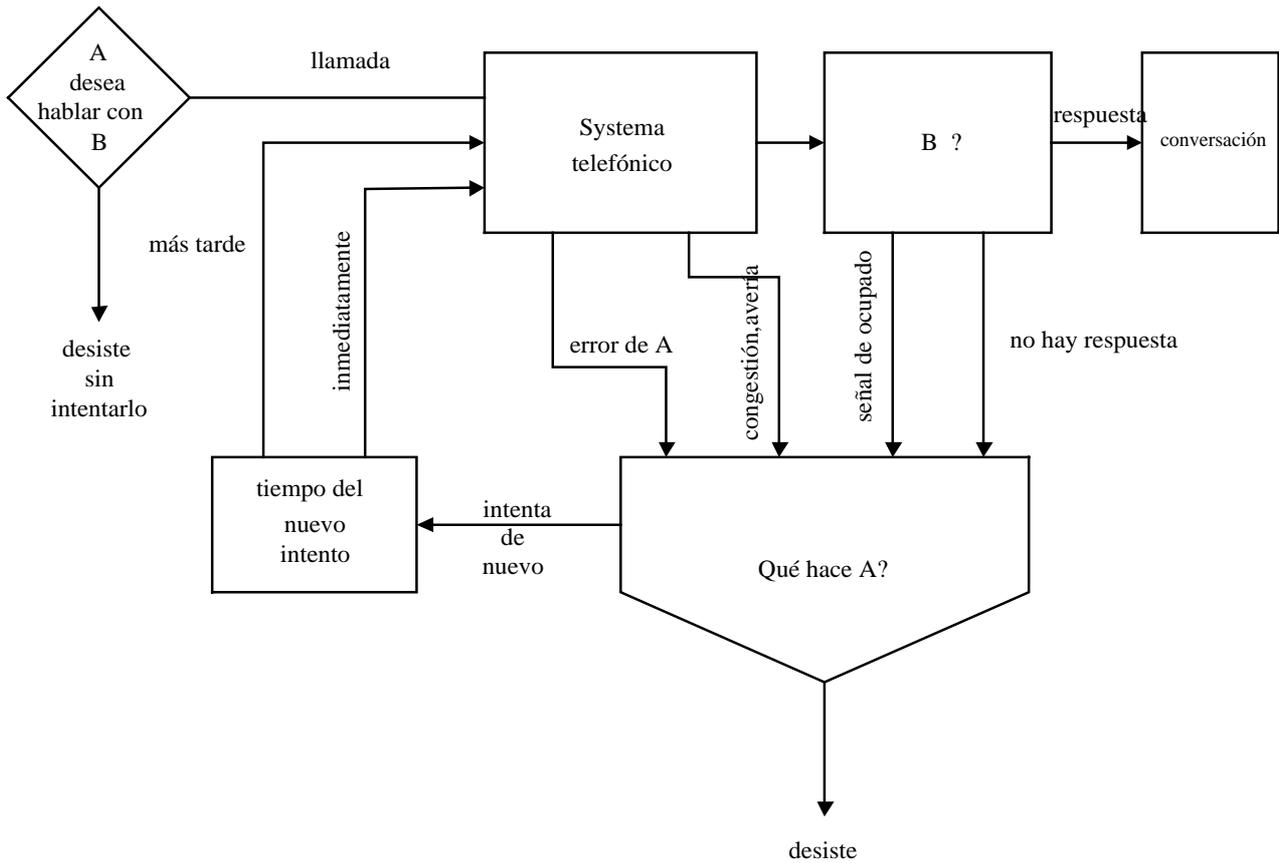


Figura TFC 2/1: Diagrama de acciones para el establecimiento de una conversación entre A y B

Se puede ver en la Figura TFC 2/1 que la conversación puede establecerse ya sea directamente o después de cierto número de fracasos. La decisión de A de hacer otro intento puede depender de:

- 1) número de intentos previos sin éxito;
- 2) A sabe que marcó el número equivocado o un número incompleto;
- 3) A no está seguro que marcó el número correcto;
- 4) A piensa que pudo haber alguna avería en el sistema telefónico;
- 5) conocimiento de los hábitos telefónicos de B si no hay contestación o da señal de ocupado;
- 6) grado de urgencia

El momento para hacer un nuevo intento se decide teniendo en cuenta:

- a) la presunción de A sobre la causa de la falla;
- b) la estimación de A acerca del momento apropiado para intentarlo otra vez;
- c) la posibilidad de A de hacer un nuevo intento;
- d) grado de urgencia.

Razones para que A no trate nuevamente:

- a) A considera inútil un nuevo intento;

b) el grado de urgencia es bajo.

Si A desiste sin hacer ningún intento, puede indicar que tiene una opinión muy pobre sobre la calidad operacional del sistema telefónico o de la posibilidad de conseguir a B. La urgencia también puede ser baja.

Se puede ver en la Figura TFC 2/1 que, de las llamadas ofrecidas a un sistema telefónico, no todas resultan en conversaciones.

Las estadísticas de un sistema telefónico con un servicio razonablemente bueno, pueden parecerse a las siguientes:

Discado incompleto o erróneo (error de A)	5 - 10 %
Congestión y averías técnicas	1 - 5 %
B está ocupado	10 - 20%
B no contesta antes que A cuelgue	10 - 15 %
<hr/>	
No hay conversación	30 - 50 %
Hay conversación	70 - 50 %
Número de intentos por cada llamada lograda	1.4

Para una red telefónica con un servicio inferior y alto tráfico por abonado, las estadísticas pueden ser las siguientes:

Discado incompleto o equivocado (error de A)	13%
Congestión y averías técnicas	25%
B está ocupado	20%
B no contesta antes que A cuelgue	10%
<hr/>	
No hay conversación	68%
Hay conversación	32%
Número de intentos por cada llamada lograda	4.8

El número de errores al marcar crece con el número de llamadas!

Duración de los diferentes tipos de llamadas

- Los intentos sin éxito usualmente ocupan el equipo de conmutación en el sistema telefónico por un tiempo más corto que las conversaciones. El tiempo promedio de ocupación de los conmutadores es, por tanto, más corto que el tiempo de conversación promedio.

