Conceptos y Elementos Básicos de Tráfico en Telecomunicaciones

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Definiciones de Tráfico en Telecomunicaciones

De acuerdo con la recomendación ITU-T B.18: en una red de telecomunicaciones, la intensidad de tráfico instantánea *A(t)* en un conjunto de elementos de red es el número de elementos ocupados en un instante dado. Por lo general se considera 1 hora.
Pueden calcularse momentos estadísticos para un periodo de tiempo dado; por ejemplo, la intensidad de tráfico media está relacionada con la intensidad de tráfico instantánea *A*(*t*) por la siguiente expresión:

$$\overline{A}(t_1,t_2)=\frac{1}{t_2-t_1}\int_{t_1}^{t_2} A(t)\mathrm{d}t$$

En las aplicaciones, el término intensidad de tráfico generalmente se emplea en este sentido de intensidad de tráfico media.
La intensidad de tráfico equivale al producto de la tasa de llegadas por el tiempo medio de ocupación.
La unidad de intensidad de tráfico empleada habitualmente es el *erlang* cuyo símbolo es E.

Erlang: Unidad de intensidad de tráfico, cuyo símbolo es E. Un erlang es la intensidad de tráfico en un conjunto de órganos, cuando sólo uno de ellos está ocupado de manera continua. Cuando el tráfico es de un (1) erlang significa que el elemento de red está totalmente ocupado durante el tiempo de medición, normalmente una hora.

NOTA – El CCIF ("Le comité consultatif international des communications téléphoniques à grande distance“predecsor del CCITT y del ITU-T) dio en 1946 el nombre de «erlang» a la unidad de tráfico, en honor del matemático danés A.K. Erlang (1878-1929), que fue el fundador de la teoría del tráfico en telefonía.

Si una línea está ocupada durante una hora entonces cursa un tráfico de 3600 llamadas-segundos que a 36 llamadas de 100 seg de duración cada una, o a cualquier otra combinación que resulte en 3600 llamadas-segundo. Si 100 usuarios solicitan una llamada con una duración promedio de 3 minutos entonces el tráfico es

$$A=\frac{100*3*60}{3600}=\frac{100*180}{3600}=5\,\text{erlang}$$

A parte del erlang también se usa el CCS (Centi Call Seconds) como unidad de tráfico. 1 CCS equivale a 100 llamadas-segundos, por lo tanto el tráfico en una línea ocupada totalmente durante una hora es de 36 CCS, por lo tanto:

$$1\,\text{erlang}=36\,\text{CCS}$$

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Definiciones de Tráfico en Telecomunicaciones

- Tráfico ofrecido
  - Tráfico que podría cursar una cantidad muy grande de elementos de red. Es el tráfico que se cursaría si no hubiesen llamadas perdidas
- Tráfico cursado
  - Es el tráfico atendido por un grupo de elementos de la red
- Tráfico de desbordamiento
  - La parte del tráfico ofrecida a un conjunto de elementos de red que no es cursada por dicho conjunto de órganos
- Tráfico Bloqueado
  - La parte del tráfico de desbordamiento que no es cursada por conjuntos subsiguientes de órganos
- Tráfico rechazado o perdido
  - La parte de tráfico bloqueado que no da como resultado reintentos de llamada. Es la diferencia entre el tráfico ofrecido y el tráfico cursado y se puede reducir aumentando la capacidad del sistema

El tráfico ofrecido es un concepto teórico y se utiliza sólo para propósitos de planificación teórica. Sólo el tráfico cursado es medida en la práctica y por supuesto depende de la capacidad de la red o sistema. Desde el punto de vista económico, la capacidad de la red siempre será menor que el tráfico ofrecido, esto se debe al carácter aleatorio de las comunicaciones lo que produce que en condiciones normales de funcionamiento sólo un porcentaje de los usuarios de la red solicitan recursos a la misma.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Unidad de Tráfico

La unidad de intensidad de tráfico se denomina Erlang, cuyo símbolo es E. Un erlang es la intensidad de tráfico en un conjunto de órganos, cuando sólo uno de ellos está ocupado de manera continua. Cuando el tráfico es de un (1) erlang significa que el elemento de red está totalmente ocupado durante el tiempo de medición, normalmente una hora.

Si una línea está ocupada durante una hora entonces cursa un tráfico de 3600 llamadas-segundos que a 36 llamadas de 100 seg de duración cada una, o a cualquier otra combinación que resulte en 3600 llamadas-segundo. Si 100 usuarios solicitan una llamada con una duración promedio de 3 minutos entonces el tráfico es:

Cantidad de llamadas

Duración en s de cada llamada

$$A = \frac{100 * 3 * 60}{3600} = \frac{100 * 180}{3600} = 5 \text{ erlang}$$

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Centi-Call Seconds - CCS

A parte del Erlang, también se usa el CCS (Centi-Call Seconds) como unidad de tráfico.

1 CCS equivale a una llamada con duración de cien segundos, de esta forma el tráfico en una línea ocupada totalmente durante una hora es de 36 CCS, por lo tanto:

$$1\,\text{erlang} = 36\,\text{CCS}$$

De esta forma si tenemos el tráfico en CCS, lo dividimos por 36 para obtener los Erlang respectivos. Si por el contrario, el tráfico está en Erlang lo multiplicamos por 36 para llevarlo a CCS.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Cálculo de Tráfico

El tráfico ofrecido depende de dos factores importantes

1.- La tasa de llegada de sesiones de comunicaciones Q [sesiones/s, sesión/min, sesión/hr]

2.- La duración promedio de cada sesión μ [s o min]

Esto se aplica por igual para llamadas de voz o para aplicaciones de datos

Si Q se expresa en sesión/min y μ en min, el tráfico promedio en erlang viene dado por

$$A = \mu Q$$

Ejemplo: Si en una red llegan 10 llamadas por mín y cada una dura en promedio 3 min, entonces el tráfico promedio ofrecido a la red es de 30 erlang.

$$A = 10[llamadas/\min]\,3[\min] = 30\,erlang$$

Si Q se expresa en sesión/hr y μ en segundos entonces el tráfico es

$$A = \frac{\mu Q}{3600}$$

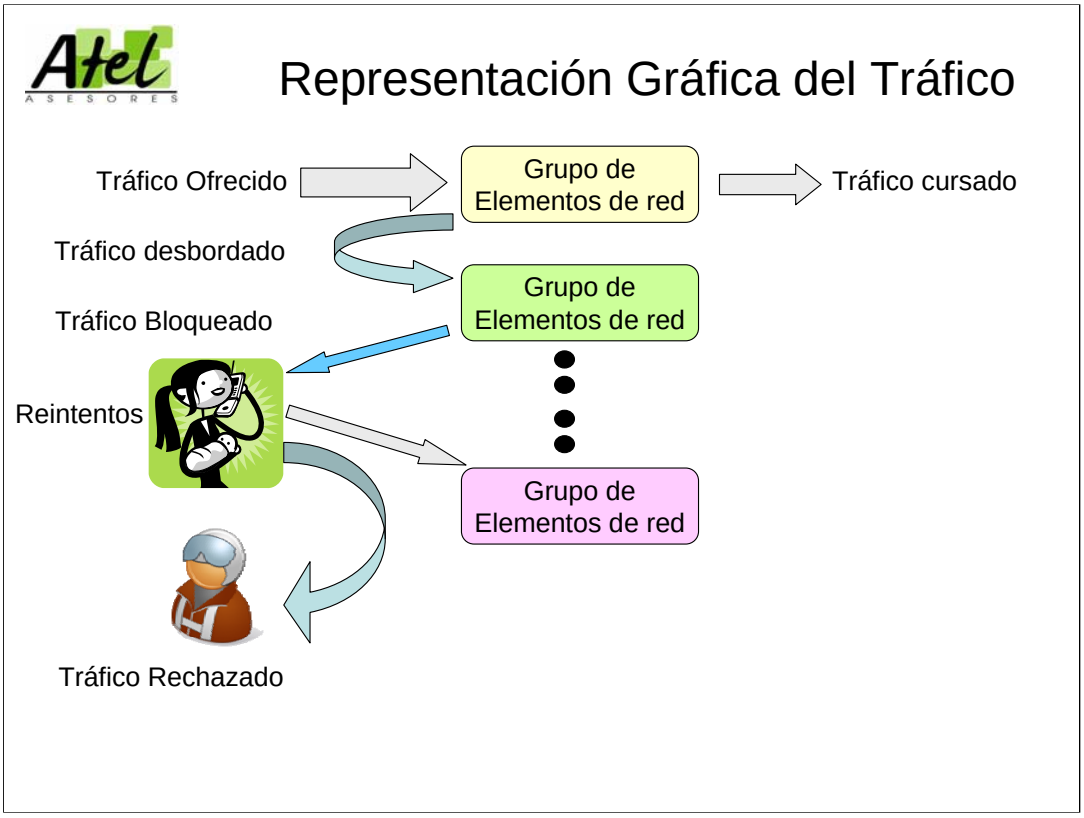
A pesar de que el erlang es la unidad de tráfico más usada, existen otras unidades de tráfico.

El cálculo de tráfico en telecomunicaciones, también conocido como teletráfico, es un compromiso entre la cantidad de recursos disponibles pero no utilizados por los usurarios, y la misma cantidad de recursos cuando todos los usuarios los soliciten, manteniendo al mínimo la cantidad de sesiones perdidas.

La estimación de la cantidad de recursos no tiene una solución única ya que depende mucho del planificar y del grado de servicio que este dispuesto a a ofrecer a los clientes.

Un canal ocupado continuamente lleva un tráfico de un Erlang.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

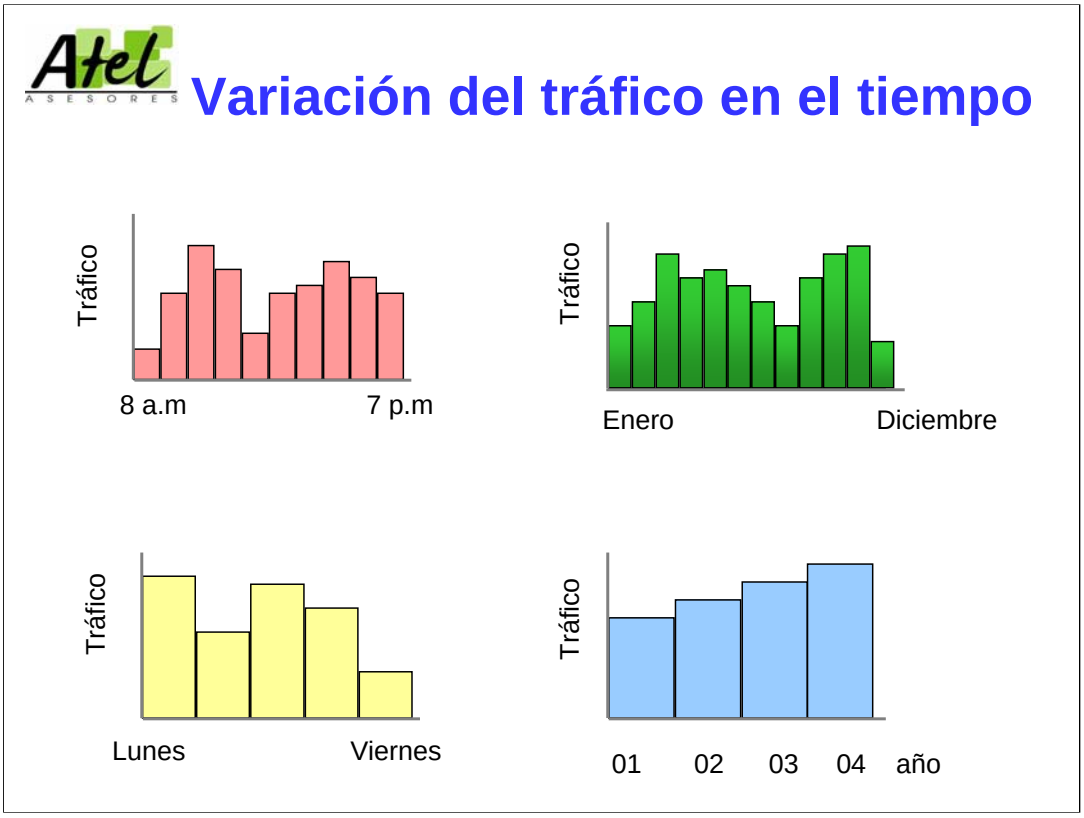


Los costos de una red de telecomunicaciones se pueden dividir en dos grandes partes: los costos de la infraestructura que depende de la cantidad de usuarios y los costos asociados al tráfico cursado por los usuarios.

