

# **VOZ, VIDEO Y TELEFONIA SOBRE IP**

**Dr. Ing. José Joskowicz**

[iosei@fing.edu.uy](mailto:iosei@fing.edu.uy)

**Instituto de Ingeniería Eléctrica, Facultad de Ingeniería**

**Universidad de la República**

**Montevideo, URUGUAY**

**Agosto 2013**

**Versión 12**

## Tabla de Contenido

Tabla de Contenido .....	2
1 Introducción.....	4
2 Voz sobre redes de datos .....	6
2.1 Codificación de voz .....	6
2.2 Transmisión de voz sobre redes de datos .....	7
2.3 RTP – Real-Time Transport Protocol .....	8
2.3.1 Versión (V) .....	10
2.3.2 CSRC count (CC - Contributing Sources Count) .....	10
2.3.3 Tipo de información (PT - Payload Type).....	10
2.3.4 Número de secuencia (Sequence Number) .....	10
2.3.5 Marca de tiempo (Time Stamp).....	10
2.3.6 Identificador del origen (SSRC - Synchronization Source Identifier) .....	11
2.3.7 Identificador del tributario (CSRC - Contributing Sources Identifier) .....	11
2.4 RTCP – RTP Control Protocol.....	14
2.5 Ancho de banda en IP para voz .....	16
3 Video sobre redes de datos .....	19
3.1 Aplicaciones de video .....	19
3.2 Codificación de video .....	19
3.3 Transmisión de video sobre redes de datos .....	20
3.3.1 Paquetización del video .....	20
3.3.2 Ancho de banda en IP para video.....	24
4 Calidad de voz en redes IP .....	27
4.1 Factores que afectan la calidad de la voz sobre redes de paquetes.....	27
4.1.1 Factor de compresión y codificación.....	27
4.1.2 Pérdida de paquetes .....	27
4.1.3 Demora .....	28
4.1.4 Eco.....	29
4.1.5 Variaciones en la demora (Jitter) .....	30
4.1.6 Tamaño de los paquetes.....	31
4.2 Estimación de la calidad de voz en redes de paquetes: ITU-T G.107 (E-Model).....	31
5 Calidad de video en redes IP .....	42
5.1 Factores que afectan la calidad del video sobre redes de paquetes.....	42
5.1.1 Factor de compresión .....	42
5.1.2 Pérdida de paquetes .....	43
5.1.3 Demora / Jitter .....	45
5.2 Estimación de la calidad de video sobre redes de paquetes: ITU-T G.1070.....	45
6 Calidad de Servicio en redes de datos.....	48
6.1 QoS en Capa 2 .....	48
6.2 QoS en Capa 3 .....	51
6.3 QoS en Capa 4 y superiores.....	52
7 Voz y Video sobre redes inalámbricas .....	53
7.1 Cobertura .....	53

7.2	Movilidad.....	53
7.3	Calidad de Servicio .....	54
7.4	Capacidad.....	56
8	Protocolos de señalización en VoIP .....	57
8.1	H.323 .....	57
8.1.1	Arquitectura H.323.....	57
8.1.1.1	Terminales.....	58
8.1.1.2	Gateways (Pasarelas) .....	63
8.1.1.3	Gatekeeper.....	64
8.1.1.4	MCU – Multipoint Control Unit (Unidad de control multipunto) .....	65
8.1.2	Ejemplos de señalización H.323 .....	66
8.2	SIP .....	72
8.2.1	Mensajería SIP .....	73
8.2.2	Arquitectura SIP .....	77
8.2.2.1	Terminales SIP.....	77
8.2.2.2	Servidores SIP .....	78
8.2.2.3	Gateway SIP .....	80
8.2.3	Ejemplos de señalización SIP.....	80
8.3	Otros protocolos de señalización en VoIP .....	82
9	Gestión de proyectos de VoIP.....	83
9.1	Iniciando un proyecto de VoIP (Procesos de Iniciación) .....	84
9.2	Planificando un proyecto de VoIP (Procesos de Planificación).....	85
9.3	Ejecutando un proyecto de VoIP (Procesos de Ejecución) .....	86
9.4	Controlando un proyecto de VoIP (Procesos de Monitoreo y Control)....	86
9.5	Finalización del proyecto de VoIP (Procesos de Cierre) .....	87
	Referencias .....	89

# 1 Introducción

Las redes de voz [1] y las redes de datos [2] presentan tecnologías muy disímiles. Por un lado, la transmisión de voz, con una historia de más de 130 años, se basa en el establecimiento de vínculos permanentes entre dos puntos, diseñados para transmitir un tipo de señal específico: la voz humana, típica señal analógica, de ancho de banda acotado, que debe llegar a destino “inmediatamente” y ser lo más inteligible posible. Por otro lado, la transmisión de datos, con una historia relativamente reciente, se basa en la transmisión de información digital, utilizando técnicas de conmutación de paquetes, donde las pérdidas y los retardos no producen generalmente consecuencias importantes.

La integración de estas dos tecnologías no parece algo sencilla, y cabe preguntarse si existe alguna ventaja en realizar el intento. Las ventajas aparecen al analizar por los menos los siguientes tres aspectos:

El primero aspecto es económico: es posible ahorrar dinero al integrar las tecnologías. Si analizamos, por ejemplo, el tráfico de voz de larga distancia internacional, se puede ver como ha habido un crecimiento sostenido año a año, mientras que los precios han bajado también sostenidamente. Sin embargo, las ganancias de los operadores se han mantenido o han aumentado, como se muestra en la Figura 1.1 (tomada de [3]). Esto es, en parte, debido a la utilización de las tecnologías de VoIP, las que están reemplazando a las tecnologías TDM para el tráfico de voz internacional. Como se puede ver en la Figura 1.2 (también tomada de [3]), el incremento porcentual año a año del tráfico internacional de VoIP ha ido creciendo a ritmos mayores al TDM.

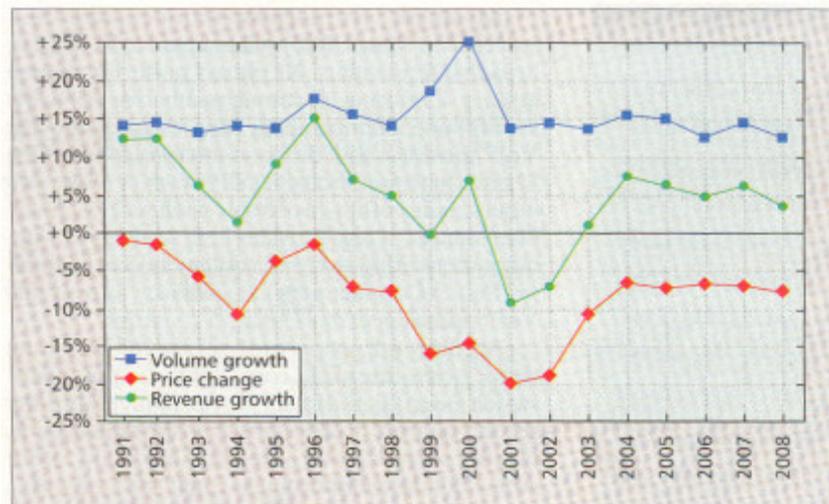


Figura 1.1

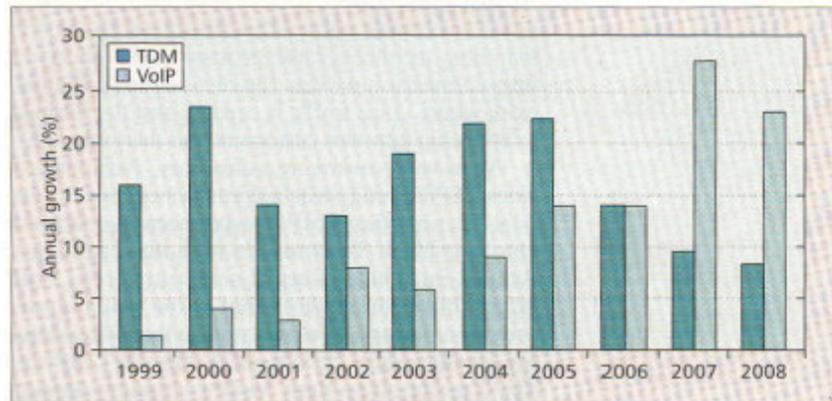


Figura 1.2

El segundo aspecto que representa una ventaja de la integración o unificación es la administración: es más sencillo administrar un único sistema que dos independientes. Históricamente, la administración de las redes telefónicas y las redes de datos eran realizadas por grupos diferentes, con perfiles diferentes, y reportando a diferentes gerencias dentro de las organizaciones. La unificación de las tecnologías está generando también una tendencia hacia la unificación en la gestión y administración, tanto en el área de los operadores telefónicos como en el ámbito corporativo. Dentro de los operadores telefónicos, las gerencias de transmisión y transporte se confunden cada vez más con las “viejas” áreas de “datos” y de “voz”. En las empresas corporativas la “Gerencia de IT” está absorbiendo, dentro de sus tareas, la administración de los sistemas telefónicos, soportados sobre las redes IP.

El tercer aspecto, y quizás a nivel del usuario presente las ventajas más relevantes, tiene que ver con la mejora en las aplicaciones. Las nuevas tecnologías permiten a los usuarios disponer de facilidades que hasta hace un tiempo no eran posibles. Algunas de estas ventajas funcionales se describen en mayor detalle en [4].

Todos los aspectos son importantes, y las empresas, tanto prestadoras de servicios, desarrolladoras, como consumidoras de tecnología están haciendo una fuerte apuesta a la integración y unificación.

En estas notas se analizan los problemas tecnológicos que aparecen al cursar tráfico de voz y video sobre redes de datos, y se presentan los protocolos más utilizados de señalización de video-telefonía sobre IP.

## 2 Voz sobre redes de datos

### 2.1 Codificación de voz

La voz es codificada digitalmente para su transmisión. Los dispositivos de codificación y decodificación se denominan CoDec (Codificadores / Decodificadores). Los procesos de codificación y decodificación, así como los estándares más utilizadas, pueden ser consultados en [5]

Los codecs pueden ser clasificados según diferentes características, entre las que se encuentran su tasa de bits (bit rates), la calidad resultante del audio codificado, su complejidad, el tipo de tecnología utilizada y el retardo que introducen, entre otros. Originalmente, los primeros codecs fueron diseñados para reproducir la voz en la banda de mayor energía, entre 300 Hz a 3.4 kHz. Actualmente este tipo de codecs son caracterizados como de “banda angosta” (narrowband). En contraste, los codecs que reproducen señales entre 50 Hz y 7 kHz se han llamado de “banda ancha” (wideband). Más recientemente, ITU-T ha estandarizado codecs llamados de banda superancha (superwideband), para el rango de 50 Hz a 14 kHz y de banda completa (fullband), para el rango de 50 Hz a 20 kHz.

A modo de referencia, la Tabla 2.1 resume los codecs más conocidos.