- Los intentos de conexión pueden ser interrumpidos en cualquier momento. Por tanto, el tiempo promedio de espera de los conmutadores al principio de la cadena de conmutación es más corto que el de los conmutadores al final de la cadena. (ver Figura TFC 2/2):

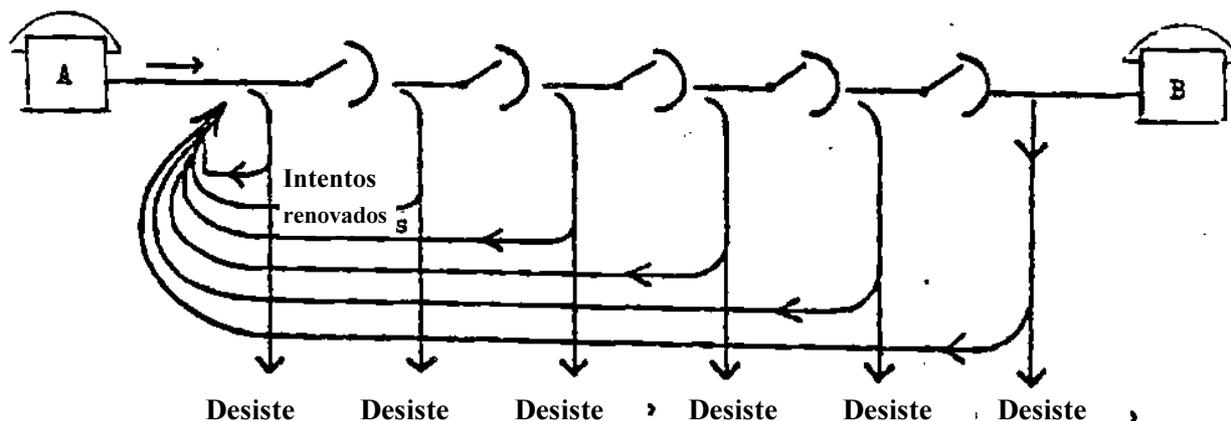


Figura TFC 2/2 : Un intento de conexión se puede interrumpir en cualquier lugar de la cadena de conmutación de A hacia B, pero cada nuevo intento siempre comienza desde el principio otra vez.

La duración de los diferentes tipos de ocupaciones de las etapas del sector varía considerablemente, como se deduce de la siguiente tabla:

Tipo de ocupación	Tiempo promedio de espera típico en segs.
Conversación, llamada local	45 - 120
Conversación, llamada de larga distancia	180 - 300
Conversación, llamada internacional	300 - 600
Un abonado, escuchando el tono cuando no hay respuesta	20
Un abonado, escuchando el tono de ocupado	5
Tiempo entre el tono de marcación y el inicio de la marcación	2
Tiempo de marcación por dígito	1.5 (con disco) 0.6 (con teclado)
Tiempo hasta el próximo intento	30

Las diferentes etapas de conmutación llevan una mezcla de diferentes tipos de llamadas, cuyos tiempos de espera y frecuencias pueden variar considerablemente.

El cambio del tiempo promedio de espera en la cadena de conexión desde el abonado que llama hasta el abonado llamado, se puede ilustrar con los valores siguientes obtenidos en una central local:

Etapas	Tiempo de ocupación promedio, en segs.
Etapas de abonado originando tráfico	80
Selector del primer grupo	100
Selector del segundo grupo	120
Etapas de abonado, tráfico de destino	150
Tiempo de conversación	220

Si se conocen los tiempos promedio de espera para los diferentes tipos de llamadas sin éxito y sus proporciones ("la mezcla de llamadas"), puede estimarse el número de intentos por llamadas con éxito, así como la tasa de llamadas completas, es decir, el porcentaje de intentos que llevan a conversaciones.

Es necesario distinguir claramente entre los siguientes cuatro conceptos de tráfico:

DEMANDA DE TRAFICO
 TRAFICO OFRECIDO
 TRAFICO CURSADO
 TRAFICO FACTURADO

La Demanda de Tráfico (A_D) es el tráfico que a los abonados les gustaría realizar si no hubiera obstáculos, como congestión, averías técnicas o estuviese ocupado el abonado B. La demanda de tráfico también puede variar con el costo de las llamadas, de modo que uno también tiene que distinguir entre la demanda de tráfico a tarifa fija y cuando la tarifa telefónica es variable. La demanda de tráfico es una cantidad hipotética que sólo puede estimarse, pero no medirse.

El Tráfico Ofrecido (A_O) es el tráfico ofrecido a un grupo, de acuerdo con una descripción teórica definida del caso de tráfico. Es consecuentemente también una cantidad hipotética que se convierte en significativa sólo si se refiere a un modelo teórico específico.

El Tráfico Cursado (A_C) es el tráfico manejado por un grupo. Se refiere a ambas, a una descripción teórica definida y a la realidad; esta última puede ser medida.

El Tráfico Facturado es la parte del tráfico cursado que es facturado al abonado. El tráfico cursado y el tráfico facturado difieren, ya que en la mayoría de los casos, el abonado sólo paga por las llamadas que fueron contestadas. Para llamadas de larga distancia e internacionales el cargo puede comenzar un poco antes, pero no antes que la llamada llegue a la primera central de tránsito. Consecuentemente, las estadísticas del tráfico facturado sólo darán una parte del tráfico real cursado.

Las siguientes expresiones pueden usarse para A_D , A_O y A_C :

$$\begin{array}{rcl}
 A_D & = & y_D \cdot h_D \\
 A_O & = & y_O \cdot h_O \\
 A_C & = & y_C \cdot h_C
 \end{array} \quad \left. \begin{array}{l} | \\ | \\ | \end{array} \right\} \text{(TFC 2.1)}$$

donde y_D , y_O e y_C son las intensidades de llamada y h_D , h_O , h_C son los tiempos promedios de espera. Su orden relativo de magnitud es:

$$\begin{array}{rcl}
 A_D & <> & A_O & \geq & A_C \\
 y_D & <> & y_O & \geq & y_C \\
 h_D & <> & h_O & \geq & h_C
 \end{array}$$

El tráfico cursado, A_C , puede dividirse en dos partes:

$$A_C = A_{CC} + A_{CF} \quad \text{(TFC 2.2)}$$

donde A_{CC} = tráfico de conversación
 A_{CF} = tráfico ocasionado por intentos frustrados.