La planificación tiene por objeto minimizar los costos manteniendo un cierto grado en la calidad del servicio a los usuarios, para ello al red debe estar bien dimensionada y se deben medir periódicamente una gran cantidad de indicadores que señalan el estado de ocupación de la red.

Cuando un elemento se ocupa, las aproximas llamadas son desbordadas a otros elementos, se produce entonces el tráfico desbordado. Cuando todos los elementos está ocupados las siguientes llamadas se bloquean. Una cierta cantidad de las llamadas bloqueadas son reintentadas, las que no lo sean se consideran como rechazadas.

Dependiendo del modelo de tráfico que usemos se pueden o no considerar reintento, si no aplica el reintento entonces todas la llamadas bloqueadas se convierten en rechazadas o perdidas.



El tráfico varía en función de diversos factores como la ubicación geográfica (urbano, suburbano, rural), si es residencial o comercial, y también varía con el tiempo. En la gráfica se muestra la variación del tráfico durante el día, la semana, un año y en varios años.

Distribución Estadística del Tráfico

- El tráfico tiene un comportamiento aleatorio por lo que es necesario el uso de distribuciones de probabilidad para analizar:
  - el patrón de llegada de sesiones [sesion/s]
  - la duración por sesión [s/sesion]
- Patrón de llegada de llamadas
  - Patrón de llegada aleatorio: Poisson
  - Patrón con picos: hyper-exponencial
  - Patrón plano: hypo-exponencial
- Distribución exponencial para la duración de las sesiones

El patrón de llegada de llamadas tiene un gran impacto en el tráfico total, por lo tanto debe escogerse con sumo cuidado.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Distribución de Poisson

- Esta distribución es discreta y se define sólo a través del parámetro μ que representa la media, y es muy útil para caracterizar eventos que ocurren en el tiempo: solicitud de llamadas, cantidad de clientes que entran a un establecimiento, cantidad de tareas que requieren CPU o I/O interface, etc.

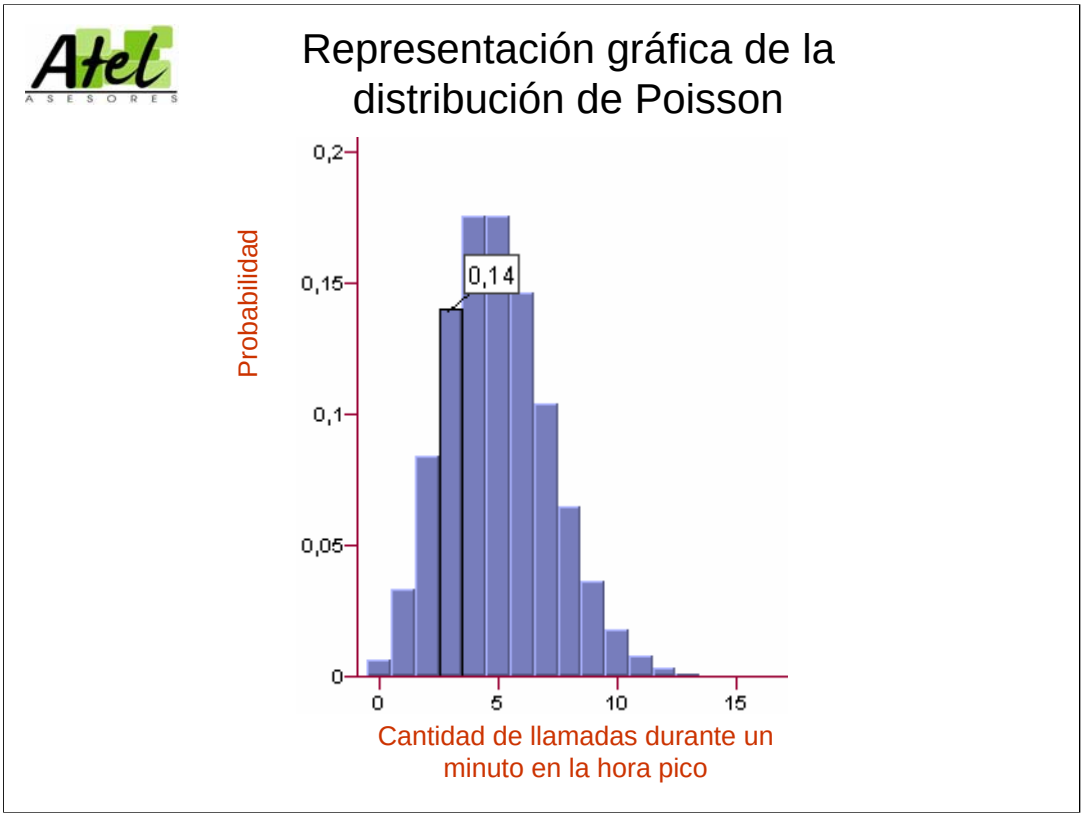
- Sea X una variable aleatoria que denota la ocurrencia de un evento, entonces la probabilidad que X sea igual *x,* donde *x* es un entero, viene dada por

$$P(X = x) = \frac{\mu^x e^{-\mu}}{x!}; x = 0,1,2,3,.....$$

En la distribución de Poisson la media y la varianza son iguales a μ. Por lo general es la distribución que se usa para representar las solicitud de sesiones de comunicaciones en las redes de telecomunicaciones.

El valor de μ se puede obtener a través de medias realizadas, modernamente todos los elementos de una red de telecomunicaciones incorporan herramientas que permiten obtener algunas estadísticas sobre la tasa de llegada y la duración promedio de las sesiones. Esos valores se utilizan entonces para realizar ajustes que mejoren el desempeño de las redes.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739



Esta gráfica la probabilidad de que se produzca una cantidad x de llamadas, por ejemplo durante un minuto en la hora pico; en el caso aquí mostrado la media μ=5, lo que equivale a 300 llamadas por hora.

En la gráfica vemos que la probabilidad de que ocurran 3 llamadas en un minuto es de 0.14. Por ejemplo, esta gráfica nos indica que la probabilidad de que se produzcan 10 llamadas por minuto es muy baja, por lo tanto a nivel de establecer la cantidad de líneas telefónicas, es preferible diseñar la red para satisfacer la cantidad promedio de llamadas, y así los costos son menores.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Distribución de Poisson (cont.)

- La distribución de Poisson supone que:
  - todos los intentos de llamada son independientes entre sí, por lo tanto ningún evento producirá una generación abrupta de las mismas
  - durante la hora pico las llamadas se distribuyen en forma aleatoria y que la probabilidad de ocurrencia de una llamada es la misma durante toda la hora pico
  - la cantidad total de llamadas es mucho mayor que 1, por lo que una sola llamada es insignificante en el patrón total de tráfico.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

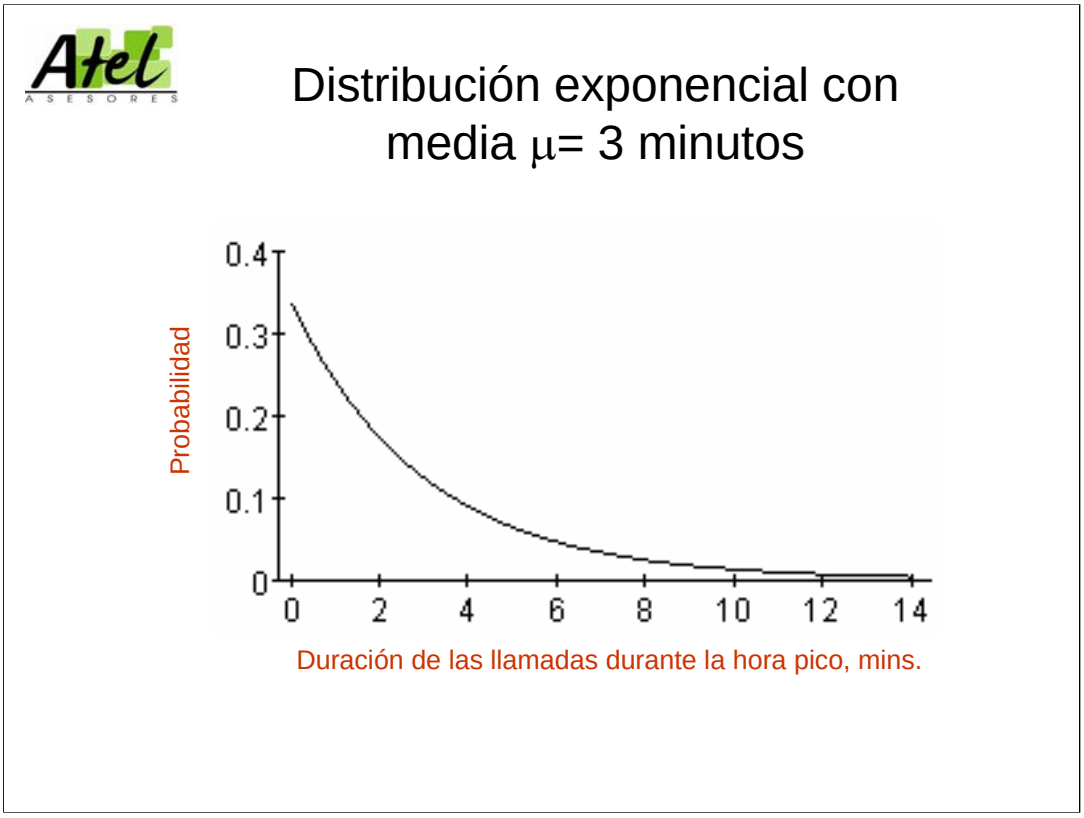
Duración de las llamadas

- Por lo general la duración de las llamadas es breve, para el caso de telefonía vocal, por lo que se considera que tienen un comportamiento como el de una distribución exponencial negativa
- La distribución exponencial. Interpretación: la cantidad de llamadas de duración *x* viene dada por:

$$P(X = x) = \frac{1}{\mu} e^{-x/\mu}$$

La probabilidad de que la duración de las llamadas sea *x*, viene dada por P(X=*x*). La media y la varianza son iguales a μ.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739



En este caso se presenta la Densidad de distribución de probabilidad de la distribución exponencial en función de la duración de las llamadas o sesiones de comunicación. El valor máximo se obtiene para x=0, y es el inverso de la media.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Grado de Servicio

Es un atributo de calidad de servicio usado en la comunicaciones telefónicas en particular, y en general en los servicios basados en conmutación de circuitos, y se refiere a la probabilidad de bloqueo en el primer intento de una llamada, durante la hora pico, y se expresa como Px donde x es menor que 1 y representa la probabilidad de bloqueo; por ejemplo P.01 significa que existe un 1% de probabilidad de bloqueo en la hora pico. Mientras más bajo es el grado de servicio es menor la probabilidad de bloqueo y por supuesto mejor el desempeño de la red. La mayoría de las redes de telefonía fija se diseñan para P.01 y las redes celulares para P0.02

El bloqueo ocurre cuando estando todos los recursos ocupados se trata de hacer una llamada la cual no puede ser atendida por la red. Cuando todas las líneas conectadas entre los nodos de la red están ocupadas también ocurre el bloque, si se trata de hacer una llamada a un abonado de otra central.

No es necesario incluir equipos de conmutación y líneas para manejar simultáneamente todos las solicitudes hechas por todos los usuarios conectados a la red, ya que en condiciones normales sólo un pequeño porcentaje del tráfico es demandado a la red. Entonces en el diseño debe haber una situación de compromiso entre la calidad de servicio y el costo asociado a los elementos de la red.

En el caso de las redes celulares la probabilidad de bloque incluye tanto los que están entrando a un sector de la BS como los que están haciendo handoff.

Teoría de Colas

- La teoría de colas tiene como objeto estudiar el problema que se presenta cuando se tiene un centro de servicio con una capacidad limitada, y un conjunto de clientes, a ser servidos, quienes llegan en forma aleatoria al centro de servicios, cuando éste está saturado el próximo cliente se envía a una lista de espera
- La solución al problema consiste en dimensionar la capacidad el centro de servicios y el tamaño de la lista de espera
- Se pueden identificar tres elementos principales
  - El centro de servicio
  - La cantidad de clientes
  - La cola de espera
- La teoría de colas debe dar como respuesta parámetros claves
  - El tiempo promedio en la cola de espera
  - El tiempo de respuesta del servicio (tiempo de espera+obtención del servicio)
  - Distribución de clientes en las colas
  - Prioridad, si existe, en las colas de espera
- Varios modelos pueden proponerse para resolver esta situación
- Ejemplo:
  - Dimensionamiento de un servidor de e-mails en función de la cantidad de usuarios, la tasa de llegada de correos y la duración promedio de los mismos.