#### Codecs de banda angosta (narrowband):

Codec	Nombre	Bit rate (kb/s)	Retardo (ms)	Comentarios
G.711	PCM: Pulse Code Modulation	64, 56	0.125	Codec “base”, utiliza dos posibles leyes de compresión: $\mu$ -law y A-law [6]
G.723.1	Hybrid MPC-MLQ and ACELP	6.3, 5.3	37.5	Desarrollado originalmente para video conferencias en la PSTN, es actualmente utilizado en sistemas de VoIP [7]
G.728	LD-CELP: Low-Delay code excited linear prediction	40, 16, 12.8, 9.6	1.25	Creado para aplicaciones DCME (Digital Circuit Multiplex Encoding) [8]
G.729	CS-ACELP: Conjugate Structure Algebraic Codebook Excited Linear Prediction	11.8, 8, 6.4	15	Ampliamente utilizado en aplicaciones de VoIP, a 8 kb/s [9]
AMR	Adaptive Multi Rate	12..2 a 4.75	20	Utilizado en redes celulares GSM [10]

**Codecs de banda ancha (wideband):**

Codec	Nombre	Bit rate (kb/s)	Retardo (ms)	Comentarios
G.722	Sub-band ADPCM	48,56,64	3	Inicialmente diseñado para audio y videoconferencias, actualmente utilizado para servicios de telefonía de banda ancha en VoIP [11]
G.722.1	Transform Coder	24,32	40	Usado en audio y videoconferencias [12]
G.722.2	AMR-WB	6.6, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, 23.85	25.9375	Estandar en común con 3GPP (3GPP TS 26.171). Los bit rates más altos tienen gran inmunidad a los ruidos de fondo en ambientes adversos (por ejemplo celulares) [13]
G.711.1	Wideband G.711	64, 80, 96	11.875	Amplía el ancho de banda del codec G.711, optimizando su uso para VoIP [14]
G.729.1	Wideband G.729	8 a 32 kb/s	<49 ms	Amplía el ancho de banda del codec G.729, y es "compatible hacia atrás" con este codec. Optimizado su uso para VoIP con audio de alta calidad [15]
RtAudio	Real Time Audio	8.8, 18	40	Codec propietario de Microsoft, utilizado en aplicaciones de comunicaciones unificadas (OCS) [16]

**Codecs de banda super ancha (superwideband):**

Codec	Nombre	Bit rate (kb/s)	Retardo (ms)	Comentarios
SILK	SILK	8 a 24	25	Utilizado por Skype [17]

**Codecs de banda completa (fullband):**

Codec	Nombre	Bit rate (kb/s)	Retardo (ms)	Comentarios
G.719	Low-complexity, full-band	32 a 128	40	Es el primer codec "fullband" estandarizado por ITU [18]

Tabla 2.1

**2.2 Transmisión de voz sobre redes de datos**

Para poder transmitir las muestras codificadas de voz sobre redes de datos, es necesario armar "paquetes". Un canal de voz consiste en un flujo de bits, dependientes del codec utilizado. Si, por ejemplo, la voz está codificada con el

codec G.711 en ley A, un canal de voz consiste en un flujo de 64 kb/s. (una “muestra” de voz, codificada con 8 bits, cada 125  $\mu$ s). Para enviar este flujo sobre una red de datos, es necesario armar “paquetes”. Si bien se podría formar un paquete con cada muestra de voz, esto generaría un sobrecarga (“overhead”) demasiado importante (recordar que cada paquete requiere de cabezales). Por otro lado, si se espera a “juntar” demasiadas muestras de voz, para formar un paquete con mínima sobrecarga porcentual, se pueden introducir retardos no aceptables. Un paquete IP puede tener hasta 1500 bytes de información. Si con muestras del codec G.711 se quisiera completar los 1500 bytes del paquete IP, se introduciría un retardo de  $125\mu\text{s} \times 1500 = 187,5 \text{ ms}$ . Esta demora no es aceptable en aplicaciones de conversaciones de voz.

Por esta razón, se toman generalmente “ventanas” de 10 a 30 ms (ver Figura 2.1). Las muestras codificadas de voz de cada una de estas ventanas se “juntan” y con ellas se arman paquetes. El tamaño de estas ventanas es configurable para algunos algoritmos de codificación, y está estandarizado para otros.

## Flujo de datos codificando la voz

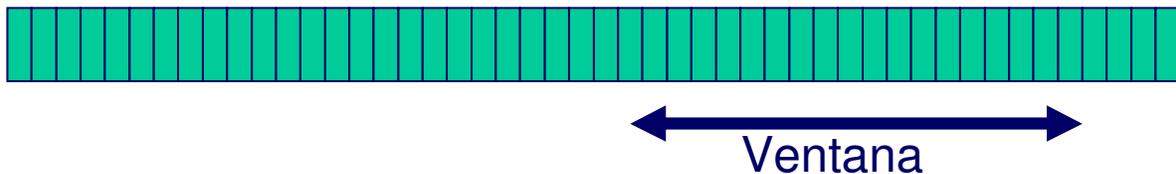


Figura 2.1

### 2.3 RTP – Real-Time Transport Protocol

El protocolo RTP, basado originalmente en el RFC 1889 y luego reemplazado por el RFC 3550 [19], establece los principios de un protocolo de transporte sobre redes que no garantizan calidad de servicio para datos “de tiempo real”, como por ejemplo voz y video.

El protocolo establece la manera de generar paquetes que incluyen, además de los propios datos de “tiempo real” a transmitir, números de secuencia, marcas de tiempo, y monitoreo de entrega. Las aplicaciones típicamente utilizan RTP sobre protocolos de red “no confiables”, como UDP. Los “bytes” obtenidos de cada conjunto de muestras de voz o video son encapsulados en paquetes RTP, y cada paquete RTP es a su vez encapsulado en segmentos UDP (Ver Figura 2.2).

RTP soporta transferencia de datos a destinos múltiples, usando facilidades de “multicast”, si esto es provisto por la red.

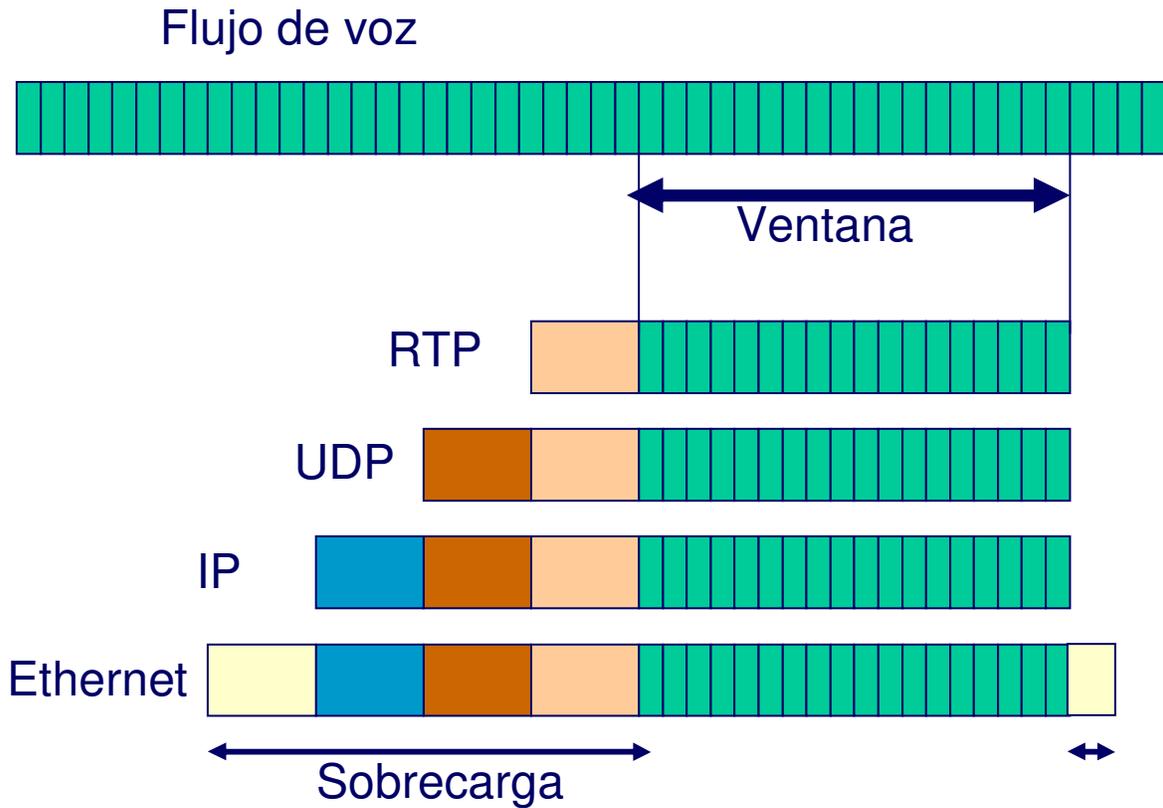


Figura 2.2

Cada paquete RTP consiste en un cabezal y los datos de voz. El cabezal contiene números de secuencia, marcas de tiempo, y monitoreo de entrega. El formato de éste cabezal es el mostrado en la Figura 2.3

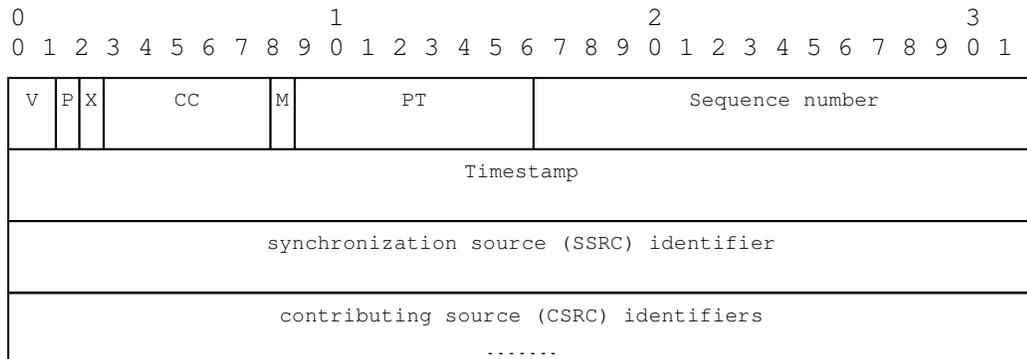


Figura 2.3

Los campos más relevantes son:

### 2.3.1 Versión (V)

La versión actual del protocolo es la 2.

### 2.3.2 CSRC count (CC - Contributing Sources Count)

El campo indica la cantidad de “fuentes” que contribuyen al audio incluido en el paquete. Pueden ser desde 0 a 15 (ver 2.3.7 )

### 2.3.3 Tipo de información (PT - Payload Type)

El campo “payload” identifica el tipo de información que viaja en el paquete. Es un campo de 7 bits, lo que permite diferenciar hasta 128 tipos de información. Este campo indica el tipo de codificación de audio o video, o el contenido de información “especial”. Los valores de este campo se definen en el RFC 3551 [20]. Algunos valores de ejemplo se muestran en la Tabla 2.2.