Podemos entonces expresar A_{CC} y A_{CF} así:

$$\begin{array}{rcl}
 A_{CC} & = & y_{CC} \cdot h_{CC} \\
 A_{CF} & = & y_{CF} \cdot h_{CF}
 \end{array} \quad \left. \begin{array}{l} | \\ | \end{array} \right\} \text{(TFC 2.3)}$$

donde generalmente $h_{CF} < h_{CC}$ or $h_{CF} \ll h_{CC}$

mientras la relación entre y_{CF} e y_{CC} depende de la tasa de completación.

Congestión

Si una llamada solicita ser servida por un grupo que ya tiene todos sus dispositivos ocupados, entonces ocurre congestión. Esto significa que la llamada no puede ser aceptada por el momento. Dependiendo del sistema utilizado, esa llamada puede ser rechazada (Sistema de Pérdida), o que se le permita esperar (Sistema de Espera). En el primer caso, el abonado que llama recibe tono de ocupado y debe hacer un nuevo intento. En el último caso, las llamadas retardadas se atenderán tan pronto como cualquiera de los dispositivos estén disponibles. Existen diferentes métodos técnicos para manejar llamadas en espera en lo referente a la elección del próximo a ser servido. Estos tienen cierto impacto en el cálculo teórico de los tiempos de espera.

Respecto a la congestión en el Sistema de Pérdida, se hace una distinción entre:

CONGESTION TEMPORAL y CONGESTION DE LLAMADAS
--

La Congestión Temporal es la parte del tiempo en que llamadas adicionales no pueden ser servidas.

La Congestión de Llamadas es la parte de las llamadas rechazadas, o forzadas a esperar en el sistema de espera.

Ambas son cantidades que pueden medirse. También pueden ser referidas a una descripción teórica definida de un caso de tráfico.

3. Variaciones de Tráfico

El tráfico varía al ritmo de la actividad en la comunidad. Por tanto, los siguientes tipos de variaciones que existen son:

A. *Variaciones durante el día*

a) Los picos de tráfico ocurren una, dos o tres veces al día durante horas normales de trabajo. (El tráfico es usualmente más bajo los sábados y días feriados)

b) El tráfico máximo no siempre ocurre a la misma hora todos los días

B. *Variaciones durante la semana*

Ciertos días de la semana pueden tener sistemáticamente mayor tráfico que los otros días.

C. *Variaciones de temporada*

Alto tráfico durante ciertas temporadas del año y tráfico bajo durante otras. Usualmente el tráfico es alto antes de un feriado grande (Año Nuevo, Navidad, etc.) y es bajo después.

D. *Tendencia de crecimiento*

El tráfico muestra generalmente una tendencia consistente a incrementarse. Este incremento no es uniforme. Por tanto, los valores alto y bajo pueden tener aumentos relativos diferentes. Es generalmente más difícil distinguir entre crecimiento y variaciones de temporada.

E. *Picos Impredecibles*

El tráfico más alto durante el año puede ocurrir durante una temporada alta, pero ocasionalmente puede ocurrir en cualquier otra época del año.