La teoría de colas es una rama de las probabilidades aplicada. Su campo de aplicación es muy extenso, entre otros podemos mencionar: redes de telecomunicaciones, sistemas de computadores, control de tráfico vial, etc.

Está basada en modelos estadísticos para determinar el patrón de llagada de los clientes al centro de servicios y la duración promedio de cada uno en ser atendido.

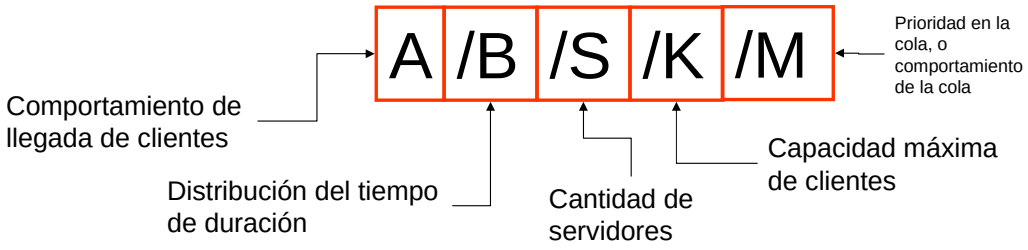
En teoría de colas el término cliente tiene un sentido muy general y puede referirse a personas que entran en establecimiento, a tareas por realizar, llamadas telefónicas, carros en la vías, y en general a cualquier agente que necesite servicios de un proceso.

La formación de colas es una manifestación clara de que el sistema no tiene capacidad para servir a todo los usuarios que pertenecen al sistema. Sin embargo, diseñar los sistemas para satisfacer todos sus usuarios generalmente resulta muy costoso, por lo que se debe establecer un compromiso entre la capacidad el sistema, la cantidad de usuarios y los lapsos de espera para obtener el servicio.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Modelos de teoría de colas

- Los diferentes modelos de teoría de colas pueden distinguirse por la manera como tratan a los clientes y se identifican por la notación de Kendall: A/B/S/K
  - A: comportamiento de llagada de los clientes. Si A=M es un proceso de Poisson
  - B: Distribución de la duración del tiempo de servicio. M si es exponencial, D si es determinística y G para una distribución general.
  - S: Cantidad de servidores
  - K: Capacidad máxima de clientes entre los servidores y la cola de espera.
- Casos particulares
  - Si la cola de espera es infinita se puede omitir el cuarto parámetro, tendremos un modelo A/B/S y se denomina sistema con retraso o sistema sin pérdidas
  - Si K=S (A/B//S/S) tenemos un sistema con pérdidas, ya que no tenemos cola de espera



La cola se produce porque la cantidad de servidores es insuficiente para servir a todos los clientes que llegan.

La prioridad en la cola obedece a 4 modalidades

• **FIFO** (first in first out): sin prioridad

• **LIFO** (last in first out): cuando un cliente termina su trabajo, se da paso al último de la cola

• **SJR** (short job first): prioridad a los servicios que duran menos

• **RR** (round robin): A cada cliente se le asigna un tiempo fijo para realizar su tarea en forma ciclica

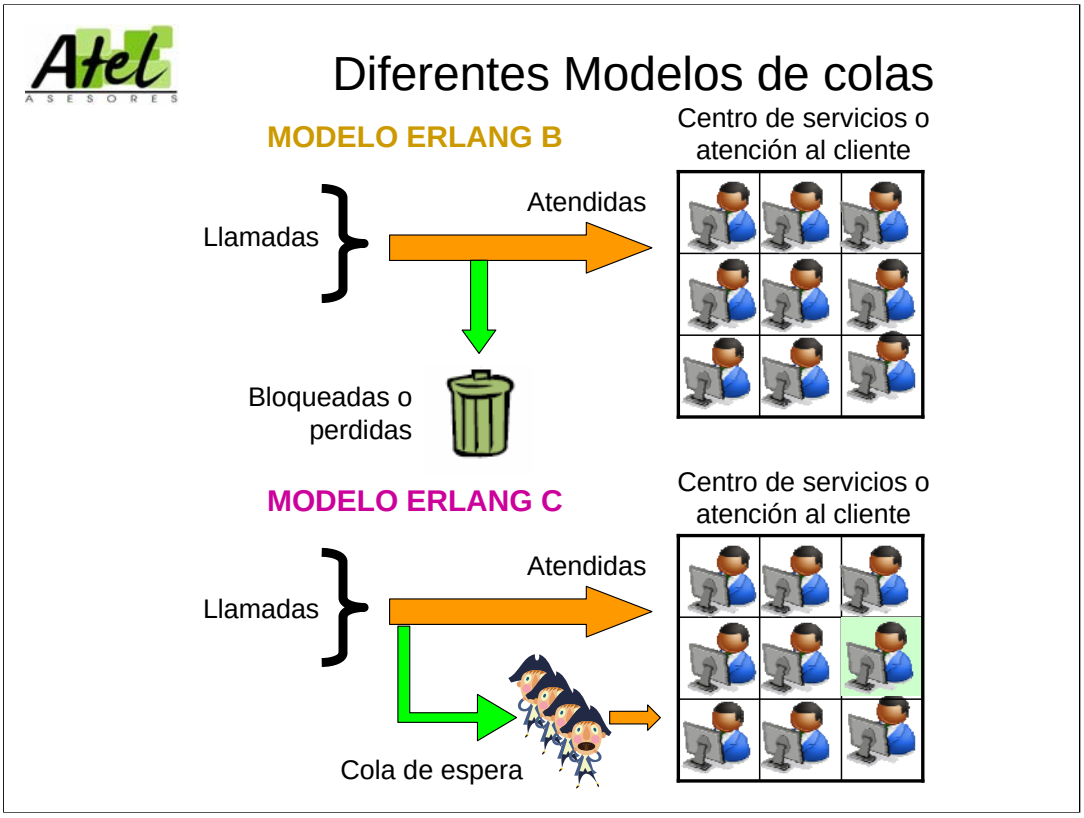
Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Modelo de Tráfico

- Los modelos de tráfico están basados en la teoría de colas
- Una vez que conocemos
  - la tasa de llegada de sesiones
  - la duración de cada sesión
  - y el grado de servicio
  - Debemos escoger un modelo de tráfico
- No existe ningún modelo que se ajuste exactamente a la realidad, pero debemos escoger uno que que se adapte lo más posible a la situación que estamos estudiando.
- Los modelos más usados son
  - Erlang B
  - extended Erlang B
  - y Erlang C
  - Todos suponen un patrón de llegada de llamadas aleatorio y la duración según una distribución exponencial. También consideran que la cantidad de fuentes que pueden originar una llamada es infinita.

Los costos de una red de telecomunicaciones se pueden dividir en dos grandes partes: los costos de la infraestructura que depende de la cantidad de usuarios y los costos asociados al tráfico cursado por los usuarios.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739



Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

El modelo Erlang B M/M/N/N

Conociendo el tráfico y la cantidad de líneas disponible, este modelo calcula la probabilidad P_B de que una llamada en su primer intento sea bloqueada y está basado en las siguientes premisas

1. La cantidad de usuarios es muy grande
2. Las llamadas llegan en forma aleatoria de acuerdo a una distribución de Poisson
3. Las llamadas se atienden según el orden de llegada
4. Las llamadas bloqueadas se pierden. Modelo con pérdidas, no hay lista de espera.
5. El tiempo de duración de las llamadas siguen una distribución exponencial

$$P_B = \frac{\frac{A^N}{N!}}{\sum_{i=0}^{N} \frac{A^i}{i!}}$$

Donde

A: es el tráfico en erlang durante la hora pico

N: cantidad de líneas del sistema

El modelo Erlang B ha sido ampliamente usado en el diseño de redes y sus resultados se encuentra tabulados pero son algo engorroso de usar; dichas tablas están diseñadas de manera que conociendo el tráfico en erlangs y el grado de servicio o probabilidad de bloqueo se pueda obtener la cantidad de líneas mínima para cursar el tráfico. Actualmente existen muchos sitios en Internet que permiten hacer dichos cálculos.

Ejemplo: Si el tráfico es de 5 erlang y el grado de servio es P0.01, entonces la cantidad de líneas necesarias es 11, con lo que obtiene un grado de servicio P0.00829.

Intensity	GOS	Trunks	Actual GOS
5 E	0.01	11	0.00829

Este es el modelo usado para el dimensionamiento de centrales telefónicas POTS tanto públicas como privadas También se usa para el caso de VoIP, ya que se espera que VoIP tenga la misma QoS de POTS.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Extracto de las tablas de Erlang.
B es la probabilidad de bloqueo en %

N/B	0.01	0.05	0.1	0.5	1.0	2	5	10
1	.0001	.0005	.0010	.0050	.0101	.0204	.0526	.1111
2	.0142	.0321	.0458	.1054	.1526	.2235	.3813	.5954
3	.0868	.1517	.1938	.3490	.4555	.6022	.8994	1.271
4	.2347	.3624	.4393	.7012	.8694	1.092	1.525	2.045
5	.4520	.6486	.7621	1.132	1.361	1.657	2.219	2.881
6	.7282	.9957	1.146	1.622	1.909	2.276	2.960	3.758
7	1.054	1.392	1.579	2.158	2.501	2.935	3.738	4.666
8	1.422	1.830	2.051	2.730	3.128	3.627	4.543	5.597
9	1.826	2.302	2.558	3.333	3.783	4.345	5.370	6.546
10	2.260	2.803	3.092	3.961	4.461	5.084	6.216	7.511
11	2.722	3.329	3.651	4.610	5.160	5.842	7.076	8.487
12	3.207	3.878	4.231	5.279	5.876	6.615	7.950	9.474
13	3.713	4.447	4.831	5.964	6.607	7.402	8.835	10.47
14	4.239	5.032	5.446	6.663	7.352	8.200	9.730	11.47
15	4.781	5.634	6.077	7.376	8.108	9.010	10.63	12.48

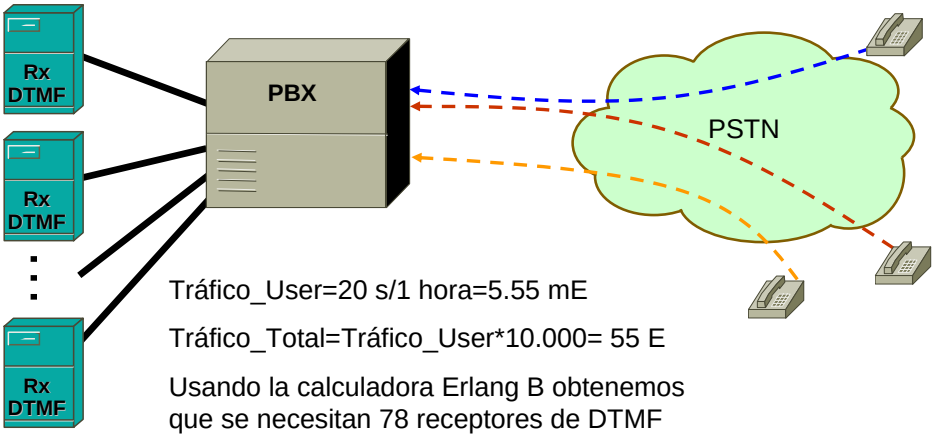
Ejemplo: Tráfico de 5 E y P0.01, Cuantos circuitos se necesitan?

Primero nos ubicamos en la columna correspondiente a P0.01, es decir B=1% y luego nos desplazamos hasta ubicar el tráfico en erlang que sea igual o inmediatamente superior a 5 E, en este caso 5.160, y así leemos en la columna de la izquierda la cantidad de 11 líneas necesarias para cursar el tráfico.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Ejemplos usando Erlang B

Consideremos un centro de atención al cliente ON LINE vía telefónica a través de tonos DTMF. El tiempo completo de la fase de marcado de un digito DTMF se asume a ser de 20 s, incluye la invitación a marcar y el anuncio de todas las opciones. Se esperan 10.000 llamadas en la hora pico y se requiere que el porcentaje de bloqueo sea inferior a 0.1%=0.001 . Deseamos calcular la cantidad de receptores de DTMF que se necesitan.