Payload Type	Formato	Medio	Clock Rate
0	PCM mu-law	Audio	8 kHz
3	GSM	Audio	8 kHz
4	G.723	Audio	8 kHz
8	PCM A-law	Audio	8 kHz
9	G.722	Audio	8 kHz
13	Confort Noise	Audio	
14	MPEG Audio	Audio	90 kHz
15	G.728	Audio	8 kHz
18	G.729	Audio	8 kHz
26	Motion JPEG	Video	90 kHz
31	H.261	Video	90 kHz
32	MPEG-1 o 2 Elementary Stream	Video	90 kHz
33	MPEG-1 o 2 Transport Stream	Video	90 kHz
34	H.263	Video	90 kHz
96 – 127	Dinámico		

Tabla 2.2

El valor 13 “Confort Noise” inidica que el paquete se trata de “ruido de confort”, utilizado junto con técnicas de detección de actividad de voz (VAD, Voice Activity Detection). Los valores 96 a 127 son dinámicos, su utilización puede depender de las aplicaciones. Por ejemplo, una utilización de los valores dinámicos es para codificar los dígitos DTMF, de acuerdo al RFC 2833 [21].

### 2.3.4 Número de secuencia (Sequence Number)

El campo correspondiente al número de secuencia es de 16 bits. Con cada paquete enviado, el emisor incrementa en uno el número de secuencia. Esto permite al receptor detectar paquetes perdidos, o fuera de orden.

### 2.3.5 Marca de tiempo (Time Stamp)

Este campo es de 32 bits. Indica el momento al que corresponde la primera muestra de la ventana de información que viaja en el paquete. Este campo es utilizado por el receptor, para reproducir las muestras con la misma cadencia con

las que fueron obtenidas. Es a su vez útil para medir el “jitter”. En audio, el campo “Time Stamp” se mide en unidades de 125  $\mu$ s (o sea, en unidades de muestreo). Si por ejemplo un paquete de 160 bytes de audio en Ley A contiene el campo TimeStamp con el valor 1, el siguiente paquete contendrá el campo TimeStamp en 160.

### **2.3.6 Identificador del origen (SSRC - Synchronization Source Identifier)**

El campo correspondiente al SSRC es de 32 bits. Típicamente cada flujo en una sesión RTP tiene un identificador diferente. El origen establece este número, asegurando que no se repita.

### **2.3.7 Identificador del tributario (CSRC - Contributing Sources Identifier)**

Pueden existir de 0 hasta 15 campos CSRC, de acuerdo al valor de CC. Esta lista identifica a cada uno de los interlocutores cuando el audio que se envía es producido en un mezclador o “mixer” (por ejemplo, cuando se envía el audio de varios participantes de una conferencia)

La Figura 2.4 muestra un ejemplo de un paquete RTP que contiene audio. En este caso, el audio fue codificado en G711 Ley  $\mu$ , y existe una única fuente que contribuye al audio (CSRC = 0).

La Figura 2.5 muestra un ejemplo de un paquete RTP que codifica el dígito DTMF “2” según el RFC 2833.

La Figura 2.6 muestra un ejemplo de un paquete RTP que codifica la señal de “Confort Noise”.