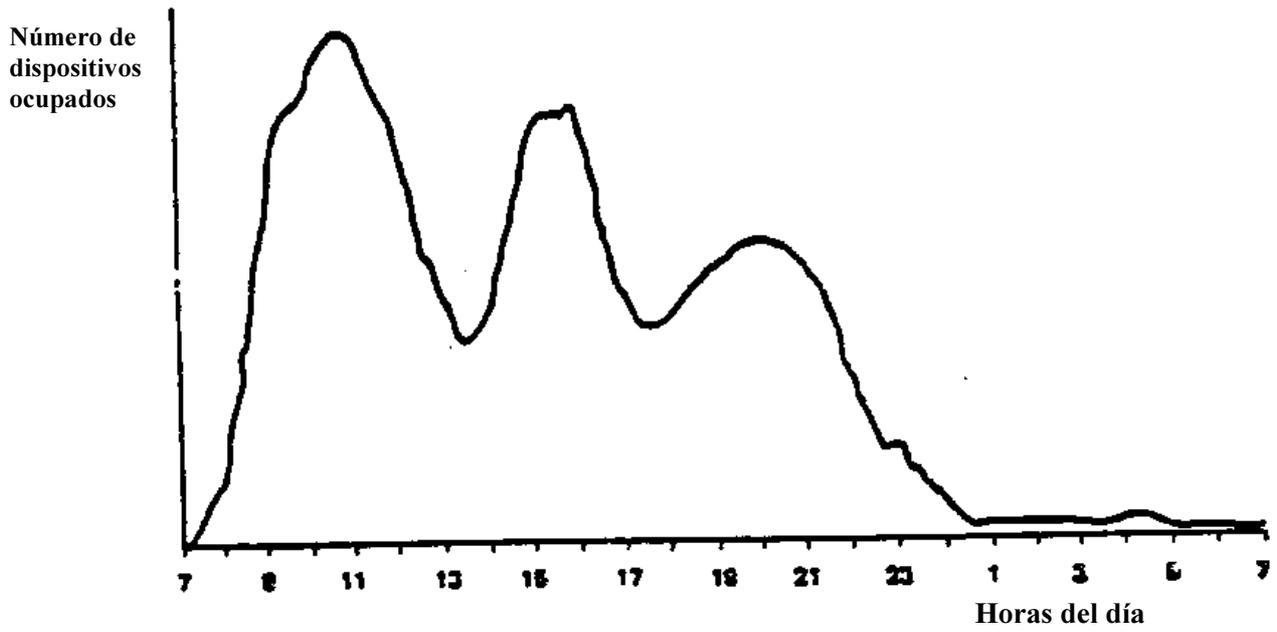


Figura TFC 3/1: Variaciones de tráfico durante un día

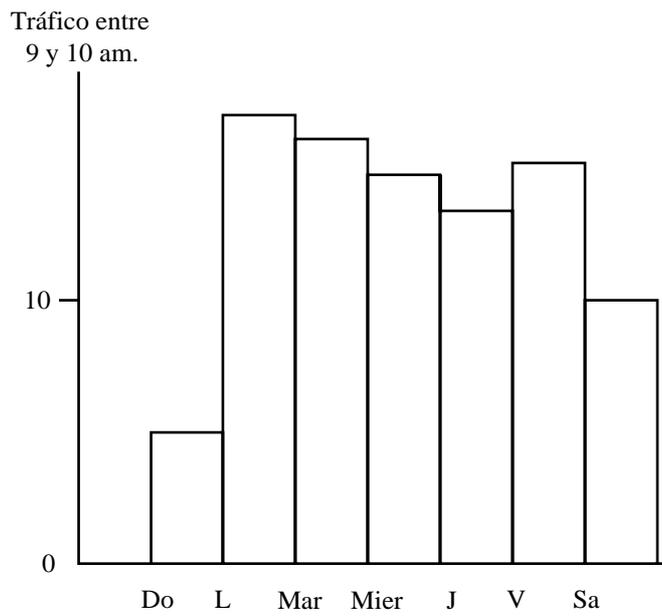


Figura TFC 3/2 : Variaciones de tráfico durante una semana

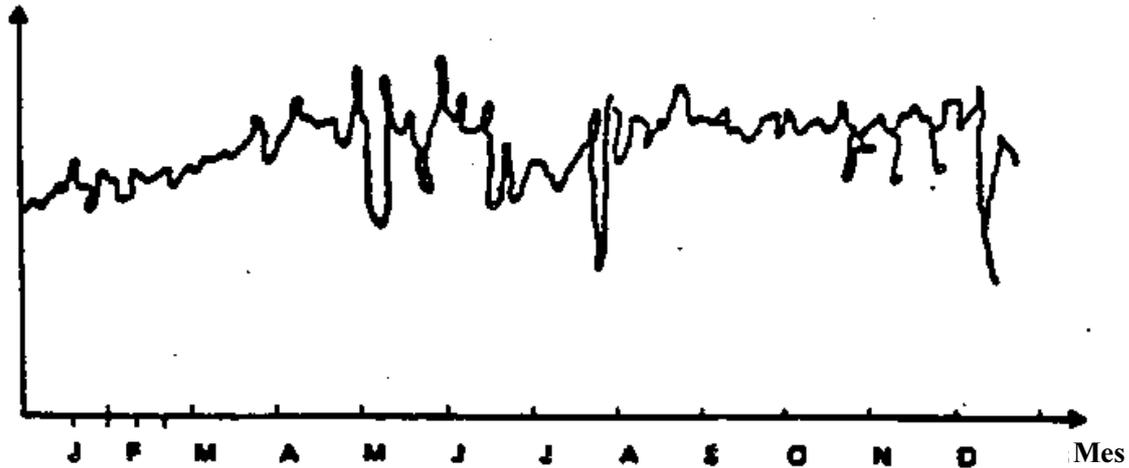


Figura TFC 3/3: Variaciones estacionales durante un año, expresadas durante la hora de tráfico pico, 9-10 a.m., cada tercer día normal de trabajo

F. *Variaciones en intervalos cortos*

Si se estudia el número de ocupaciones simultáneas en un grupo es estudiado por un intervalo de tiempo corto, como una hora, se encontrará que varía de manera irregular. Estas variaciones pueden describirse como variaciones aleatorias. Dependiendo de la hora del día, el número de ocupaciones puede mostrar una tendencia a aumentar con el tiempo, disminuir o permanecer aproximadamente a un nivel constante.

La teoría del tráfico se basa en las condiciones durante la hora pico, donde las variaciones generalmente ocurren alrededor de un nivel medio, con menos tendencia a aumentar o disminuir que en otros momentos del día.

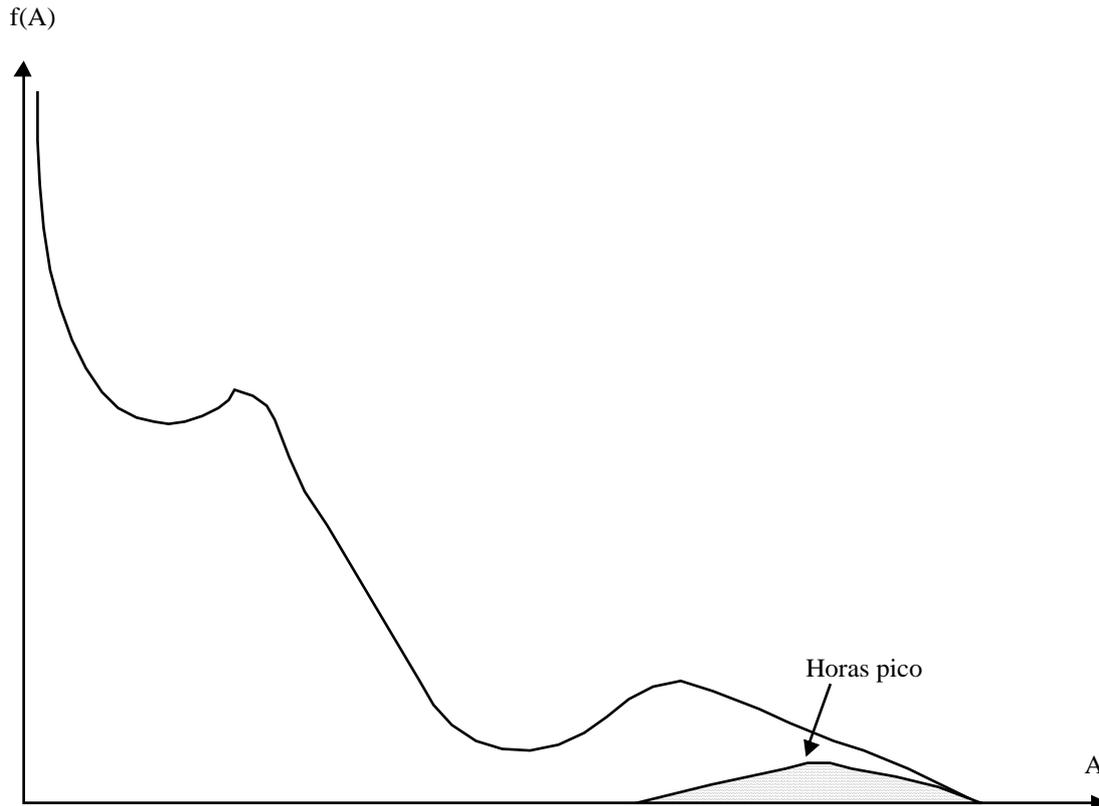
El tráfico muestra consecuentemente variaciones aleatorias, así como más o menos variaciones sistemáticas. Para entender el significado de los valores de tráfico dados en erlangs es necesario, por tanto, tener la información de cómo ellos son observados. Muestras tomadas durante temporadas altas, temporadas menos altas y durante temporadas bajas pueden tener significado muy diferente.