Tráfico_User=20 s/1 hora=5.55 mE

Tráfico_Total=Tráfico_User*10.000= 55 E

Usando la calculadora Erlang B obtenemos que se necesitan 78 receptores de DTMF

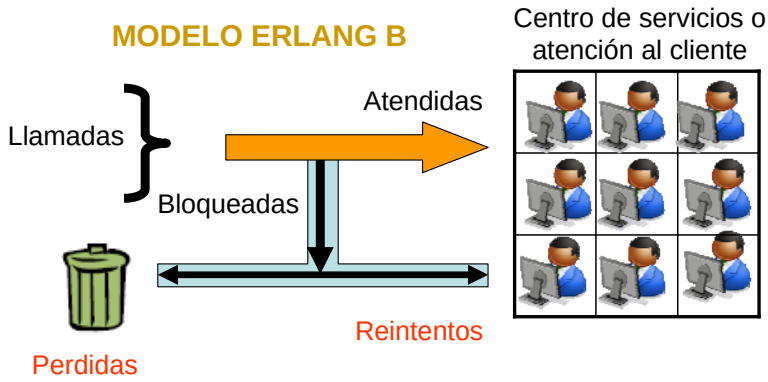
En este caso las llamadas bloqueadas se pierden ya que no se considera ningún sistema de colas de espera. Si los 78 receptores de DTMF están ocupados existe una probabilidad del 0,1% que la siguiente llamada se pierda.

En internet existen varios sitios que tienen implementado una calculadora de Erlang B, por ejemplo http://www.erlang.com/calculator/erlb/.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

El modelo Extended Erlang B

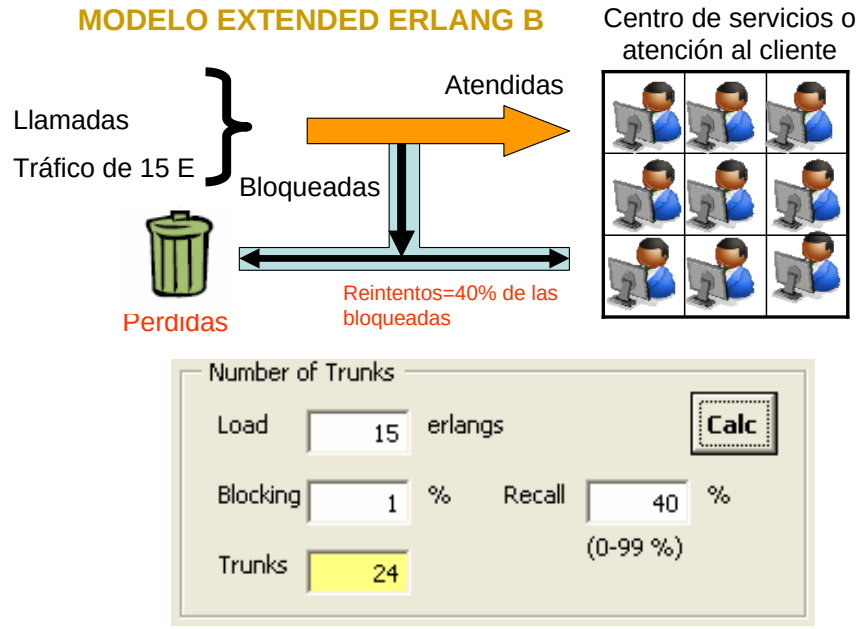
En este modelo se asume que un porcentaje de las llamadas bloqueadas son reintentadas de nuevo y el otro porcentaje se pierde.



En este modelo de tráfico un porcentaje de las llamadas bloqueadas son reintentadas por usuarios insistentes e ingresan de nuevo al sistema, el porcentaje restante se convierte en llamadas perdidas. Es una modificación del modelo Erlang B. Es un modelo más realista ya que muchos usuarios vuelven a intentar la llamada hasta que sean atendidos.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

El modelo Extended Erlang B



En la página

www.infotel-systems.com/Downloads/trafcalc_with_instructions.xls está una calculadora Extended Erlang B.

Necesitamos conocer el Tráfico total, la probabilidad de bloqueo y el porcentaje de las bloqueadas que reintentan.

Sin embargo, no existe ninguna formula para determinar el porcentaje de reintento. Se puede aplicar la siguiente regla: si no se tiene ninguna idea usar 50%, y si cree que todos los bloqueados reintentaran use 99%.

El modelo Erlang C M/M/N/∞/FIFO

El modelo Erlang C se usa para calcular la probabilidad que una llamada sea colocada en la cola de espera y permite dimensionar un grupo de servidores bajo el esquema de colas FCFS (First Come First Serve) o FIFO, donde las llamadas en espera se van atendiendo a medida que los servidores se van liberando. Se asume que:

- La cantidad de fuentes es muy grande
- Las llamadas llegan en forma aleatoria de acuerdo a una distribución de Poisson
- Las llamadas se atienden según el orden de llegada
- Las llamadas bloqueadas se colocan en una cola de espera de tamaño infinito
- El tiempo de duración de las llamadas es de acuerdo a una distribución exponencial

La probabilidad de que una llamada sea retrasada y puesta en la cola de espera es:

$$P(>0) = \frac{\dfrac{A^N N}{N!(N-A)}}{\displaystyle\sum_{i=0}^{N-1} \frac{A^i}{i!} + \frac{A^N N}{N!(N-A)}}$$

Donde

N: cantidad total de servidores

A: Tráfico ofrecido al grupo de servidores, en erlang

La formula Erlang C permite calcular la probabilidad de que un llamada sea retrasada y por lo tanto puesta en la cola de espera.

Es el modelo usado para el dimensionamiento de los call center a fin de determinar la cantidad de agentes necesarios para garantizar una cierta calidad de servicio. La llamadas entrantes son respondidas inmediatamente si hay agentes disponibles, sino son enviadas a una cola de espera. No hay prioridad y las llamadas se atienden según el patrón FIFO. Se supone que los usuarios nunca abandonan la cola, salvo una vez que son atendidos. También es conocido como un sistema sin pérdidas.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

El modelo Erlang C M/M/N/∞/FIFO

El tiempo promedio de retraso o de espera en la cola, con una duración promedio de una llamada de μ, es:

$$D = P(>0) * \frac{\mu}{(N - A)}$$

La probabilidad de que el retraso sea superior a un tiempo t

$$P(delay > t) = P(delay > 0)e^{\frac{-(N-A)t}{\mu}}$$

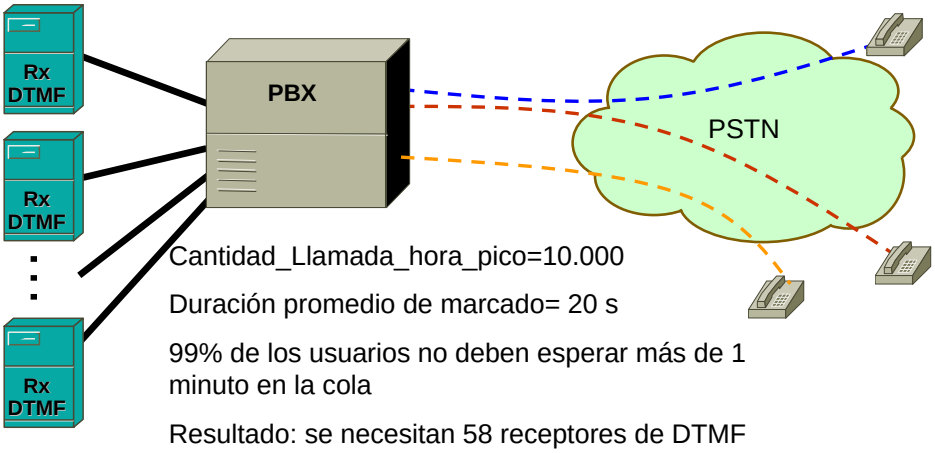
Estas tres ecuaciones pueden usarse para resolver el siguiente problema: obtener la cantidad de agentes N para que un cierto porcertanje x de las llamadas sea atendidas en menos de T minutos, si cada llamada dura en promedio D minutos y se reciben M llamadas en la hora pico.

En internet hay muchos sitios con calculadoras Erlang C donde se pueden hacer las diversas combinaciones de la variables ya mencionadas. En general existen 5 parámetros y debemos conocer 4 de ellos.

1. Cantidad de llamadas en la hora pico
2. Duración promedio de las llamadas en la hora pico. El producto de la cantidad de llamadas por su duración es la carga o tráfico ofrecido a la red en erlangs
3. Cantidad de agentes o servidores
4. Tiempo promedio de retardo. Es el tiempo promedio de retardo considerando todas las llamadas, tanto las que se atienden inmediatamente como las que se retardan.
5. Porcentaje de llamadas que debe ser atendido en un tiempo inferior t_d

Ejemplos usando Erlang C

Vamos a resolver el ejemplo anterior pero considerando que existe una cola de espera para ser atendido. Se asume que el 99% de los usuarios no debe esperar más de 1 minuto en la cola.



Cantidad_Llamada_hora_pico=10.000

Duración promedio de marcado= 20 s

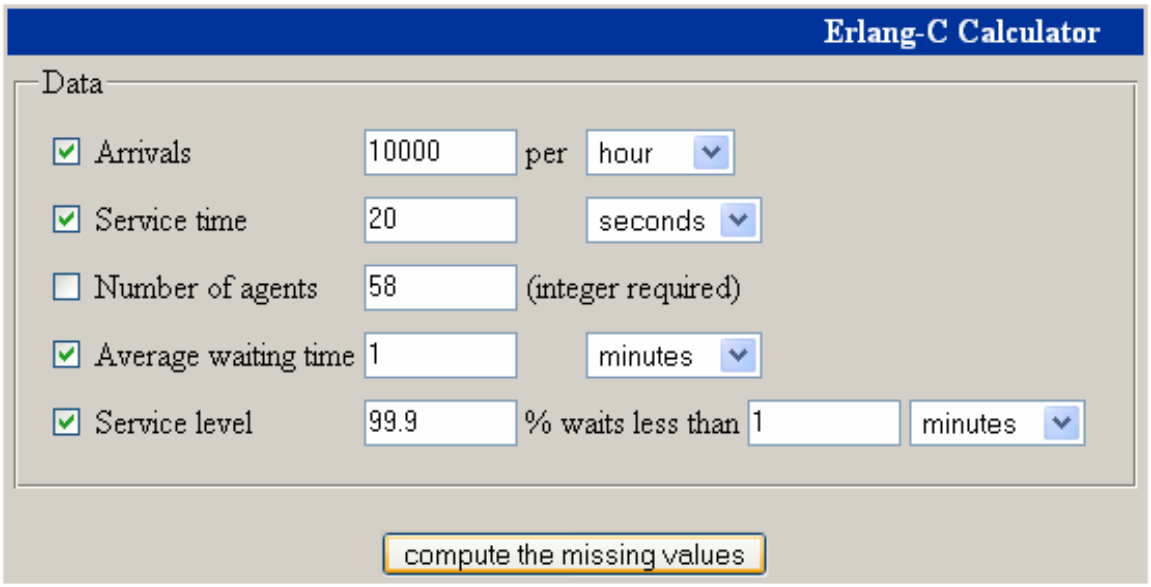
99% de los usuarios no deben esperar más de 1 minuto en la cola

Resultado: se necesitan 58 receptores de DTMF

Los resultados fueron obtenidos de la siguiente dirección: http://www.math.vu.nl/~koole/ccmath/.

Vemos que se necesitan menos receptores que en el caso de Erlang B, la cantidad de recursos es menor pero los usuarios deben esperar como máximo un minuto para ser atendidos. La reducción es de un 25.64%.

Service Time: Es el tiempo durante el cual se usan los recursos de la red.