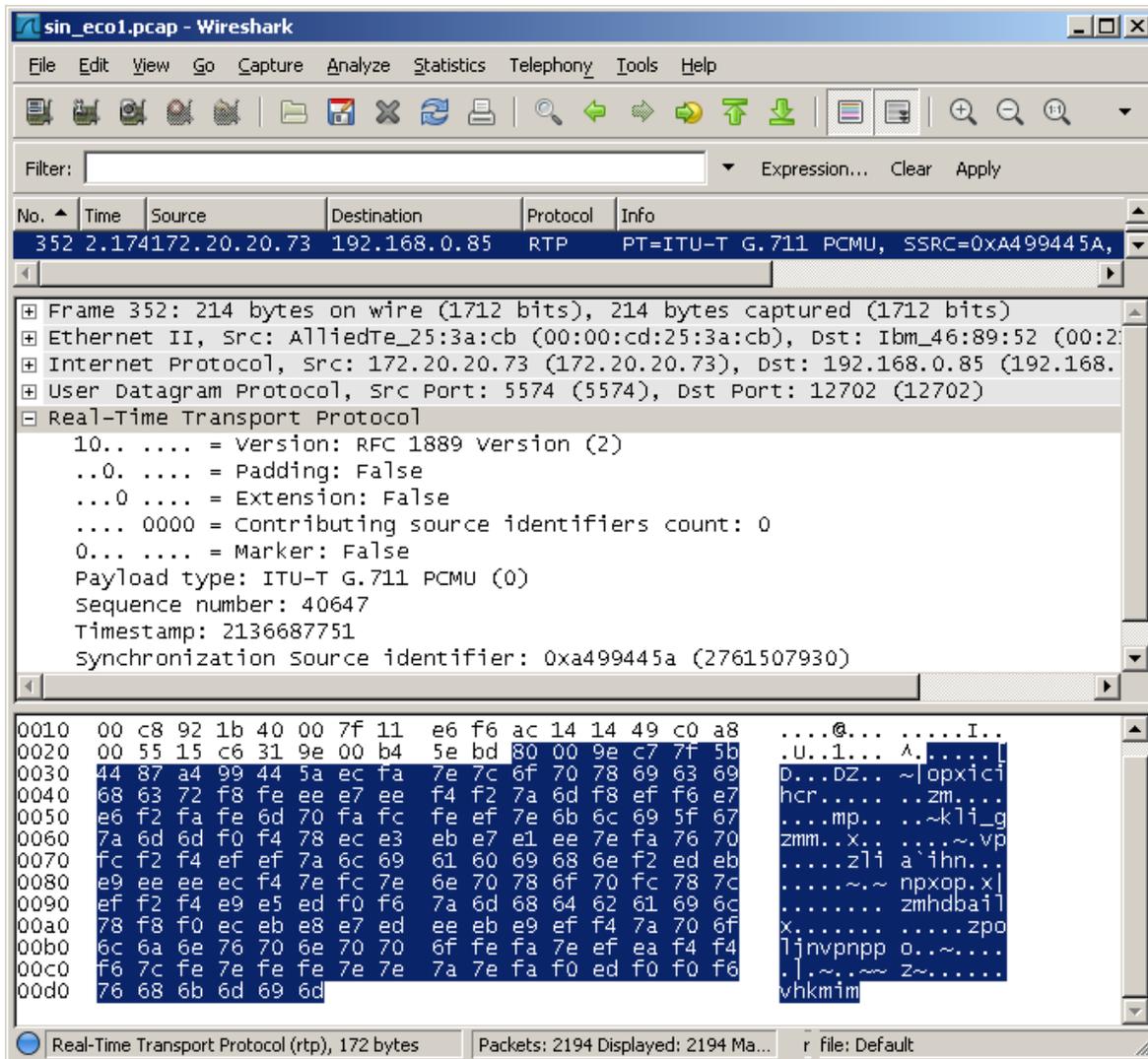


Figura 2.4

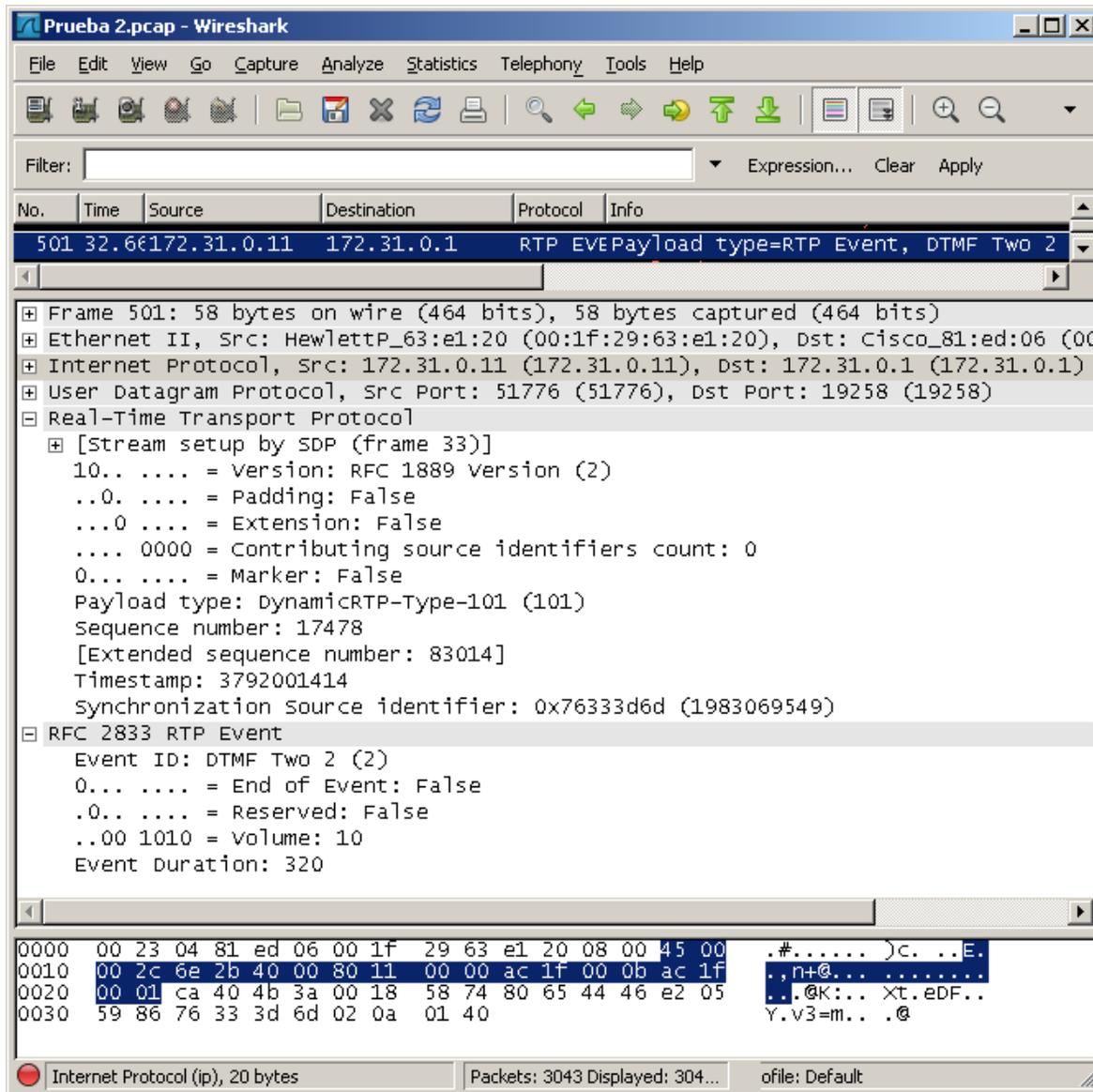


Figura 2.5

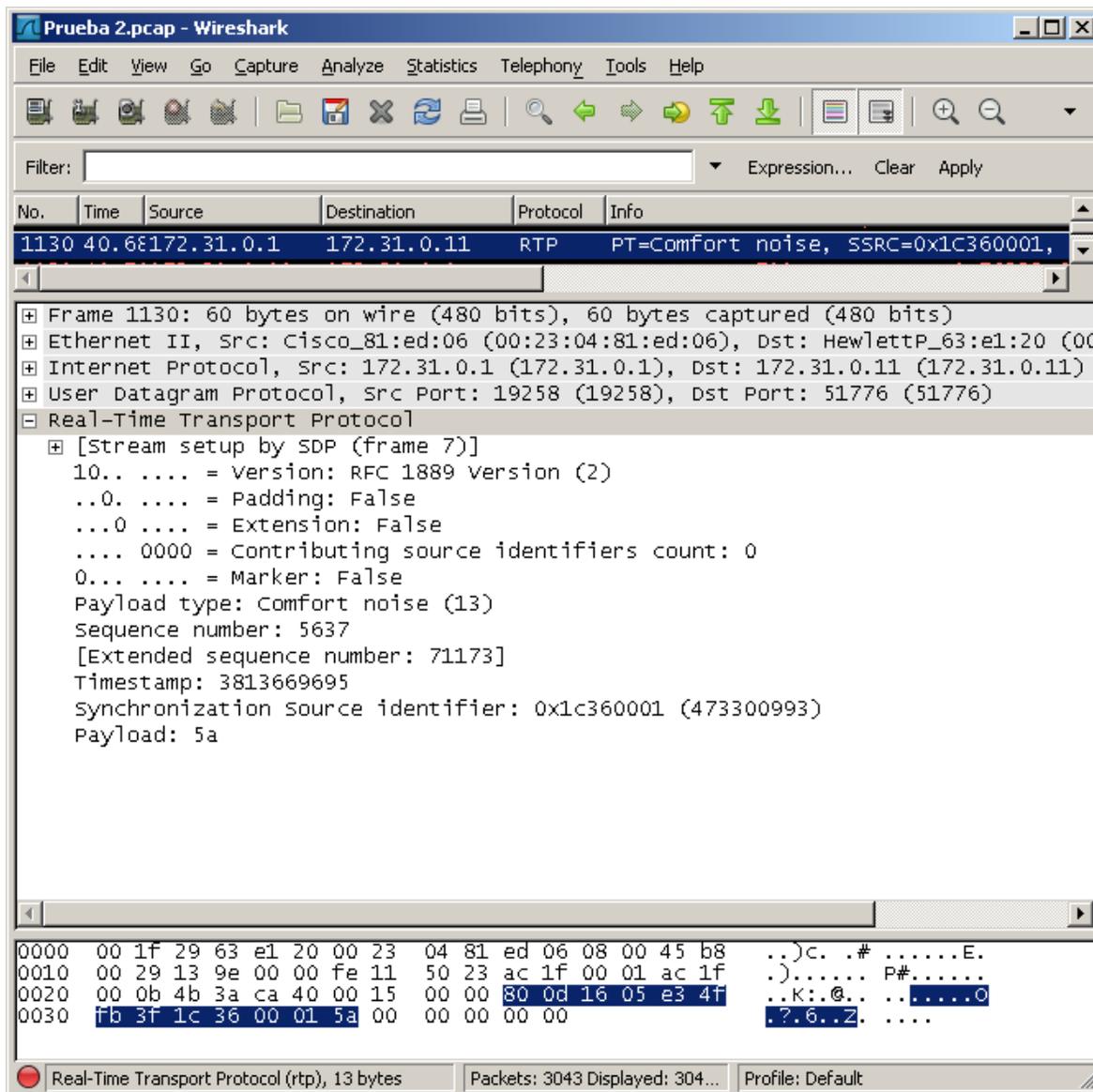


Figura 2.6

## 2.4 RTCP – RTP Control Protocol

El RFC 3550 establece, además del protocolo RTP, un protocolo de control, RTCP, encargado de enviar periódicamente paquetes de control entre los participantes de una sesión. El protocolo RTCP tiene las siguientes funciones principales:

- Proveer realimentación acerca de la calidad de los datos distribuidos (por ejemplo, de la calidad percibida de VoIP). Esta realimentación permite adaptar dinámicamente la codificación, o tomar acciones tendientes a

solucionar problemas cuando se detecta degradación en la calidad de la comunicación

- Transporte del CNAME (Canonical Name) de cada originador. Este identificador permite asociar varios flujos RTP con el mismo origen (por ejemplo, flujos de audio y video provenientes del mismo emisor)
- Adaptar dinámicamente la frecuencia de envío de paquetes de control RTCP de acuerdo al número de participantes en la sesión. Dado que los paquetes se deben intercambiar “todos contra todos”, es posible saber cuantos participantes hay, y de esta manera calcular la frecuencia de envíos de esto paquetes.

Los paquetes RTCP pueden ser de los siguientes tipos:

- SR (Sender Report): Envía estadísticas de los participantes “origen” (sender)
- RR (Receiver Report): Envía estadísticas de los participantes “destino” (receivers)
- SDES (Source Description): Envía ítems de descripción del origen
- BYE: Indica el fin de la participación en el intercambio de mensajes RTCP
- APP: Funciones específicas para las aplicaciones participantes

En la Figura 2.7 se muestra un ejemplo de un paquete RTCP. En este ejemplo se envían en el mismo paquete un SR (Sender Report) y un SDES (Source Description).

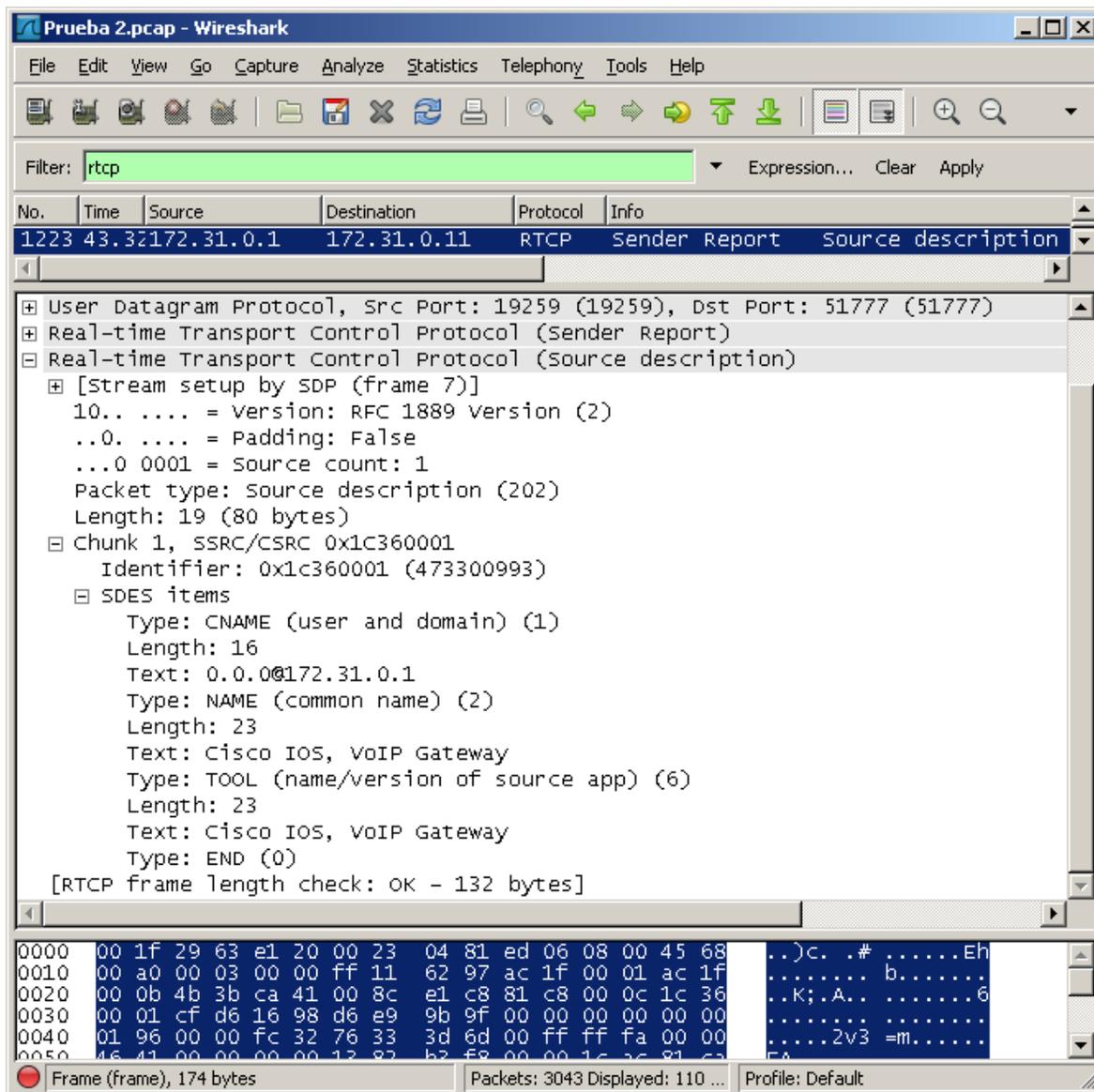


Figura 2.7

## 2.5 Ancho de banda en IP para voz

Dado que para el envío de voz sobre redes de datos es necesario armar “paquetes”, el ancho de banda requerido dependerá de la “sobrecarga” (“overhead”) que generen estos paquetes.

Como se ha visto, para el envío de voz sobre redes de paquetes se utiliza el estándar RTP. Éste protocolo a su vez se monta sobre UDP, el que a su vez se monta sobre IP, el que, en la LAN, viaja sobre Ethernet.

Esta suma de protocolos hace que el ancho de banda requerido para el tráfico de voz sobre Ethernet sea bastante mayor al ancho de banda del audio. A continuación se presenta un ejemplo, para el codec G.711.