La HORA PICO es el período de 60 minutos durante el día cuando el tráfico es mas alto a LARGO PLAZO (LONG RUN) (hora pico consistente en el tiempo)

La HORA PICO MAS ALTA (BUSIEST HOUR) es el período de 60 minutos, todos los días cuando el tráfico es más alto.

Consecuentemente, la hora pico y la hora pico más alta pueden no coincidir, y la hora pico más alta puede ocurrir en diferentes momentos del día.

La variación del tráfico durante un período de tiempo (por ej. el tráfico de cada hora durante el año) se puede describir mediante un histograma (Figura TFC 3/4).



**Figura FTC 3/4 Variación de tráfico durante todas las 8760 horas de un año.
El área sombreada es la contribución de las horas pico**

Si el tráfico durante todas las 8760 horas de un año se representa en el histograma, se encuentra que hay un número considerable de horas con tráfico muy bajo. Los valores más altos se derivan mayormente de las horas pico (área sombreada del diagrama), pero las horas fuera de las horas pico pueden también contribuir a la parte alta de la curva de distribución.

El promedio de tiempos de espera también puede tener tanto variaciones diarias como estacionales, pero de menor magnitud que las variaciones de tráfico.

La presente teoría de tráfico sólo se puede usar para cálculos prácticos de congestión cuando el tráfico está en equilibrio, que corresponde a grandes rasgos a las condiciones durante los picos de tráfico.

4. Principios de dimensionamiento

Debido a la estructura y diseño de los sistemas de telecomunicaciones, la provisión de equipo no puede hacerse para igualar todas las variaciones de tráfico de tiempo corto. Las varias partes del sistema deben, por eso, ampliarse en pasos suficientemente largos para durar cierto período, usualmente al menos 6 meses hasta 5 años. Los pasos de ampliación deben ser lo suficientemente largos para evitar la congestión perturbadora durante todo el período

Para cada período de ampliación se hace una proyección de tráfico que debe describir cómo el tráfico va a variar y a crecer. La provisión depende de los siguientes factores:

- 1) Los ingresos del tráfico telefónico dependerán del tráfico cursado, usualmente sobre el número de conversaciones establecidas.
- 2) Los gastos dependen del grado de servicio que se desea dar a los abonados durante condiciones de tráfico pico.
- 3) La mayor parte del volumen de tráfico cursado se maneja cuando prácticamente no hay congestión.

- 4) Congestionamientos verdaderamente altos ocurren sólo en pocas ocasiones.
- 5) Estas ocasiones de congestión alta ocurren más a menudo al final de un período de ampliación que al principio.
- 6) No es económicamente factible dimensionar un sistema telefónico para que nunca ocurra congestión.
- 7) Un grado de servicio mejorado usualmente resulta en cierto crecimiento de tráfico cuando los abonados se dan cuenta que es más fácil conseguir las llamadas (salto por mejoramiento de servicio o estímulo de tráfico).
- 8) Los factores humanos siempre causan más fallas que el sistema telefónico. Por “factor humano” se entiende: errores del abonado que llama, el abonado al que se llama está ocupado o no hay contestación.
- 9) El abonado individual no reacciona mucho a la congestión cuando ésta es menor, digamos, de un 10 %.
- 10) El tráfico pico, por lo general y comparativamente, no es afectado por cambios menores en las tarifas. Esto se debe a que el grueso del tráfico durante las horas pico proviene de abonados de negocios y no de personas privadas.

El problema de dimensionamiento:

Cuánta congestión puede permitirse durante unas pocas horas del año?

Durante un período dado, las variaciones de tráfico durante todas las horas del año pueden describirse como se muestra en la figura TFC 4/1:

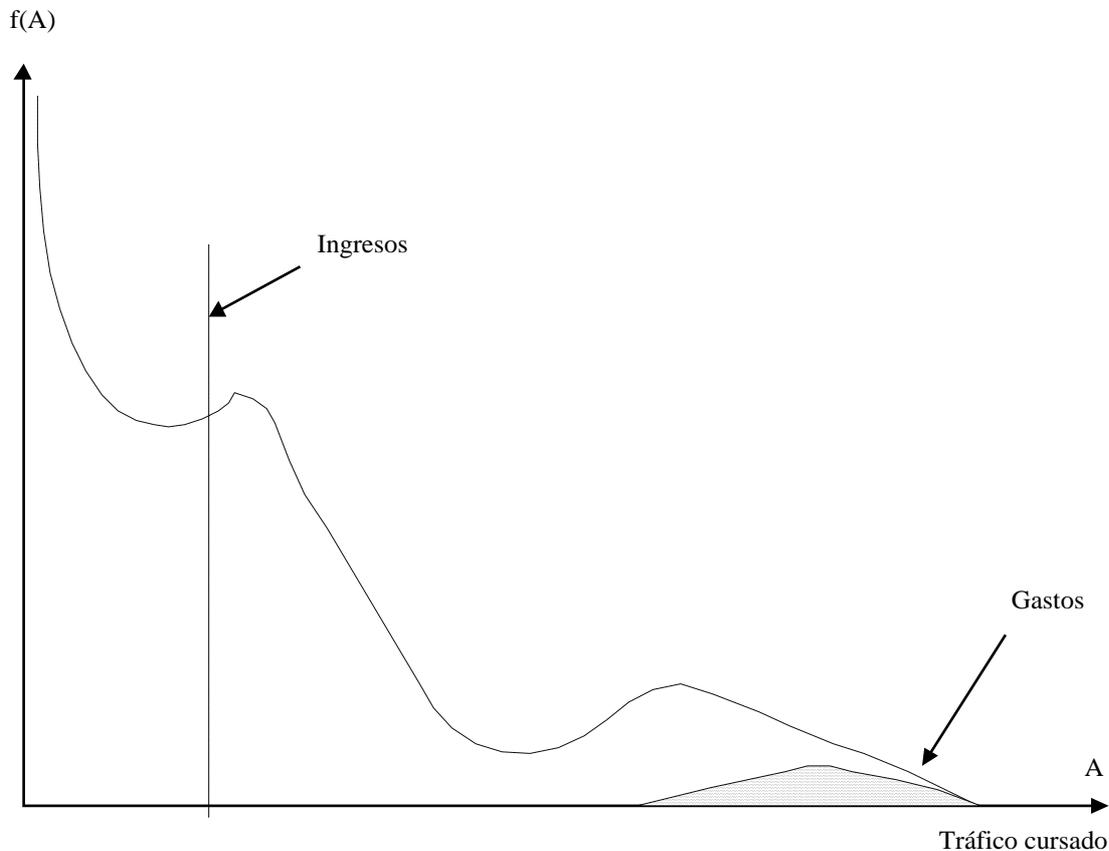


Figura TFC 4/1: Variaciones de tráfico (todas las horas) durante un período de ampliación

Si cierto número de circuitos se utiliza para llevar el tráfico arriba presentado, la congestión variará como se muestra en la figura 4/2:

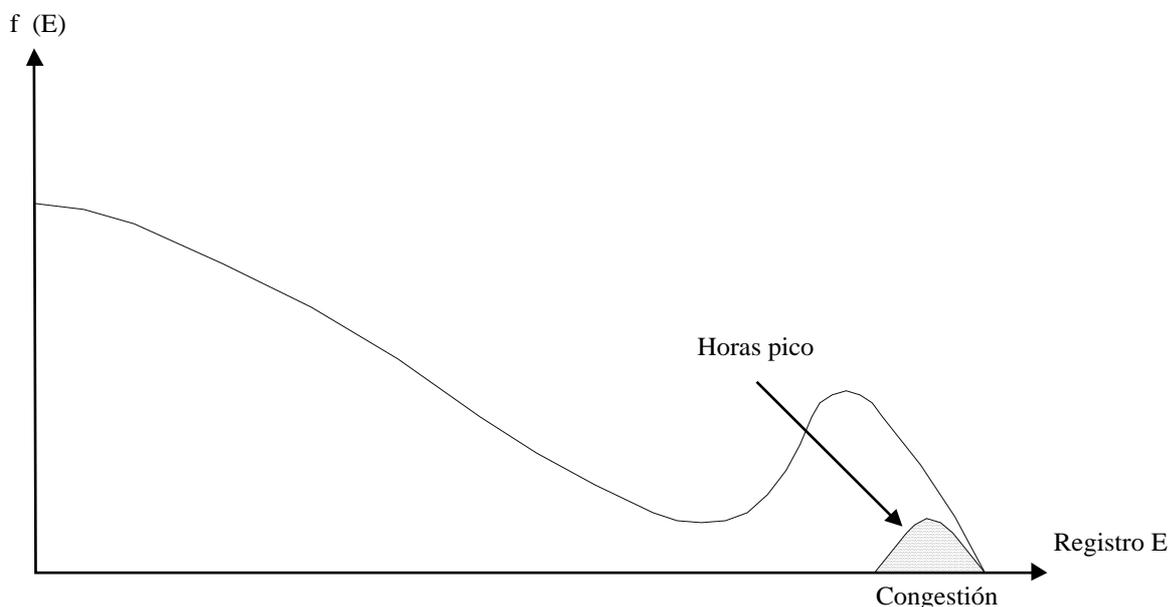


Figura TFC 4/2 Variación de la congestión (todas las horas) durante un período de ampliación

Los ingresos son aproximadamente proporcionales al tráfico cursado. Esto significa que el ingreso es proporcional al área entre $f(A)$ y el eje x en la Figura TFC 4/1.

Los gastos dependen principalmente con cuánta frecuencia se permite a la congestión exceder ciertos niveles dados.

Si se proveen más dispositivos en grupo, la congestión mensurable ocurrirá con menos frecuencia y, por supuesto, la alta congestión también será un raro suceso. Se entiende de los diagramas en las figuras TFC 4/1 y TFC 4/2, que un cambio menor en el número de dispositivos de un grupo no ocasiona mucho cambio en los ingresos, mientras que los gastos pueden ser más directamente proporcionales al número de dispositivos provistos. El dimensionamiento, por tanto, se convierte en un asunto de balance entre los gastos y los requisitos del grado de servicio, donde el grado de servicio es expresado en relación a cuán frecuentemente se permite a la congestión exceder ciertos valores.

Este es el enunciado general del principio de dimensionamiento. Para ponerlo en práctica, los administradores telefónicos deben ceñirse a lo que es observable. Ya que actualmente el tráfico cursado y la congestión no pueden observarse continuamente, uno debe limitarse a lo que es realmente mensurable.

Un procedimiento común es medir el tráfico de hora pico durante 5 o 10 días específicos durante una temporada de tráfico alta, cada año. Más mediciones se pueden hacer con equipo muy caro. Generalmente, la congestión también se mide simultáneamente. Para controlar el funcionamiento del servicio diario, también pueden registrarse fuera de estos períodos de medición valores que exceden ciertos niveles altos dados.