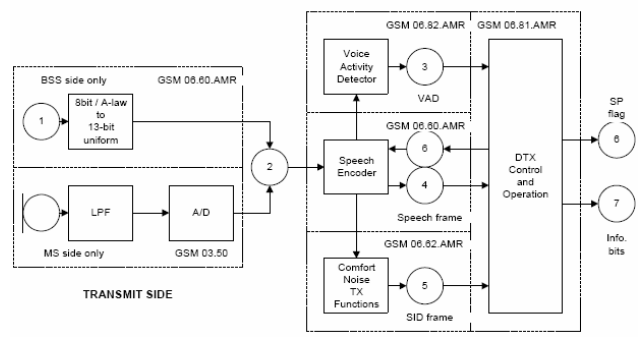


Modelo de Tráfico para VoIP

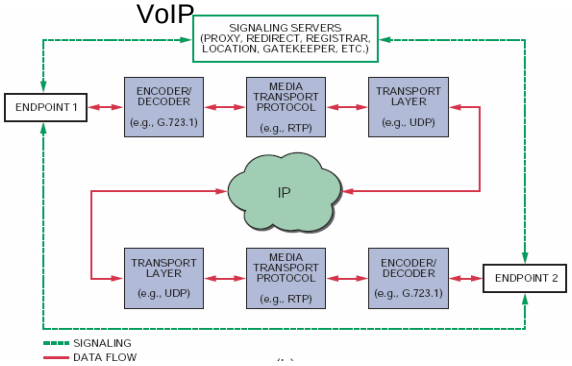
Generalmente cuando hablamos de VoIP, casi instintivamente pensamos en la trasmisión de voz sobre Internet; sin embargo, como su nombre lo indica, VoIP se refiere a la transmisión de voz paquetizada cuyos paquetes usan el protocolo IP en capa 3. La red de transporte puede ser una red privada, o por supuesto internet. Cuando usamos VoIP en Internet no podemos garantizar la calidad de servicio, la QoS se logra sólo si tenemos una red privada con control de los recursos.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Esquema General de un Sistema de VoIP



Procesos realizados en VoIP



Generalmente cuando hablamos de VoIP, casi instintivamente pensamos en la trasmisión de voz sobre Internet; sin embargo, como su nombre lo indica, VoIP se refiere a la transmisión de voz paquetizada cuyos paquetes usan el protocolo IP en capa 3. La red de transporte puede ser una red privada, o por supuesto internet. Cuando usamos VoIP en Internet no podemos garantizar la calidad de servicio, la QoS se logra sólo si tenemos una red privada con control de los recursos.

Diferentes Aplicaciones

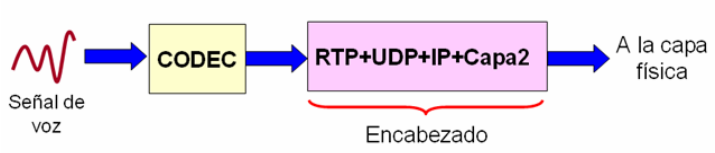
Class	Application	Bandwidth Guideline		Latency Guideline		Jitter Guideline	
1	Multiplayer Interactive Gaming	Low	50-85 kbps	Low	< 150 msec	Low	<100 msec
2	VoIP & Video Conference	Low	4-384 kbps	Low	< 150 msec	Low	<50 msec
3	Streaming Media	Low to High	5 kbps to 2 Mbps	N/A		Low	<100 msec
4	Web Browsing & Instant Messaging	Moderate	10 kbps to 2 Mbps	N/A		N/A	
5	Media Content Downloads	High	> 2 Mbps	N/A		N/A	

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

LA CODIFICACIÓN DE LA VOZ

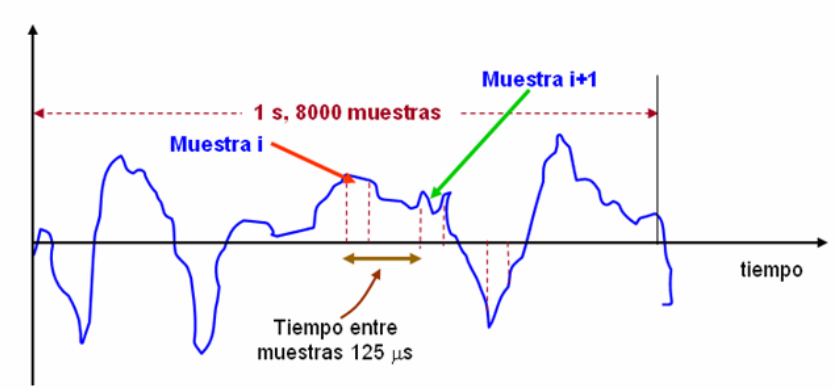
La primera fase en la arquitectura de VoIP es la codificación de la voz, que consiste en transformar la voz analógica en paquetes de voz por medio de un CODEC para luego enviarla a las capas inferiores, este procese se puede observar en la Fig. 3.2, tanto el CODEC como los protocolos de los protocolos están estandarizados y todas sus características son conocidas.

Hay una gran cantidad de técnicas de codificación de la voz cada uno con diferentes requerimientos en cuando a la tasa de bits (G.711, G.722, G.722.1, G.723.1, G.728, G.729, y AMR); incluyendo el overhead de los protocolos, las tasas de bits pueden estar entre 5 Kbps y 64 Kbps en cada sentido, es decir para el DL y par el UL.



EL CODEC

Cualquiera que sea el CODEC siempre realiza una serie de funciones claves de acuerdo con sus propios parámetros. La frecuencia de muestreo es 8 KHz, por lo que se toman 8000 muestras en un segundo, ver. Fig. 3.3



Un CODEC tiene definido los siguientes parámetros:

- Duración de la Trama: T_{trama}
- Cantidad de muestras en cada trama: $M_{muestras}$
- Cantidad de tramas en el payload: N_{tramas}
- Cantidad de bits que contiene cada muestra: típicamente 8 bits (1 Byte)

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

La Trama

Una trama está formada por varias muestras cuya cantidad depende de cada CODEC.
En un tiempo igual a la duración de una trama existirán $M_{muestras}$*8 bits.
El payload que produce el CODEC está formado por N_{tramas} comprimidas usando un factor de compresión FC. De manera que la cantidad total de bits contenida en el payload es:

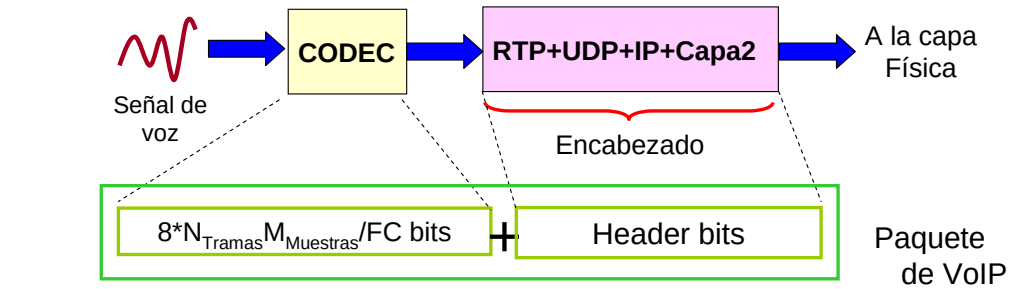
$$Payload = \frac{8 * N_{tramas} * M_{muestras}}{FC}$$

El payload es la carga útil, para que el mismo sea transportado es necesario agregarle los encabezados de las capas inferiores, como ya mencionamos estos encabezados tienen un tamaño determinado que lo llamaremos *Header*. La suma del Payload y del Header producen el paquete, a continuación se muestra la ecuación del tamaño del paquete en bits.:

$$Paquete = Payload + Header = \frac{8 * N_{tramas} * M_{muestras}}{FC} + Header \quad \text{bits}$$

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Estructura de un paquete de VoIP



En cada paquete de voz hay N_{trama} cada una de las cuales dura un tiempo igual a T_{trama}, así que cada paquete debe enviarse en un tiempo $T_{paquete}$:

$$T_{paquete} = N_{trama} T_{trama} \quad \text{s}$$

La tasa de envío de paquetes $R_{paquete}$ es el inverso de $T_{paquete}$ dado por

$$R_{paquete} = \frac{1}{T_{paquete}} = \frac{1}{N_{trana} T_{trama}} \quad \text{paquetes/s}$$

Estos cálculos están hechos bajo la premisa de que la frecuencia de muestreo es de 8 KHz y que cada muestra se codifica con 8 bits. El tamaño del header se expresa en bytes y FC es el Factor de Compresión.

La diferencia entre la capacidad total del paquete y la usada por el payload, es la que se asigna al header. En todos los caso el header es superior al payload, pero sin el header lo paquetes no llegarían a destino.

Tasa de Bits

Entonces la tasa de bits $Bit\_Rate_{paquete}$ necesaria para enviar un paquete es igual a la tasa de envío de paquetes multiplicada por la cantidad de bits de cada paquete:

$$Bit\_Rate_{paquete} = \frac{\frac{8 * N_{tramas} * M_{muestras}}{FC} + Header}{N_{trama} T_{trama}}$$

Si se toman 8000 muestras en un segundo, entonces en el tiempo correspondiente a una trama, se toma $M_{muestra}$:

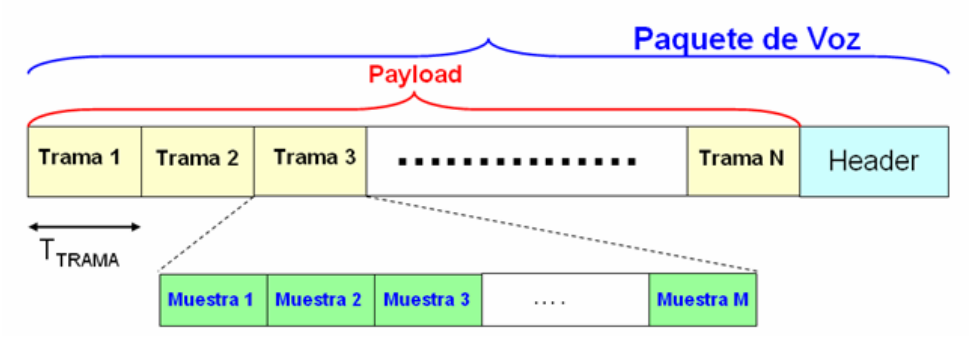
$$M_{muestras} = 8000 * T_{trama}$$

Y finalmente la tasa

$$Bit\_Rate_{paquete} = \frac{64000 * N_{tramas} * T_{trama} / FC + Header}{N_{trama} T_{trama}}$$

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Estructura detallada de un paquete de VoIP



$$Bit\_Rate = 64000/FC + \frac{Header}{N_{trama} T_{trama}}$$

Esta ecuación nos indica que la tasa de bits necesaria para llevar los paquetes de VoIP es la suma de la tasa de bits para llevar las muestras más la tasa de bits para llevar el encabezado. Dado que dicha ecuación es general y válida para cualquier CODEC, también observamos que la diferencia entre los CODECs es el factor de compresión FC.
El header depende de cada tecnología de capa 2, ya que los encabezados de RTP, UDP e IP son constantes. Si no conocemos directamente el factor de compresión FC, podemos reemplazar 64000/FC directamente por la tasa del CODEC.

En la expresión arriba mostrada hay que considerar el hecho de tanto LTE como WiMAX pueden comprimir el encabezado a fin de aumentar la eficiencia en el uso de los recursos.

CODEC AMR

- AMR (Adaptive Multirate) es un codec optimizado para aplicaciones de voz y es el más importante para aplicaciones de redes inalámbricas
- Adoptado por el 3GPP desde 1998
- Es el CODEC típico usado en redes inalámbricas 3G
- Incluye
  - un detector de actividad de voz (VAD)
  - un generador de ruido de confort
  - un sistema para combatir los errores y las perdidas de paquetes en la transmisión
- El CODEC Multirate tiene 8 modos de operación con tasas que van desde 4.75 Kbps hasta 12.2 Kbps
  - La cantidad de bits de del payload del paquete de voz depende de la tasa de bits
- El Header contiene 19 bits
- RESUMEN DE LAS CARACTERISTICAS DEL CODEC AMR
  - Frecuencia de muestreo 8 KHz
  - Tramas de voz de 20 ms cuando hay actividad, corresponden a 160 muestras. Se analizan las 160 muestras para extraer los parámetros del modelo CELP
  - Estos parámetros se codifican y se transmiten
  - En el decodificador se decodifican los parámetros y se genera una voz sintetizada
  - Hangover. 7 tramas de voz: 140 ms.

Existe una gran variedad de esquemas de codificación para voz, por ejemplo G.711, G.722, G.722.2, G.723.1, G.728, G.729 y AM; cada uno de estos modos requiere uncancho de banda diferente para transmitir las tramas de voz. Típicamente para VoIP se necesitan entre 5 Kbps y 64 Kbps incluyendo los headers en ambos sentidos.

AMR está basado en el esquema de codificación conocido como “Algebraic Code Excited Linear Prediction Coder” ACELP y está definido en el estándar ETSI TS 126 090 V9.0.0 y se conoce popularmente como el CODEC GSM. EL CODEC puede ajustar se tasa de bits automáticamente en cualquier momento dependiendo de las condiciones del canal.