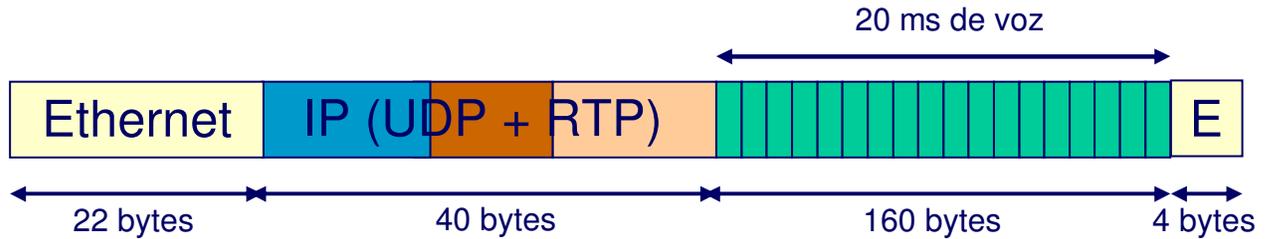


Figura 2.8

Para una ventana de 20 ms, y con codificación de audio G.711 Ley A, se obtienen 160 bytes de voz por trama (Ver Figura 2.8)

$$\text{Bytes de voz/trama} = 64 \text{ kb/s} * 20 \text{ ms} / 8 = 160 \text{ bytes}$$

El paquete IP (incluyendo los protocolos RTP y UDP) agrega 40 bytes adicionales

$$\text{Bytes de paquete IP} = 160 + 40 = 200 \text{ bytes}$$

La trama Ethernet agrega otros 26 bytes:

$$\text{Bytes de Trama Ethernet} = 200 + 26 = 226 \text{ bytes}$$

En este ejemplo, cada 20ms se generan 226 bytes que se deben enviar por la LAN. Esto equivale a un ancho de banda de 90,4 kb/s (compárese con los 64 kb/s del flujo de audio)

$$\text{Ancho de banda LAN} = 226 * 8 / 20 \text{ ms} = 90.4 \text{ kb/s}$$

Es de hacer notar que este cálculo fue hecho para el envío de audio en una dirección. Pueden utilizarse técnicas de “supresión de silencio”, en las que no se envían paquetes cuando no hay audio. En este caso, el ancho de banda en cada dirección es poco más de la mitad del cálculo anterior.

Por lo visto anteriormente, el ancho de banda de la voz paquetizada en la LAN depende del tamaño de la “ventana” (típicamente 10, 20 o 30 ms) y el codec utilizado.

La Tabla 2.3 muestra los anchos de banda unidireccionales necesarios utilizando redes IP sobre Ethernet para algunos codecs de banda angosta.

Tipo de Codec	Duración de Trama (ms)	Bytes de voz/Trama	Bytes de paquete IP	Bytes de trama Ethernet	Ancho de Banda en LAN (kbps)
<b>G.711 (64 kb/s)</b>	10	80	120	146	116,8
	20	160	200	226	90,4
	30	240	280	306	81,6
<b>G.729 (8 kb/s)</b>	10	10	50	76	60,8
	20	20	60	86	34,4
	30	30	70	96	25,6
<b>G.723.1 (6.3 kb/s)</b>	30	24	64	90	23,9
<b>G.723.1 (5.3 kb/s)</b>	30	20	60	86	22,9

Tabla 2.3

Como se puede ver en la tabla, el ancho de banda puede variar notablemente, dependiendo del codec y la ventana seleccionada.

Análisis similares se pueden realizar para otros codecs, tanto de banda angosta, como de banda ancha, super ancha o completa.

## 3 Video sobre redes de datos

### 3.1 Aplicaciones de video

El video es utilizado en diversos tipos de aplicaciones, las que a su vez, tienen diversos requerimientos. La TV es, quizás, la aplicación de video más conocida. Sin embargo, existen en forma cada vez más difundida un nuevo conjunto de aplicaciones de video, entre las que se encuentran la video telefonía, los servicios de video conferencia, la distribución de video a demanda a través de Internet y la IP-TV, por mencionar los más relevantes. Cada una de estas aplicaciones tiene sus características propias en lo que respecta a requerimientos de calidad, velocidades, etc.

En el área corporativa, se nota una creciente importancia de la video-telefonía y las video-conferencias.

La video-telefonía es una aplicación típicamente punto a punto, con imágenes del tipo “cabeza y hombros”, y generalmente poco movimiento. Por otra parte, es una aplicación altamente interactiva, donde los retardos punta a punta juegan un rol fundamental en la calidad conversacional percibida.

Las aplicaciones de video conferencias son típicamente punto a multi-punto. Al igual que la video telefonía, generalmente tienen poco movimiento. Además de la difusión del audio y el video es deseable en estas aplicaciones poder compartir imágenes y documentos. La interactividad también es típicamente un requisito, aunque podrían admitirse retardos punta a punta un poco mayores que en la video telefonía, ya que los participantes generalmente están dispuestos a “solicitar la palabra” en este tipo de comunicaciones.

Las aplicaciones de difusión (IP-TV o video a demanda por Internet, por ejemplo) son unidireccionales, y los retardos desde la emisión a la recepción pueden ser más grandes. Sin embargo, esto tiene impacto en eventos de tiempo real, como por ejemplo, campeonatos deportivos.

### 3.2 Codificación de video

Los procesos de codificación y decodificación de video, así como los estándares más utilizados, pueden ser consultados en [5]

A modo de referencia, la Tabla 3.1 resume algunos de los codecs más conocidos.

Característica	MPEG-1	MPEG-2	MPEG-4	H.264/MPEG-4 Part 10/AVC
Tamaño del macro-bloque	16x16	16x16 (frame mode) 16x8 (field mode)	16x16	16x16
Tamaño del bloque	8x8	8x8	16x16 8x8, 16x8	8x8, 16x8, 8x16, 16x16, 4x8, 8x4, 4x4
Transformada	DCT	DCT	DCT/DWT	4x4 Integer transform
Tamaño de la muestra para aplicar la transformada	8x8	8x8	8x8	4x4
Codificación	VLC	VLC	VLC	VLC, CAVLC, CABAC
Estimación y compensación de movimiento	Si	Si	Si	Si, con hasta 16 MV
Perfiles	No	5 perfiles, varios niveles en cada perfil	8 perfiles, varios niveles en cada perfil	3 perfiles, varios niveles en cada perfil
Tipo de cuadros	I,P,B,D	I,P,B	I,P,B	I,P,B,SI,SP
Ancho de banda	Hasta 1.5 Mbps	2 a 15 Mbps	64 kbps a 2 Mbps	64 kbps a 150 Mbps
Complejidad del codificador	Baja	Media	Media	Alta
Compatibilidad con estándares previos	Si	Si	Si	No

Tabla 3.1

### 3.3 Transmisión de video sobre redes de datos

#### 3.3.1 Paquetización del video

Las secuencias (Elementary Streams) son paquetizadas en unidades llamadas PES (Packetized Elementary Streams), consistentes en un cabezal y hasta 8 kbytes de datos de secuencia. Estos PES a su vez, son paquetizados en pequeños paquetes, de 184 bytes, los que, junto a un cabezal de 4 bytes (totalizando 188 bytes) conforman el “MPEG Transport Stream” (MTS) y pueden ser transmitidos por diversos medios.

En redes IP, el transporte del video se realiza mediante los protocolos RTP y RTCP, ya descritos en 2.3. El RFC 2250 [22] establece los procedimientos para transportar video MPEG-1 y MPEG-2 sobre RTP. Varios paquetes MTS de 188 bytes pueden ser transportados en un único paquete RTP, para mejorar la eficiencia. Esto se puede observar en la siguiente figura, donde se muestra un paquete RTP con 7 paquetes MTS (ISO 13818), cada uno de ellos de 188 bytes. A su vez, cada paquete MTS contiene un cabezal de 4 bytes y 184 bytes de “contenido”, con el video codificado en MPEG-2, como se ve en la segunda figura.

Test6\_720p\_rtp\_MPEG2.pcap - Wireshark

File Edit View Go Capture Analyze Statistics Telephony Tools Help

Filter: Expression... Clear Apply

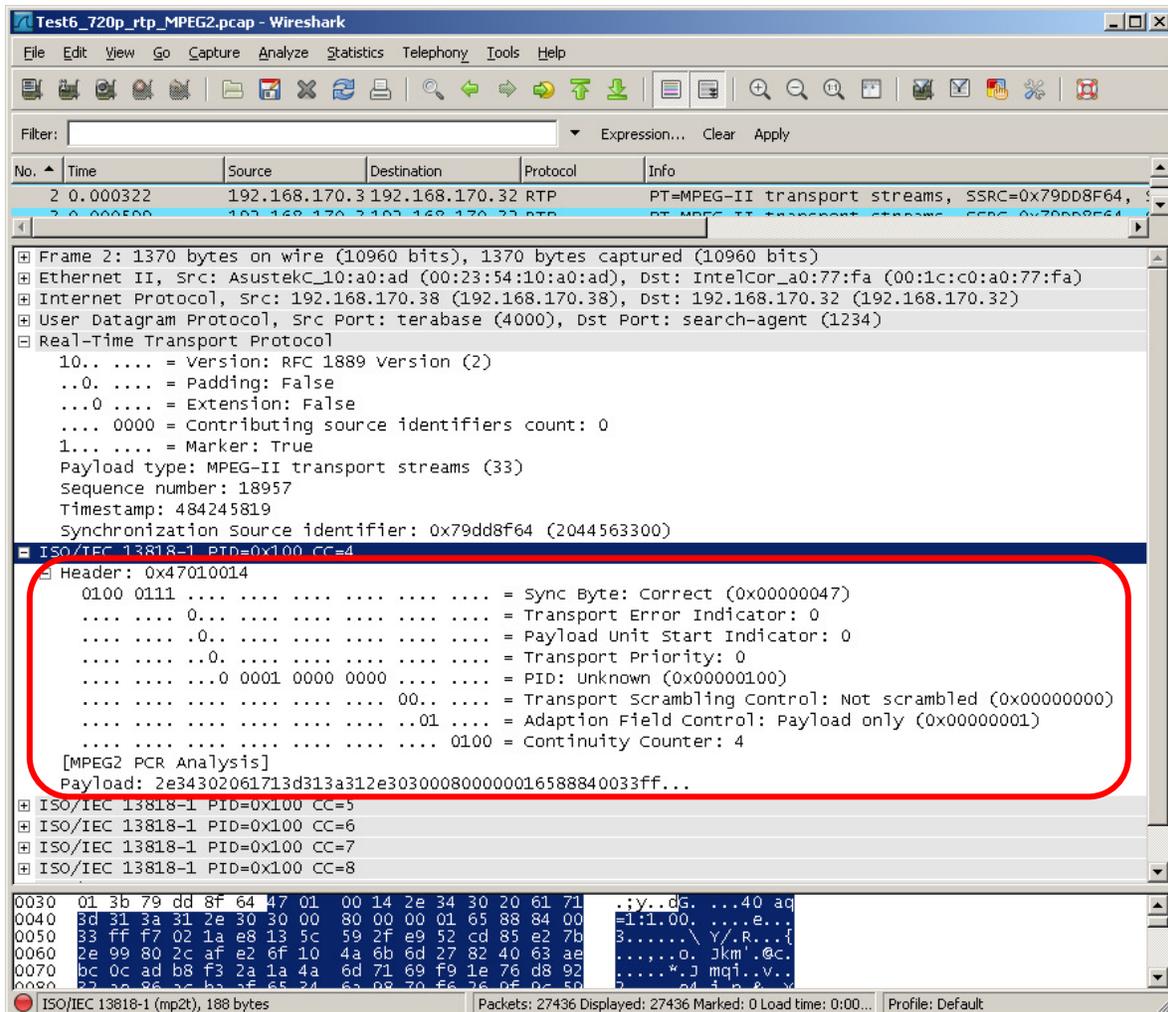
No.	Time	Source	Destination	Protocol	Info
2	0.000322	192.168.170.3	192.168.170.32	RTP	PT=MPEG-II transport streams, SSRC=
3	0.000599	192.168.170.3	192.168.170.32	RTP	PT=MPEG-II transport streams, SSRC=
4	0.000662	192.168.170.3	192.168.170.32	RTP	PT=MPEG-II transport streams, SSRC=
5	0.000827	192.168.170.3	192.168.170.32	RTP	PT=MPEG-II transport streams, SSRC=