Las mediciones anuales son generalmente usadas como base para proyectar el tráfico. Ya que se sabe que estos períodos medidos generalmente no alcanzan los picos de tráfico más altos del año, la provisión de conmutadores y troncales se escoge para una congestión baja, generalmente entre 0.1 y 5%. Esto se espera proteja contra congestiones altas perturbadoras (como 20% o más) que ocurren muy frecuentemente en períodos de ampliación, siempre que el pronóstico sea correcto.

El mecanismo de períodos de provisionamiento, incremento del tráfico previsto y el grado de servicio resultante se ilustran en la figura 4/3 donde, por supuesto, el grado de servicio tiene que referirse a un nivel de tráfico definido.

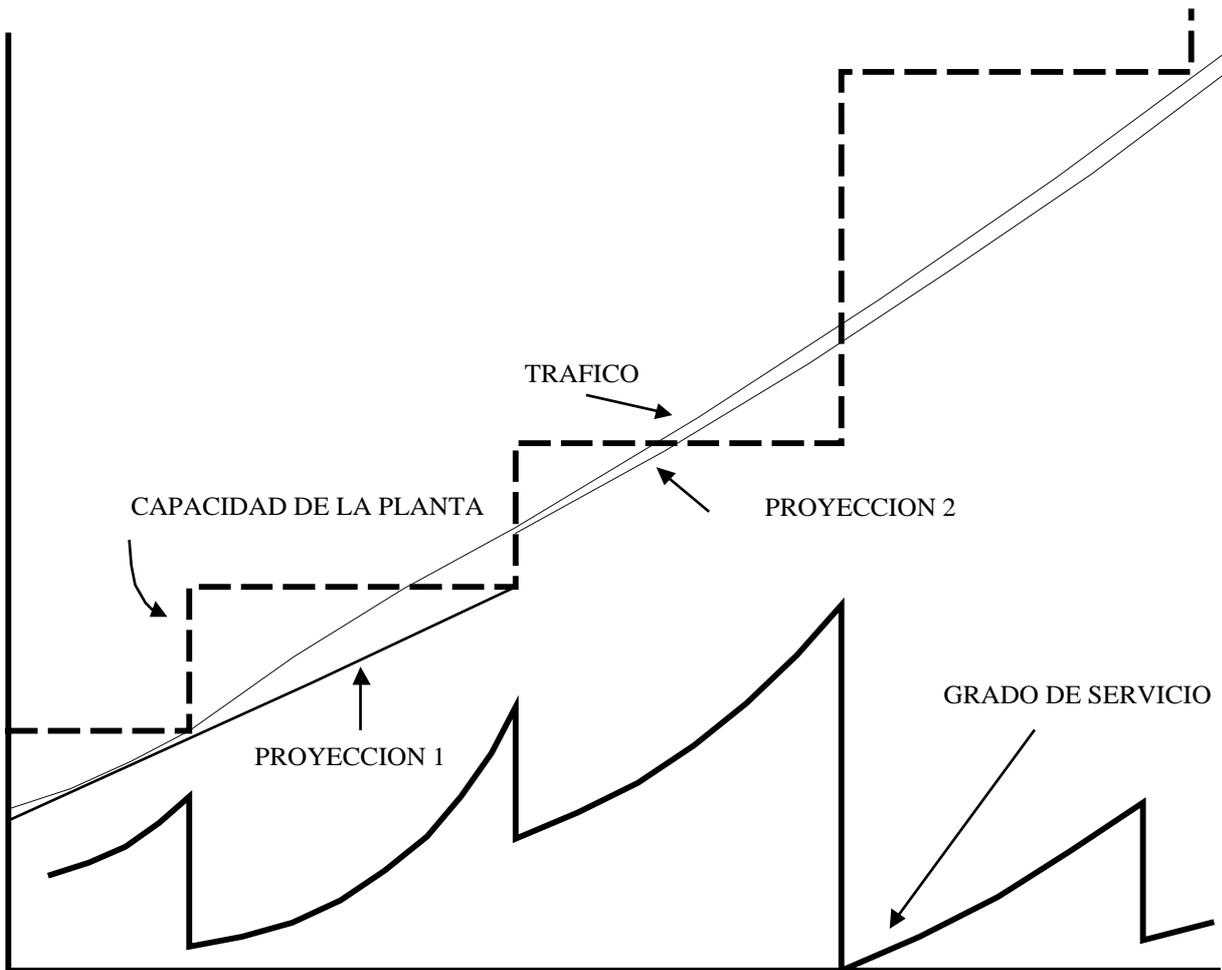


Figura TFC 4/3: Efecto del período de provisión en el grado de servicio

Elección de períodos de ampliación

La elección de pasos y períodos de ampliación apropiados es de gran importancia para el mejor uso económico del capital disponible para inversiones. Esto es así, debido a que las líneas y el tráfico telefónico continúan creciendo.

La elección de períodos de ampliación depende de un número de factores. Los más importantes son:

- tasa de crecimiento;
- costo del equipo adicional;
- unidades de ampliación técnicamente adecuadas;
- tiempos de entrega;
- tiempo requerido para la instalación de una ampliación;
- coordinación con otras ampliaciones;
- flexibilidad respecto a errores de proyección;
- costos de instalación;

- interés del capital invertido;
- ingresos de tráfico;
- límites de inversión del capital.

Se entiende que el equipo caro debe extenderse en pequeños pasos mientras que el menos caro puede ampliarse en pasos más largos. Los pasos de ampliación óptimos, sin embargo, dependen del programa total de ampliación para todo el país completo y no se puede considerar para cada caso individual. No obstante, a veces los recursos limitados para inversiones pueden obligar a la administración telefónica a usar pasos de ampliación menos económicos.