CODEC ADAPTIVE MULTI-RATE (AMR)

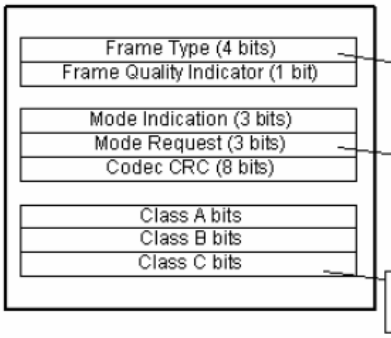
AMR es un CODEC de audio estandarizado por el 3GPP y por supuesto es usado en todas sus tecnologías, así como en otras tecnologías; debido a sus prestaciones y eficiencia es el CODEC típico que se usa en redes móviles. AMR está basado en Algebraic Code Excited Linear Prediction Coder (ACELP) y está definido en el estándar ETSI TS 126 090 V9.0.0. AMR ofrece una amplio rango de bit rates que va desde 4.75 Kbps hasta 12.2 kbps; incorpora un sistema de detección de actividad de voz (VAD) y un generador de ruido de confort para reducir la cantidad de bits enviados durante los periodos de silencio. AMR es el cuarto CODEC definido para GSM y tiene la ventaja de ir adaptando el nivel de protección de errores a las condiciones del canal y del tráfico.

Codec mode	Source codec bit-rate
AMR_12.20	12,20 kbit/s (GSM EFR)
AMR_10.20	10,20 kbit/s
AMR_7.95	7,95 kbit/s
AMR_7.40	7,40 kbit/s (IS-641)
AMR_6.70	6,70 kbit/s (PDC-EFR)
AMR_5.90	5,90 kbit/s
AMR_5.15	5,15 kbit/s
AMR_4.75	4,75 kbit/s
AMR_SID	1,80 kbit/s

La Tabla muestra los modos del CODEC AMR, incluyendo aquel donde se inserta el ruido de confort denominado AMR_SID Silence Descriptor (Comfort Noise Frame).

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Estructura Genérica de AMR



Cantidad de bits en una trama de voz en función de la tasa del CODEC

Frame Type	AMR codec mode	Total number of bits	Class A	Class B	Class C
0	4,75	95	42	53	0
1	5,15	103	49	54	0
2	5,90	118	55	63	0
3	6,70	134	58	76	0
4	7,40	148	61	87	0
5	7,95	159	75	84	0
6	10,2	204	65	99	40
7	12,2	244	81	103	60

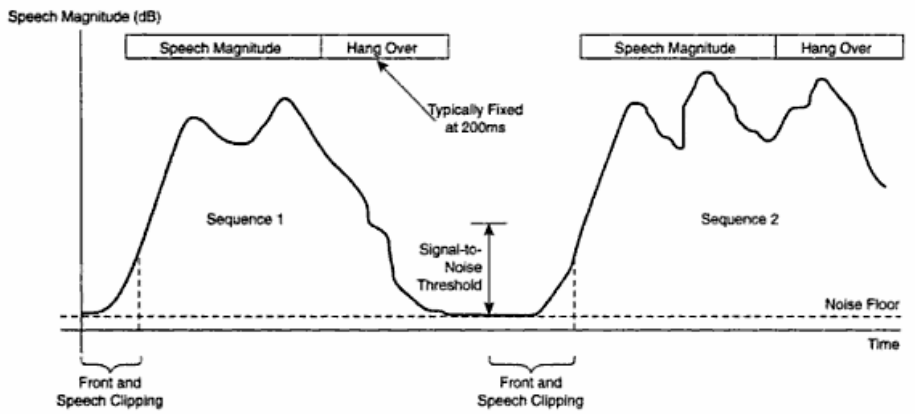
La trama genérica del CODEC tiene una duración de 20 ms, y está formada por la concatenación del AMR Header, AMR Auxiliary Information y del AMR Core Frame, donde podemos ver que el paquete de voz o de ruido de confort tiene un encabezado de 19 bits.

La trama genérica del CODEC AMR está formada por la concatenación del AMR Header, AMR Auxiliary Information y del AMR Core Frame.

En el AMR Core se envían los bits de voz o de ruido de confort, dependiendo de si el CODEC detecta o no actividad de voz. Los bits de core se clasifican en tres clases: A, B y C. Cada clase obedece a la sensibilidad de dichos bits a los errores. Los bits A son los más sensibles a los errores y errores en estos bits típicamente resultan en una trama dañada a la cual se deben aplicar los métodos de corrección de errores antes de decodificarlas, esta clase de bits es protegida por el CRC. La clase B y C contienen bits menos sensibles cuyos errores pueden degradar la calidad de la voz pero aun se puede decodificar la trama. Cuando se tramite el ruido de confort, los bits respectivos se envían todos en la Clase A; ni la Clase B ni la C se usan en dicho caso.

Voice Activity Detector (VAD)
Detección de Actividad de Voz

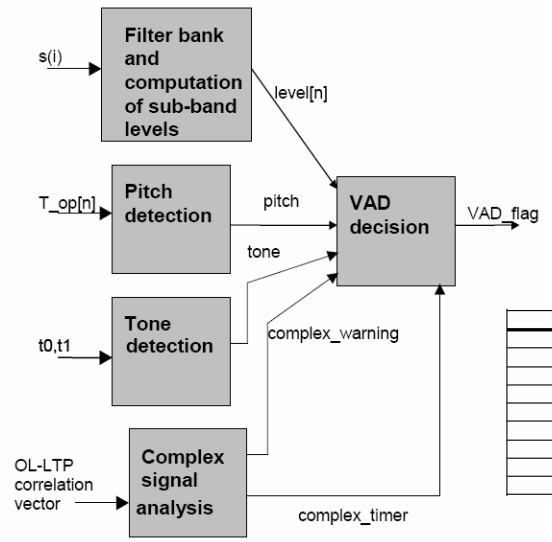
- VAD es un algoritmo que decide si el tipo de señal a la entrada del CODEC es voz, audio o si es ruido. La decisión está basada en un análisis riguroso de la señal de entrada.
- La salida del VAD es una bandera que indica si en la trama hay o no señales de audio. Cuando el VAD observa una caída de la amplitud de la señal de voz, espera un tiempo denominado Hangover, antes de detener el proceso de empaquetamiento de la voz.
- La decisión toma en cuenta tanto la trama actual así como tramas anteriores.
- Al detectar que no hay actividad de voz, se procede a enviar ruido de confort



El ruido de confort se transmite en una trama de menor tamaño y a una menor tasa en comparación con las tramas de voz,con lo que se usa de manera más eficiente el ancho de banda.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

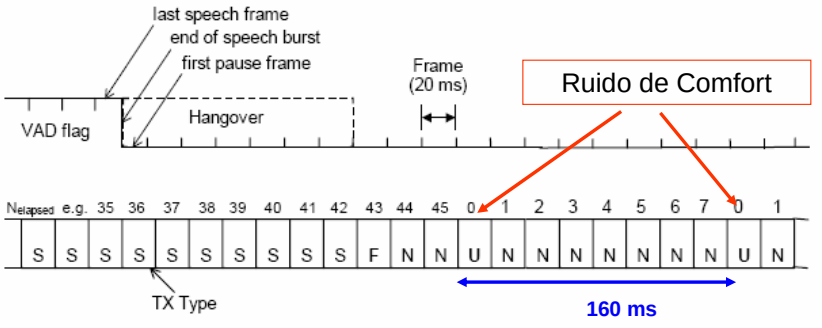
Posible Diagrama en Bloques de un VAD



Band number	Frequencies
1	0 - 250 Hz
2	250 - 500 Hz
3	500 - 750 Hz
4	750 - 1000 Hz
5	1000 - 1500 Hz
6	1500 - 2000 Hz
7	2000 - 2500 Hz
8	2500 - 3000 Hz
9	3000 - 4000 Hz

El sistema usa parámetros del codificador de voz para determinar si existe una señal de voz o audio en la entrada del CODEC. Las muestras de la trama de entrada se dividen en sub-bandas por medio de un banco de filtros, y luego se calcula el nivel de señal en cada una de dichas sub-bandas.

Factor de Actividad de Voz y Ruido de Confort



("S" = SPEECH; "F" = SID_FIRST; "U" = "SID_UPDATE; "N" = NO DATA,
Nelapsed: No. Of elapsed frames since last SID_UPDATE)

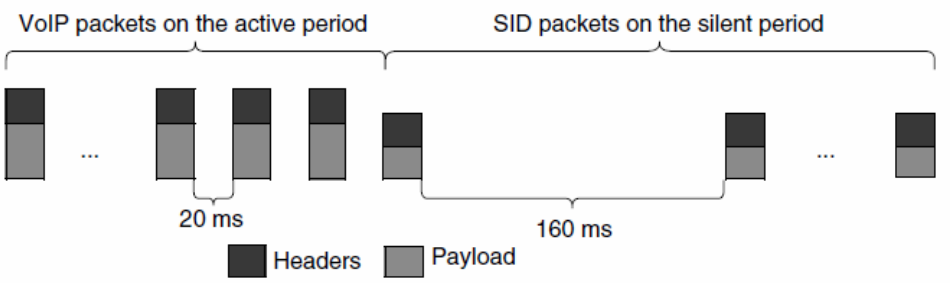
Aquí se muestra el proceso cuando VAD detecta que no hay señal de voz a la entrada del CODEC. Después de la última trama de voz, se inicia el periodo de Hangover cuya duración es de 7 tramas, después se envía una trama SID="F" que significa que es la primera trama SID (Silence Detector) después del Hangover. Después se dejan pasar dos tramas sin enviar información y luego se envía una trama de actualización U (Ruido de Comfort). La incorporación del sistema VAD crea un tráfico ON/OFF, es decir tramas de voz con tramas de ruido de confort intercaladas.

Las tramas de voz son las tramas "S" y cada una lleva una cantidad de bits que depende de la tasa de codificación, de acuerdo a la tabla mostrada anteriormente. Para el caso particular de Full Rate (12.2 Kbps) cada trama S lleva 244 bits correspondientes a los parámetros de la voz de acuerdo a CELP.

Después que se ha enviado el primer SID_First se deben calcular y enviar las tramas SID_UPDATE cada 8 trama, es decir espaciadas 160 ms**.** El primer SID_UPDATE debe enviarse tres tramas después de haber enviado la trama SID_FIRST. Cada trama SID_UPDATE transporta 35 bits.

Al detectarse que no hay actividad de voz, después de la última trama de voz se inicia el periodo de Hangover cuya duración es de 7 tramas, después se envía la primera trama SID (Silence Detector) denominada SID_First ="F", que significa que es la primera trama SID después del Hangover. Después se dejan pasar dos tramas sin enviar información, son las tramas N="No Data", y luego se envía una trama de actualización U="SID_Update" que es la lleva el ruido de confort. Las tramas U se envían cada 8 tramas, es decir cada 160 ms y cada una transporta en total 39 bits]. Es necesario aclarar que cuando nos referimos a la tasa del CODEC sólo se consideran los bits de voz, es decir los 12.2 Kbps del AMR FR solo incluyen los 244 bits de la trama de voz no incluye los 19 bits de encabezado; éstos últimos deben considerarse para calcular los recursos que consume un paquete de VoIP.

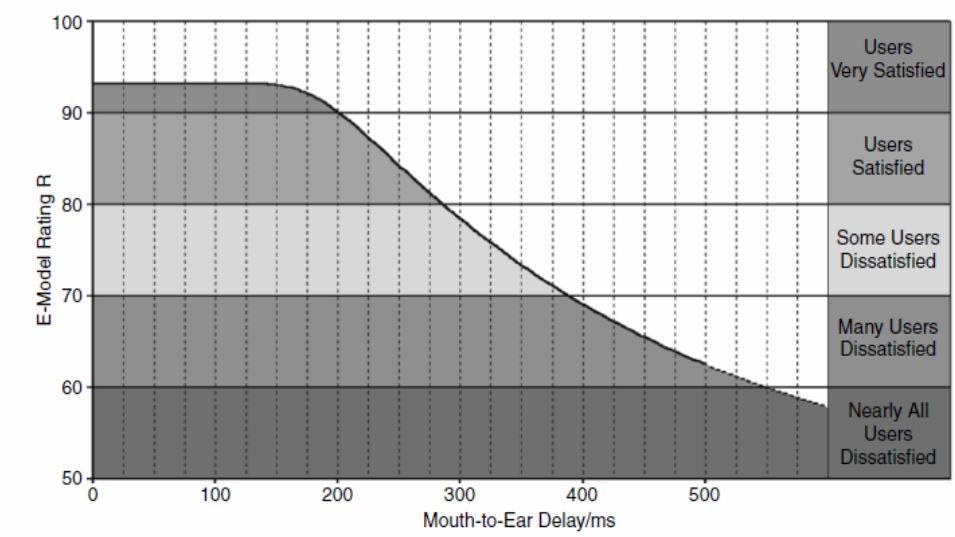
Tráfico de VoIP



Los paquetes de voz se generan cada 20 ms. El sistema de detección de actividad de voz, envía paquetes de silencio cada 160 ms cuando no hay actividad. Si se asume que el delay del core network es de 100 ms, quedan 150 ms para la capa MAC haga todos los procedimientos necesarios.