Frame 2: 1370 bytes on wire (10960 bits), 1370 bytes captured (10960 bits)

- Ethernet II, Src: AsustekC\_10:a0:ad (00:23:54:10:a0:ad), Dst: IntelCor\_a0:77:fa (00:1c:c0:a0:77:fa)
- Internet Protocol, Src: 192.168.170.38 (192.168.170.38), Dst: 192.168.170.32 (192.168.170.32)
- User Datagram Protocol, Src Port: terabase (4000), Dst Port: search-agent (1234)
- Real-Time Transport Protocol
  - 10... .. = Version: RFC 1889 version (2)
  - ..0... .. = Padding: False
  - ...0... .. = Extension: False
  - .... 0000 = Contributing source identifiers count: 0
  - 1... .. = Marker: True
  - Payload type: MPEG-II transport streams (33)
  - Sequence number: 18957
  - Timestamp: 484245819
  - Synchronization source identifier: 0x79dd8f64 (2044563300)
  - ISO/IEC 13818-1 PID=0x100 CC=4
  - ISO/IEC 13818-1 PID=0x100 CC=5
  - ISO/IEC 13818-1 PID=0x100 CC=6
  - ISO/IEC 13818-1 PID=0x100 CC=7
  - ISO/IEC 13818-1 PID=0x100 CC=8
  - ISO/IEC 13818-1 PID=0x100 CC=9
  - ISO/IEC 13818-1 PID=0x100 CC=10

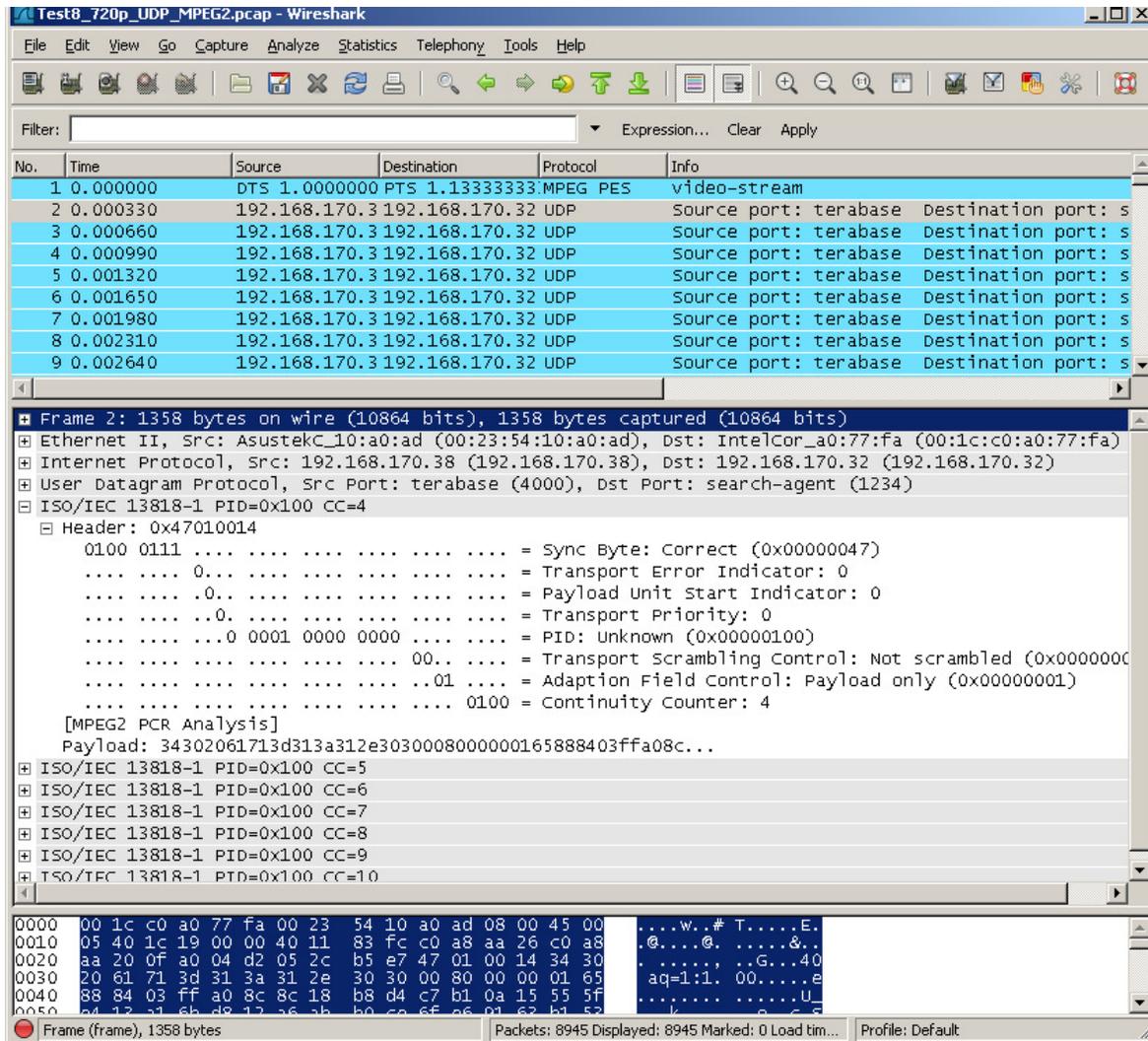
0000 00 1c c0 a0 77 fa 00 23 54 10 a0 ad 08 00 45 00 ...w..# |....E.  
 0010 05 4c 04 b6 00 00 40 11 9b 53 c0 a8 aa 26 c0 a8 .L....@. .S...&..  
 0020 aa 20 0f a0 04 d2 05 38 a8 bf 80 a1 4a 0d 1c dd .....8 ....J...  
 0030 01 3b 79 dd 8f 64 47 01 00 14 2e 34 30 20 61 71 .;y..dG. ...40 aq  
 0040 3d 31 3a 31 2e 30 30 00 80 00 00 01 65 88 84 00 =1:1.00. ....e...  
 0050 22 ff f7 02 13 08 12 5c 50 2f 00 52 cd 85 02 7b ? \ / n . f

File: "D:\jose\vigo\tesis\videos\Opticom\Test6\_720p\_rtp\_MPEG2... Packets: 27436 Displayed: 27436 Mark... Pr ile Default

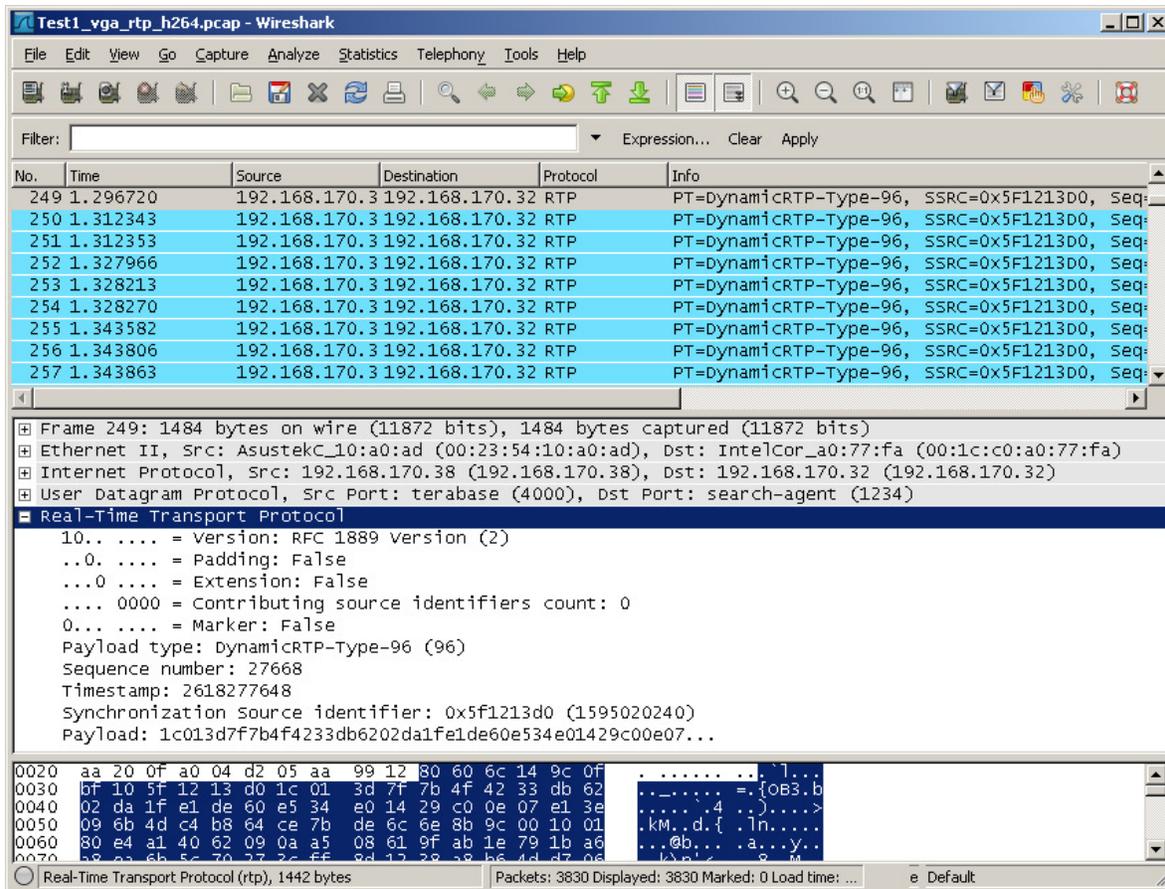


Algunos sistemas no utilizan el protocolo RTP, sino que incluyen los paquetes de MTS directamente en el paquete UDP, como se muestra en la siguiente figura. Si bien esta forma de transmisión tiene menos sobrecarga (“overhead”), ya que no se envían los bytes del protocolo RTP, el envío del video paquetizado dentro de RTP tiene varias ventajas [23] :

- El envío de video sobre dentro del protocolo RTP está estandarizado en el RFC 2250 y por el grupo DVB-IP [24], lo que asegura mejor interoperabilidad.
- El protocolo RTP utiliza el protocolo RTCP para como realimentación de la calidad. No hay mecanismos similares estandarizados sin se envía los MTS directamente sobre UDP
- RTP contiene nativamente información de reloj, que puede ser utilizado por el destino para decodificar apropiadamente los flujos multimedia.
- Muchos equipos de redes pueden darle prioridad al protocolo RTP frente a otros protocolos, gestionando de esta manera la QoS (Calidad de Servicio). Esto no es así si los MTS se encapsulan directamente debajo de UDP.