5. Casos de Conexión

Una conexión a través de un sistema telefónico se da en un número de selecciones sucesivas. Para el cálculo y establecimiento (assessment) de las características de tráfico en cada paso de selección, deben definirse los siguientes factores:

- a) entrada de tráfico a la etapa;
- b) agrupación de la etapa;
- c) cómo maneja la etapa las llamadas entrantes (método de cacería);
- d) el procedimiento para tratar con las llamadas sin éxito.

Equilibrio

El supuesto usual en los cálculos es el equilibrio estadístico - "estacionario" el tráfico no está ni creciendo ni decreciendo y el proceso de tráfico ha continuado por tanto tiempo que el estado inicial no es significativo.

Entrada de tráfico

La entrada de tráfico se define por la intensidad de la llamada y por la distribución del tiempo de ocupación.

Intensidad de llamada:

Como regla, el tráfico se describe de acuerdo al número de fuentes simples (una fuente simple sólo puede producir una llamada a la vez). Para la intensidad de llamada, los supuestos usuales son:

BE: Bernoulli o entrada Engset

Número limitado de fuentes, intensidad de llamada decreciendo con el número de ocupaciones:

BE : Bernoulli or Engset input

Número limitado de fuentes, intensidad de llamada decreciendo con el número de ocupaciones:

$$y(p) = (N - p) \cdot \beta \quad (\text{TFC 5.1})$$

N = número de fuentes

p = número de ocupaciones

β = intensidad de llamada por fuente cuando está libre (y la fuente ocupada tiene intensidad de llamada 0)

P = entrada de Poisson

Número infinito de fuentes, intensidad de llamada independiente del número de ocupaciones:

$$y(p) = y; \quad y = \lim (N\beta) = \text{finito}$$
$$N \rightarrow \infty$$
$$\beta \rightarrow 0 \quad (\text{TFC 5.2})$$

NB : Entrada de tipo binomial negativa.

Se asume que la intensidad de la llamada aumenta con el número de ocupaciones:

$$y(p) = a + b \cdot p \quad (\text{TFC 5.3})$$
$$y(p) = (N - p) \beta \quad a = N\beta$$
$$b = -\beta$$

Es decir, NB se obtiene de BE si $\beta < 0$ y $N < 0$.

NB a veces se usa para describir el tráfico con características especiales.

Distribución de espera:

Para la distribución del tiempo de ocupación, usualmente se asume distribución exponencial negativa.

$$f(t) = \frac{1}{s} e^{-\frac{t}{s}} \quad (\text{TFC 5.4})$$

El tiempo de ocupación constante también puede existir. En la práctica ocurren con frecuencia casos que incluyen ambos, tiempos de espera constantes y tiempos de espera exponencialmente distribuidos.

La descripción del tráfico de entrada debe basarse en condiciones reales, como el número de llamadas simultáneas que pueden producir las entradas.

Agrupación

Dependiendo de la disponibilidad entre entradas y salidas, un arreglo de agrupación se clasifica como sigue:

1. Grupo de disponibilidad total: cada entrada tiene acceso a cada salida.
2. Gradación (Grading): una entrada tiene acceso sólo a un número limitado de salidas (disponibilidad limitada). Las entradas con acceso a exactamente las mismas salidas, constituyen un grupo de entrada. Al menos algunas salidas (circuitos de salida) son comunes a más de un grupo de entrada. Ningún grupo de entrada es independiente de los otros. Es usado cuando se requieren más circuitos que a los que el selector individual tiene acceso.
3. Sistema de enlace: la conexión de la entrada a la salida tiene lugar sobre una o más etapas de enlace. Se aplica selección condicional. Selección condicional = selección efectuada, sólo si la entrada puede encontrar un trayecto libre hacia las salidas deseadas. Requiere información referente a enlaces libres y ocupados y salidas en el sistema de enlace y el conocimiento de cómo pueden combinarse los dispositivos en diferentes etapas; es decir, requiere de un marcador o control procesador. Es usado particularmente en sistemas de barras cruzadas y en sistemas de memoria de programas.
4. Sistema de enlace gradual: combinación de 2 ó 3.

Métodos de cacería

Hay principalmente dos métodos de cacería:

1. Cacería secuencial: cazar por una salida libre siempre comienza en la misma posición; se toma la primera salida.
2. Cacería aleatoria: cada salida libre tiene la misma probabilidad de ser tomada.

El método de cacería es menos significativo para grupos de disponibilidad total.

Usualmente son aconsejables diferentes diseños de sistemas de gradación (grading) y sistemas de enlace para diferentes métodos de cacería.

Llamadas sin éxito

Dependiendo, cuando se encuentra la congestión, de que el abonado pueda esperar hasta ser servido o deba hacer un nuevo intento, los sistemas telefónicos se clasifican en:

1. sistemas de pérdida;
2. sistemas de espera.

Para sistemas de pérdida se puede asumir, por ejemplo, que:

- a) las llamadas sin éxito hacen que no aumente la intensidad de llamadas desde las fuentes bloqueadas;
- b) las llamadas sin éxito dan lugar a intentos renovados con una probabilidad dada.

El supuesto a) lleva a modelos de tráfico simple.

Para sistemas de espera se puede asumir que:

- a) las llamadas sin éxito esperan hasta ser servidas;
- b) las llamadas sin éxito son abandonadas con una probabilidad dada;
- c) las llamadas sin éxito esperan por un tiempo máximo dado.

El supuesto a) lleva a los modelos de tráfico más simples.

Disciplina de la disposición en cola

Dependiendo de cómo las llamadas de espera se aceptan fuera de la cola en un sistema de espera, debe hacerse una distinción entre los siguientes métodos de disposición en cola:

- a) fila ordenada (primero en llegar, primero en ser servido);
- b) fila aleatoria: cada llamada en espera tiene la misma probabilidad de ser la próxima en ser servida, independientemente de cuánto tiempo ha estado esperando;
- c) disposición en cola prioritaria: cada llamada en fila tiene un número de prioridad dependiendo de donde viene o dependiendo de la entrada en la que está esperando. La llamada con la prioridad más alta se sirve primero.

El método a) usualmente lleva a los modelos matemáticos más simples, mientras que c) puede ser el método más usado en la práctica.