SID: Silence Descriptor

VoIP pertenece a la categoría de tráfico conversacional y por lo tanto el retardo debe mantenerse en límites aceptables. El retardo promedio entre boca-oido esta en el orden de 250 ms. A continuación se muestra una gráfica indicando la opinión de los usuarios en función del retardo.

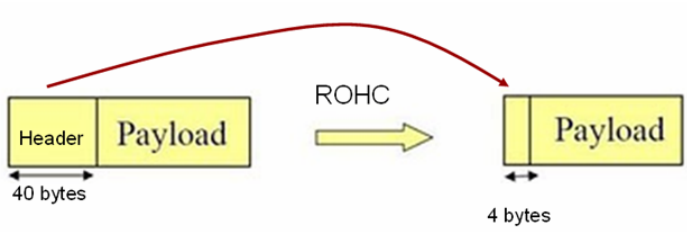


E-model rating en función del retardo boca-oido (ms). *Fuente*: ITU-T Recommendation G. 114. “One way transmission time.” (05/2003)

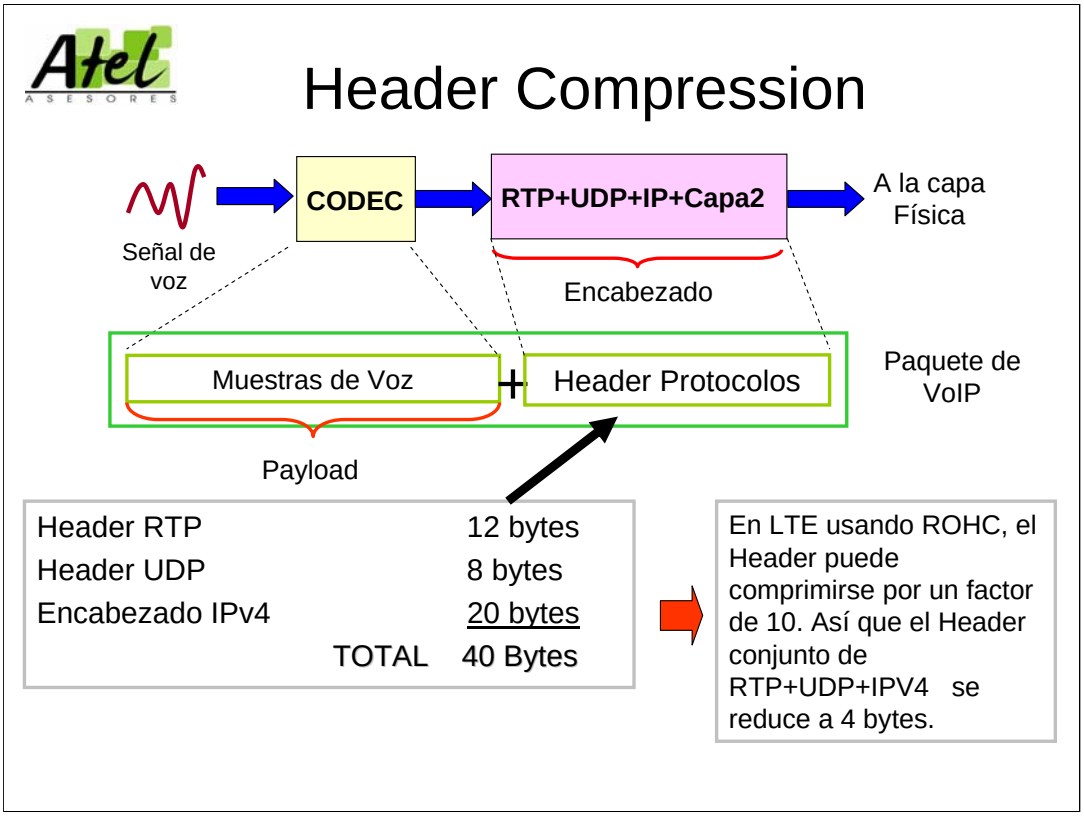
Header Compression

ROHC (Robust Header Compression RFC 3095)

ROHC es un estándar de la IEFT. En LTE es obligatorio sólo para los UE que soportan VoIP, los cuales deben soportar al menos un método de compresión para RTP, UDP e IP. En el RFC 4995 la IETF describe otros protocolos de compresión. A continuación se presentan los protocolos de compresión de Headers soportados por LTE.



Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739



El proceso de compresión de los encabezados permite ahorrar recursos del sistema, básicamente ancho de banda y al enviar paquetes de menor tamaño también se reduce la latencia.

En LTE el encabezado de los protocolos se puede comprimir hasta por un factor de 10, de esta manera se ahorra en ancho de banda y se mejora la latencia de la red. Es bueno recordar que según el estándar, los UE que soporten VoIP deben también soportar ROHC. La compresión de los encabezados se hace en capa 2, por lo que el header de esta última no se comprime. Sólo se comprime el encabezado de las capas superiores; el paquete de voz que sale del CODEC incluye las muestras de voz y un encabezado propio de cada CODEC, dicho encabezado no se comprime. En el caso de VoIP, se comprimen los headers de RTP, UDP y IP.

Capa 2 LTE

La capa 2 depende de cada estándar y cada uno agrega un encabezado particular. En esta parte sólo analizaremos el caso de LTE, pero igualmente la metodología es válida para cualquier otra tecnología.

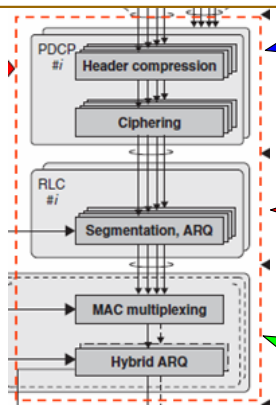
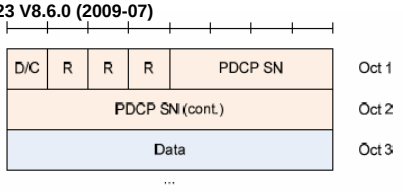


Figure 6.2.3.1: PDCP Data PDU format for DRBs using a 12 bit SN

ETSI TS 136 323 V8.6.0 (2009-07)



RLC ETSI TS 136 322 V8.6.0 (2009-07)

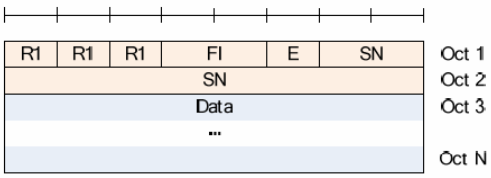
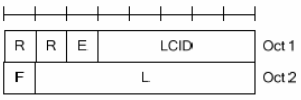


Figure 6.2.1.3-2: UMD PDU with 10 bit SN (No LI)



ETSI TS 136 321 V8.6.0 (2009-07), pp. 32

PDCP Header: pueden ser 2 o 1 byte. Si el PDCD SN (Sequence Number) ocupa 12 bits el header son 2 byte; si el SN ocupa 7 bits entonces el header es 1 byte. Se toma el peor caso: 2 bytes.

RLC Header: pueden ser 2 o 1 byte. Si el SN (Sequence Number) ocupa 10 bits el header son 2 byte; si el SN ocupa 5 bits entonces el header es 1 byte. Se toma el peor caso: 2bytes.

MAC Header: por cada SDU MAC se genera un subheader que puede ser de 2 o 3 bytes. Para nuestros efectos se consideraran 2 bytes.

En total anivel de capa 2 son 6 bytes, o pudieran ser 4 bytes si consideramos 1 byte cada PDCP y otro para RLC.

Encabezados de las capa 2 y capa 1 de LTE

Capa	Encabezado Corto bytes	Encabezado Largo bytes
PDCP	1	2
RLC	1	2
MAC	1	3
Física	2	2
Total	5 bytes	9 bytes

La tabla muestra un resumen de los encabezados de las capa 2 y 1 de LTE. Nótese que tenemos dos tipos de encabezados en función de la longitud de los números de secuencias en cada subcapa. Cada paquete, sea de voz o de ruido de confort, debe llevar estos 5 o 9bytes de las dos últimas capas. Para nuestro análisis escogeremos el peor caso que sería cuando el encabezado total es de 9 bytes.

CODECs AMR Full Rate 12.2 Kbps Periodos de Actividad de Voz

- La trama AMR Full Rate, cuando hay actividad de voz, está compuesta por un header de 19 bits y un payload de 244 bits, la cual es enviada cada 20 ms
- Agregando un bit de relleno se obtienen 264 bits que equivale a 33 bytes
- HEADER Capa 2 y Capa 1 en LTE

	MAXIMO	MINIMO
– PDCP (Seguridad)	2 bytes	1 byte
– RLC	2bytes	1 byte
– MAC	3 bytes	1 bytes
– Capa 1 (CRC)	2 bytes	2 bytes
Sub-total	9 bytes	5 bytes

- Capas Superiores 4 bytes (RTP+IUDP+IP)
- TOTAL: 33+4+9=46 bytes (tomando el caso de 9 bytes para capa 1 y 2)

$$Tasa\_bits\_Voz = \frac{46*8}{0.02} = 18.4\,kbps$$

Esta es la tasa de bits que se necesita para una conexión de VoIP en una dirección usando el CODEC AMR 12.2 Kbps, y equivale al peor caso; si las condiciones del canal mejoran el CODEC puede usar automáticamente una modo con una menor tasa, en cuyo caso la tasa de transmisión de VoIP disminuiría. De igual manera es esta tasa no se está considerando los periodos de actividad de la voz que evidentemente reducirán dicha tasa. No se incluye overhead debido a HARQ ya que en VoIP la capa RLC trabaja en Unacknowledged Mode (UM).

Este método para estimar el encabezado aplica para cualquier protocolo de capa 2 y méteo de compresión del header, sólo se necesita conocer las particularidades de cada protocolo.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

CODECs AMR Full Rate 12.2 Kbps Periodos de Silencio

- La trama U del AMR Full Rate, en los periodos de silencio, está formada por 39 bits de payload y 19 de header, la cual es enviada cada 160 ms
- HEADER Capa 2 yCapa 1 en LTE
  - 9 bytes
- Capas Superiores 4 bytes
- TOTAL: 8+4+9=21 bytes

Frame Type Index	FQI	AMR TX_TYPE or RX_TYPE	Total number of bits	Class A			Class B	Class C
				SID Type Indicator (STI)	Mode Indication mi(i)	Comfort Noise Parameter s(i)		
8	1	SID_UPDATE	39	1 (= "1")	3	35	0	0
8	1	SID_FIRST	39	1 (= "0")	3	35 (= "0")	0	0
8	0	SID_BAD	39	1	3	35	0	0

$$Tasa\_bits\_Silencio = \frac{20.25*8}{0.16} = 1.0125\,kbps$$

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

Datos Experimentales VoIP

- CODEC AMR
- Base de experimentos
  - 10 conversaciones de 15 de duración cada una
  - Periodos de actividad de voz (ON): 1026 ms
  - Periodos de silencio (OFF): 1171 ms
  - Tasa de actividad de voz: $T_{ON}/T_{Total}=0,467$

Nota: Estos tiempos fueron obtenidos a partir del instante en que cambia la bandera del VAD

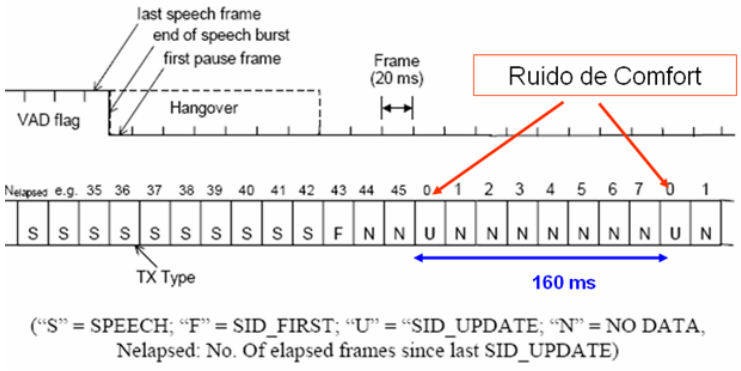
Estos resultados experimentales reportados en la literatura se usarán como base para establecer los parámetros del modelo de tráfico para VoIP usando un CODEC AMR Full Rate.