Los RFC 3016 [25] y RFC 3640 [26] establecen los procedimientos para transportar flujos de audio y video MPEG-4. El RFC 3984 [27] establece los procedimientos para transportar flujos de video codificados en H.264. En la siguiente figura se muestra un paquete de video codificado en H.264 dentro de RTP. En este caso se utiliza el “paylod tipe” dinámico, con el número 96. El paquete tiene 1430 bytes de “payload” con el video codificado en H.264.



### 3.3.2 Ancho de banda en IP para video

Como se ha visto, la codificación digital de video utiliza algoritmos de compresión, los que generan codificación de largo variable y flujos de ancho de banda también variables. Para una aplicación determinada, el ancho de banda requerido en una red IP dependerá del tipo de codificación utilizada (MPEG-1, 2, 4, H264, etc.), del tamaño de la pantalla (SD, CIF, QCIF, etc), del tipo de cuantización seleccionado y del movimiento y textura de la imagen. Al ancho de banda propio de la señal de video se le debe sumar la sobrecarga de los paquetes IP, UDP y RTP y para la LAN, de las tramas Ethernet. A diferencia de la codificación de audio de tasa de bits constantes, donde los anchos de banda pueden calcularse en forma exacta en base únicamente al codec utilizado, la codificación de video es estadística, y depende de la imagen transmitida, por lo que los cálculos son también aproximados y estadísticos. En video, generalmente, se puede establecer la tasa de bits o ancho de banda deseado, y el codec varía dinámicamente sus parámetros de codificación para alcanzar el ancho de banda establecido, a expensas de modificar la calidad.

A modo de ejemplo, en la Figura 3.1 se muestra como varía la calidad percibida en función del ancho de banda para diferentes secuencias de video, codificadas en MPEG-2, y en resolución CIF (352 x 288) a 25 cuadros por segundo. En esta gráfica, la calidad percibida se mide como “MOS”, donde valores de 5 se corresponden a “excelente calidad” y valores de 1 se corresponden con muy “mala calidad”. Cada línea se corresponde con una secuencia de video diferente, tomadas de la página del VQEG [28]. En este caso, para formatos CIF, se puede ver que para anchos de banda superiores a 1.75 Mb/s la calidad es “excelente” para cualquier secuencia de video. Sin embargo, para anchos de banda inferiores a 1 Mb/s la calidad percibida depende fuertemente del contenido del video (movimiento, textura de la imagen, etc.)

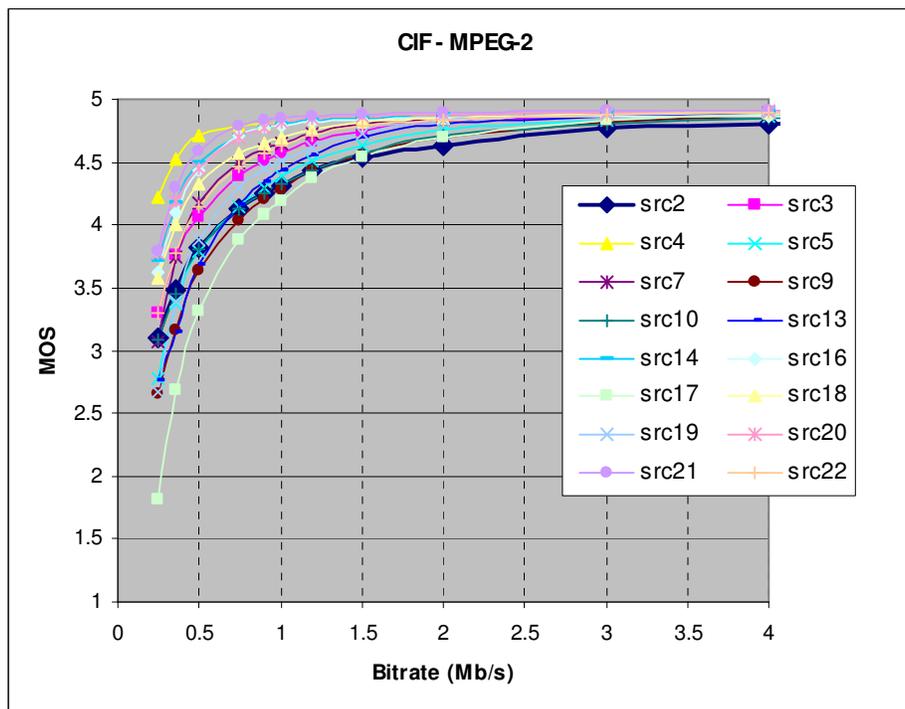


Figura 3.1

Por otra parte, en la Figura 3.2, tomada de [29], se muestra como varía el ancho de banda requerido usando diversos codificadores, en función de la calidad de la imagen, para una secuencia de video particular (“Tempete”, src22), en resolución CIF a 15 Hz. Puede verse como para una misma calidad (medida en este caso como PSNR), MPEG-2 requiere de aproximadamente el doble de ancho de banda que H.264/AVC.

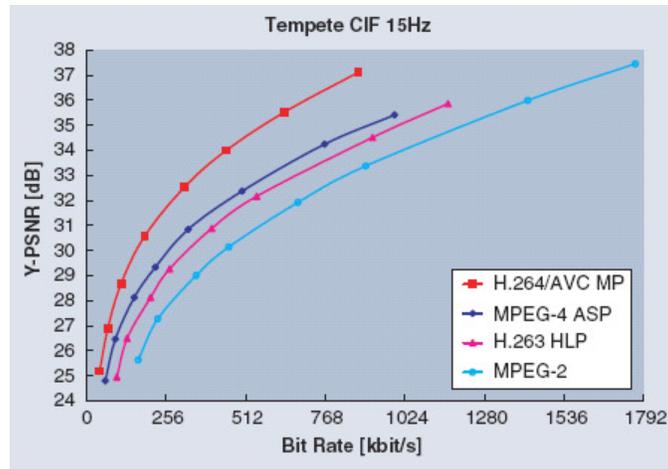
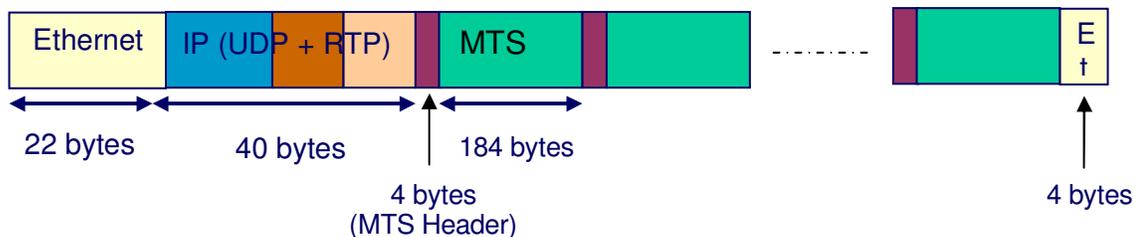


Figura 3.2

Como se puede ver, el ancho de banda de las señales de video puede variar notoriamente, desde valores cercanos a los 64 kb/s (para baja resolución de pantalla, imágenes con poco movimiento, baja cantidad de cuadros por segundo), hasta varios Mb/s (para resoluciones medias o altas).

En redes IP, el “overhead” o sobrecarga depende de la forma de encapsulado utilizada. Como se mencionó anteriormente (Ver 2.5) los protocolos IP/UDP/RTP tienen 40 bytes, y Ethernet otro 26 bytes. En el caso de MPEG-2, utilizando MTS encapsulados en RTP, se pueden incluir hasta 7 paquetes MTS dentro de un mismo paquete IP. Cada MTS tiene 4 bytes de cabezal y 184 bytes de contenido. Por tanto, un paquete IP con MPEG-2 está formado como se muestra en la siguiente figura:



En un paquete IP se pueden incluir  $7 \times 184 = 1288$  bytes de contenido MPEG-2, y por otra parte hay  $40 + 4 \times 7 = 68$  bytes de cabezales a nivel de capa 3 (IP) y 94 bytes de cabezales a nivel de capa 2 (en Ethernet). Por lo tanto, el ancho de banda de MPEG-2 transportado en RTP es 5.3% ( $68/1288$ ) mayor que el ancho de banda propio del video en capa 3 (IP) y 7.3 % ( $94/1288$ ) mayor que el ancho de banda propio del video en capa 2 (Ethernet).

En el caso de H.264 encapsulado directamente sobre RTP, sin utilizar TS, se pueden enviar hasta 1430 bytes de “payload” en un paquete IP/UDP/RTP, por lo que el ancho de banda en capa 3 es 2.8% ( $40/1430$ ) mayor que el del propio video codificado y en capa 2 es 4.6% ( $66/1430$ ) mayor que el del propio video codificado.

## 4 Calidad de voz en redes IP

La VoIP enfrenta problemáticas propias de las redes de datos, que se manifiestan como degradaciones en la calidad del servicio (QoS) o la calidad de la experiencia (QoE) percibida por los usuarios. Estas degradaciones pueden deberse por ejemplo a retardos, jitter (diferencia de retardos) y pérdida de paquetes, entre otros factores. Para que la tecnología de VoIP pueda ser utilizada tanto a nivel corporativo como a nivel de operadores telefónicos, es esencial garantizar una calidad de voz aceptable. Un análisis acerca de la medida de la calidad de voz se puede ver en [30].

Se analizarán a continuación los factores específicos que afectan la calidad de voz percibida sobre redes de paquetes.

### 4.1 Factores que afectan la calidad de la voz sobre redes de paquetes

Realizaremos una pequeña discusión acerca de los parámetros que influyen en la calidad de la voz transmitida a través de la red de datos:

#### 4.1.1 Factor de compresión y codificación

Para poder transmitir la voz a través de una red de datos, es necesario realizar previamente un proceso de digitalización y codificación, el que puede degradar la señal de voz original, debido a la utilización de técnicas de compresión (Ver 2.1).

#### 4.1.2 Pérdida de paquetes

A diferencia de las redes telefónicas, donde para cada conversación se establece un vínculo “estable y seguro”, las redes de datos admiten la pérdida de paquetes. Esto está previsto en los protocolos “seguros” de alto nivel, y en caso de que ocurra, los paquetes son reenviados. En los protocolos diseñados para tráfico de tiempo real generalmente no se recibe confirmaciones de recepción de paquetes, ya que si el canal es suficientemente seguro, estas confirmaciones cargan inútilmente al mismo.

En aplicaciones de voz y video, el audio es “encapsulado” en paquetes y enviado, sin confirmación de recepción de cada paquete.