Conociendo el tiempo promedio que dura la actividad de voz y sabiendo que los paquetes se generan cada 20 ms, podemos calcular ;la cantidad de paquetes de voz enviada durante el tiempo en que se habla; dado que después que se detecta la falta de actividad vocal se deja pasar el tiempo de hangover equivalente a 7 tramas de voz, entonces al tiempo T_{ON} debemos sumar 140 ms.

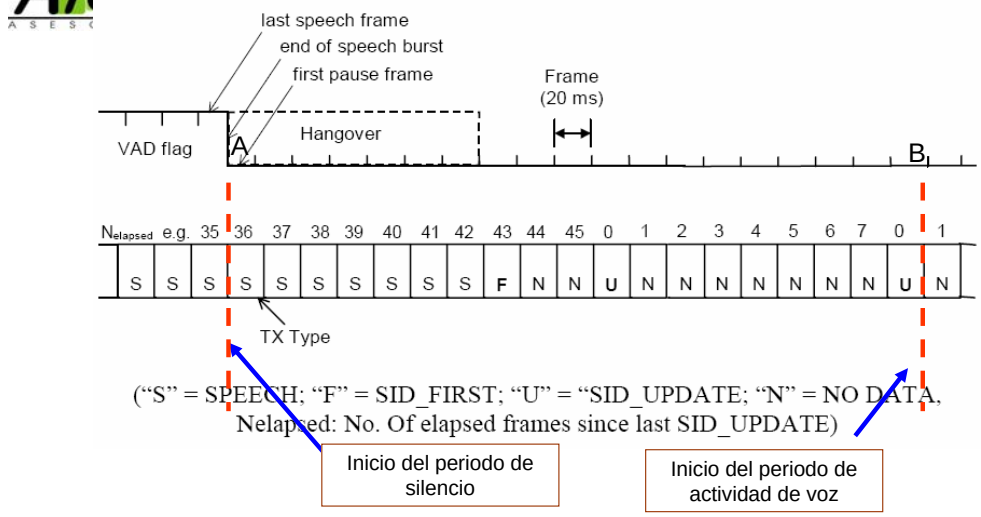
En el caso de los periodos de silencio, debemos considerar que los mismos se envían cada 160 ms.

Cantidad total de bits del payload para cada modo del CODEC AMR

El CODEC AMR adapta la tasa de bits en función de la calidad del canal, y también en función de la actividad de voz. La Fig. muestra un diagrama temporal de un CODEC AMR ajustando la tasa de bits.



("S" = SPEECH; "F" = SID_FIRST; "U" = "SID_UPDATE; "N" = NO DATA, Nelapsed: No. Of elapsed frames since last SID_UPDATE)



El VAD levanta la bandera cuando no hay actividad de voz, pero después aún se envían 7 tramas de hangover; así que para calcular la cantidad equivalente de paquetes de voz enviados se debe sumar el tiempo de hangover de 140 ms. De igual manera al tiempo de silencio debemos quitarle el tiempo de hangover.

$$P_{ON} = \frac{T_{ON} + 140\text{ms}}{20\text{ms}} = 58\ \text{paquetes} \qquad P_{OFF} = \frac{T_{OFF} - 140\text{ms}}{160\text{ms}} = 6.5\ \text{paquetes}$$

Como los tiempo T_{ON} y T_{OFF} se calculan con el cambio de la bandera del VAD, debemos considerar el tiempo de hangover ya que la cantidad de datos enviada en los periodos de actividad y de silencio son distintas en cada caso. Al tiempo T_{ON} se le debe sumar el tiempo de Hangover y se le debe restar la misma cantidad al tiempo T_{OFF}.

Tasa de Bits Promedio

$$Tasa\_bits\_Confort = \frac{Total\_bits\_confort}{160\,\text{ms}} = 1.013\,\text{Kbps}$$

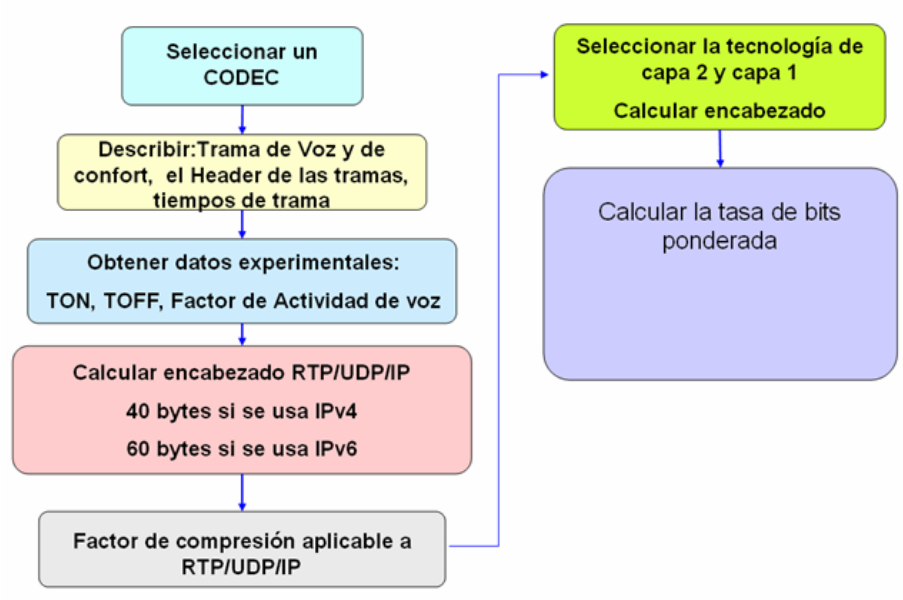
$$Tasa\_bits\_Voz = \frac{Total\_bits\_voz}{20\,\text{ms}} = 18.4\,\text{Kbps}$$

$$R_{b-VoIP} = \frac{Tasa\_bits\_Voz * T_{VOZ} + Tasa\_bits\_Confort * T_{CONFORT}}{T_{ON} + T_{OFF}} = 10.241\,\text{Kbps}$$

Al final obtenemos 10.24 Kbps como tasa promedio de bits para VoIP usando el CODEC AMR 12.2 Kbps. Esta sería la tasa de bits promedio que consume un usuario de VoIP usando un CODEC AMR Full Rate cuando consideramos que los SN (Sequence Number) de capa2 tienen la mayor longitud de acuerdo a los estándares del 3GPP, es decir analizamos el peor escenario. Esta metodología puede aplicarse a cualquier CODEC y cualquier otra tecnología de capa 2 y capa 1, por supuesto los parámetros aquí calculados serán diferentes en cada caso.

Aquí se muestran los valores promedios, pero cuando se están transmitiendo tramas de voz se deben reservar 18.4 Kbps, mientras que para las tramas de ruido de confort se deben reservar apenas 1.013 Kbps. A título de planificar y cuantificar los recursos necesarios para atender una cierta cantidad de usuarios de VoIP es recomendable usar la tasa pico en los periodos de actividad de voz.

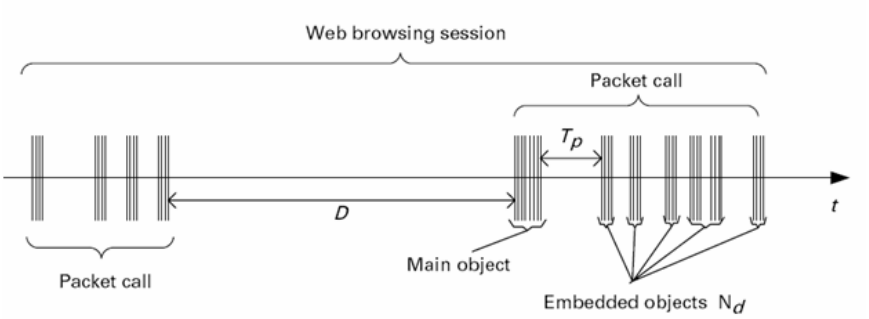
Procedimientos para calcular la tasa de bits en VoIP



Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739

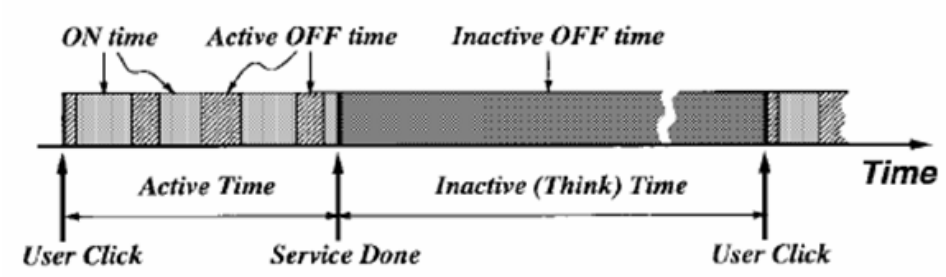
Modelo de Tráfico para Navegación WEB

Se observa que el tráfico http es de tipo ON/OFF, representado por tiempos de download mientras se bajan las páginas, y periodos de lectura por parte del usuario. Estos periodos intercalados de actividad e inactividad representan la interacción entre la red y el usuario. En la figura "Packet Calls" representa los paquetes de las páginas solicitadas, mientras que el "Reading Times" se refiere a los periodos durante le cual el usuario lee, analiza las páginas y toma decisiones con relación a las acciones a seguir, por ejemplo hacer click en un link, imprimir la página, guardar la página , etc.



El tráfico WEB se caracteriza por el tamaño promedio del objeto principal SM, el tamaño de los objetos embebidos SE, la cantidad de objetos embebidos ND, el tiempo de lectura D y el tiempo y el tiempo de análisis Tp, que se refiere que tarda el usuario para seleccionar un link dentro del la página principal.

Atel Asesores C.A atel.asesores@cantv.net www.atelasesores.com.ve Telf. 0212-9417771 0416-6306739



Traza de paquetes en una sesión WEB, donde se muestran los periodos de actividad e inactividad. Incluso los periodos de actividad se caracterizan por instantes de recepción de los paquetes e instantes en los que no se recibe nada, este comportamiento es típico de la conmutación de paquetes y depende sólo de la red, no es atribuible al usuario.

Basado en observaciones realizadas, el 76% de los paquetes usan un MTU (Maximum Transmission Unit) de 1500bytes, mientras que el 24% restante usa un MTU de 576 Bytes, incluyendo los 40 bytes de TCP/IPv4.

Parámetros del modelo de tráfico HTTP

Parameter	Statistical characterization
Main object size S_M	Truncated lognormal distribution, mean = 10710 bytes, standard deviation = 25032 bytes, minimum = 100 bytes, maximum = 2 Mbytes (before truncation) PDF: $f_x = \frac{1}{\sqrt{2\pi}\sigma x} e^{\frac{-(\ln x - \mu)^2}{2\sigma^2}} \quad x > 0 \sigma = 1.37, \mu = 8.37$
Embedded object size S_E	Truncated lognormal distribution, mean = 7758 bytes, standard deviation = 126168 bytes, minimum = 50 bytes, maximum = 2 Mbytes (before truncation) PDF: $f_x = \frac{1}{\sqrt{2\pi}\sigma x} e^{\frac{-(\ln x - \mu)^2}{2\sigma^2}} \quad x > 0 \quad \sigma = 2.36, \mu = 6.17$
Number of embedded objects per page N_D	Truncated Pareto distribution, mean = 5.64, maximum = 53 (before truncation) PDF: $f_x = \frac{\alpha k^\alpha}{\alpha + 1}, k \le x < m \quad f_x = \left(\frac{k}{m}\right)^\alpha, x = m$ $\alpha = 1.1, \quad k = 2, \quad m = 55$ Note: subtract k from the generated random value to obtain N_D
Reading time D	Exponential distribution with a mean = 30 seconds PDF: $f_x = \lambda e^{-\lambda x} \quad x \ge 0 \quad \lambda = 0.033$
Parsing time T_p	Exponential distribution with mean = 0.13 seconds PDF: $f_x = \lambda e^{-\lambda x} \quad x \ge 0 \quad \lambda = 7.69$

Los modelos de Tráfico para http no suministran información sobre la tasa de bits que necesitaría la aplicación, y es normal ya que http no tiene calidad de servicio debido a que no es en tiempo real. Los operadores hacen un estimado sobre la tasa de bits con la cual esperan que los usuarios estén satisfechos. Por ejemplo, para usuarios residenciales actualmente 512 Kbps es una tasa aceptable, mientras que para las empresas se puede usar 2 Mbps para las de mayor demanda y 1 Mbps para las otras.