Si el porcentaje de pérdida es pequeño, la degradación de la voz también lo es. Los porcentajes de pérdida admisibles dependen de otros factores, como por ejemplo la demora de transmisión y el factor de compresión de la voz.

Existen técnicas para hacer menos sensible la degradación de calidad en la voz frente a la pérdida de paquetes. La más sencilla consiste en simplemente repetir el último paquete recibido.

También cuentan como “perdidos” los paquetes que llegan a destiempo o fuera de orden.

Existen métodos para mitigar el efecto de la pérdida de paquetes. Un ejemplo se describe en el Anexo I de la Recomendación ITU-T G.711 [31], donde se detalla un método de cancelación de paquetes perdidos (PLC, Packet Loss

Concealment). Este método propone regenerar la forma de onda del paquete perdido en base a información extraída de la señal previa a la pérdida del paquete.

### 4.1.3 Demora

Un factor importante en la percepción de la calidad de la voz es la demora. La demora total está determinada por varios factores, entre los que se encuentran

- Demora debida a los algoritmos de codificación  
En forma genérica, cuanto mayor es la compresión, más demora hay en el proceso (los codecs requieren más tiempo para codificar cada muestra).

Algoritmo de muestreo/compresión	Demora típica introducida
G711 (64 kb/s)	125 $\mu$ s
G.728 (16 kb/s)	2.5 ms
G.729 (8 kb/s)	10 ms
G.723 (5.3 o 6.4 kb/s)	37.5 ms
RTAudio (8 kb/s)	40 ms

Tabla 4.1

- Demoras de procesamiento  
Es el tiempo involucrado en el procesamiento de la voz para la implementación de los protocolos. Generalmente puede ser despreciado.
- Demoras propias de la red (latencia)  
Las demoras propias de la red están dadas por la velocidad de transmisión de la misma, la congestión, y las demoras de los equipos de red (routers, switches, etc.)

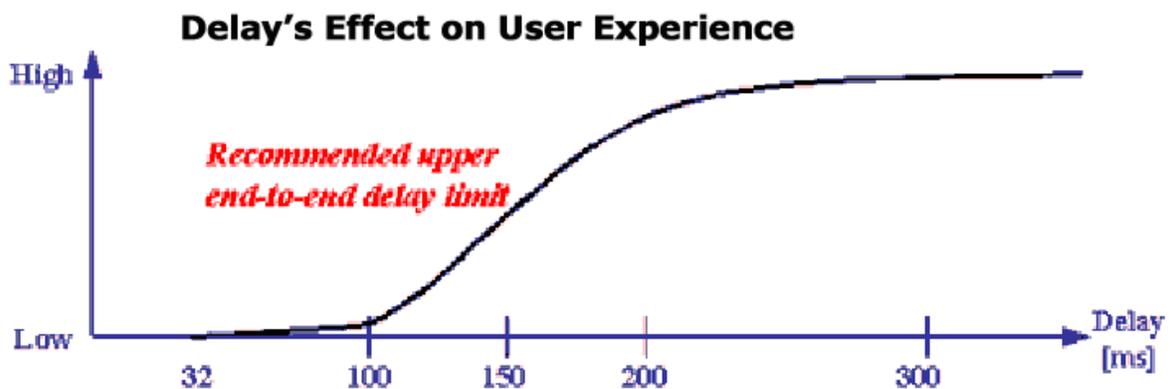


Figura 4.1

Las demoras no afectan directamente la calidad de la voz, sino la calidad de la conversación. Como se puede ver en la Figura 4.1, hasta 100 ms son generalmente tolerados, casi sin percepción de los interlocutores. Entre 100 y 200

ms las demoras son notadas. Al acercarse a los 300 ms de demora, la conversación se vuelve poco natural. Pasando los 300 ms la demora se torna crítica, haciendo dificultosa la conversación.

Un efecto secundario, generado por las demoras elevadas, es el eco. El eco se debe a que parte de la energía de audio enviada es devuelta por el receptor. En los sistemas telefónicos este efecto no tiene mayor importancia, ya que los retardos o demoras son despreciables, y por lo tanto, el “eco” no es percibido como tal.

Cuando la demora de punta a punta comienza a aumentar, el efecto del eco comienza a percibirse.

#### 4.1.4 Eco

Si el tiempo transcurrido desde que se habla hasta que se percibe el retorno de la propia voz es menor a 30 ms, el efecto del eco no es percibido. Asimismo, si el nivel del retorno está por debajo de los  $-25$  dB, el efecto del eco tampoco es percibido. En las conversaciones telefónicas habituales, generalmente existe un retorno de la propia voz en niveles audibles (mayores a  $-25$  dB), pero la demora es mínima, por lo que este retorno no es percibido como eco.

El retorno que produce el eco se produce en diferentes elementos de la red, varios de los cuales se esquematizan en la siguiente figura.

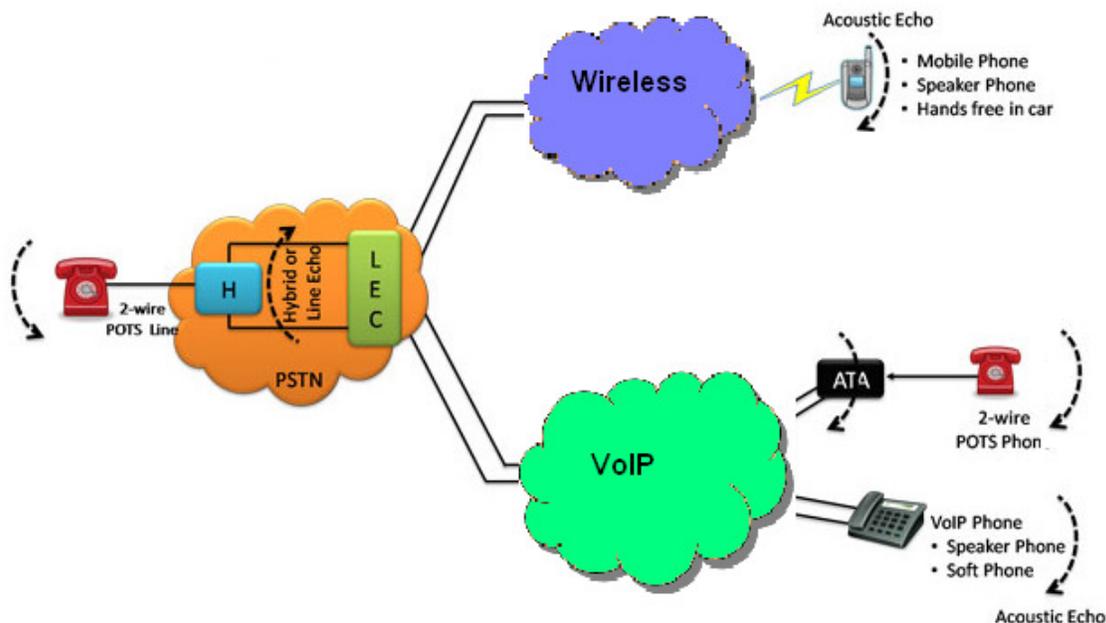


Figura 4.2

Los teléfonos analógicos pueden generar retorno en sus “híbridas”. Las “híbridas” de las tarjetas de abonado también pueden generar retorno. Los teléfonos

celulares tienen el micrófono muy cerca del auricular, y pueden generar retorno acústico. Los teléfonos IP de hardware pueden generar retorno acústico, si se utilizan en “manos libres”. Los teléfonos IP de software pueden generar retorno en la tarjeta de sonido del PC, o en las diademas.

Todos estos retornos pueden ser percibidos como eco, si las demoras entre su generación y su escucha es apreciable. Como las redes IP tienen retardos de punta a punta muy superiores a los existentes en las redes TDM, todos estos retornos se pueden percibir como eco, y deben ser evitados o cancelados.

En la Recomendación ITU-T G.168 se describen las características que deben tener los sistemas digitales diseñados para realizar “cancelación de eco”. Estos sistemas funcionan como se muestra en la Figura 4.4. Mediante un procesamiento digital se evalúa si parte de la señal en el camino de “recepción” (Receive Path) se ha introducido en el camino de “Transmisión” (Send Path), con cierto retardo. Si esto es detectado, la señal del canal de “Transmisión” es procesada, restándole la estimación de la señal correspondiente al eco. Luego de este proceso, la señal pasa por un proceso no lineal, que suprime las señales que están por debajo de cierto umbral.

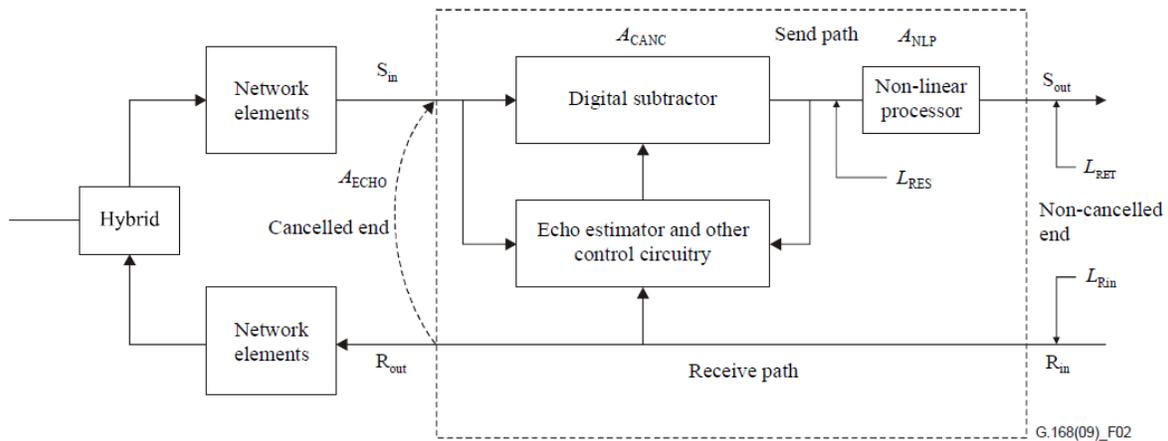


Figura 4.3

La mayoría de los sistemas que utilizan VoIP disponen de canceladores de eco en algún punto del camino de audio.

#### 4.1.5 Variaciones en la demora (Jitter)

El “jitter” es la variación en las demoras (latencias). Por ejemplo, si dos puntos comunicados reciben un paquete cada 20 ms en promedio, pero en determinado momento, un paquete llega a los 30 ms y luego otro a los 10 ms, el sistema tiene un “jitter” de 10 ms.

El receptor debe recibir los paquetes a intervalos constantes, para poder regenerar de forma adecuada la señal original. Dado que el “jitter” es inevitable, los receptores disponen de un “buffer” de entrada, con el objetivo de “suavizar” el efecto de la variación de las demoras. Este buffer recibe los paquetes a intervalos variables, y los entrega a intervalos constantes.

Es de hacer notar que este “buffer” agrega una demora adicional al sistema, ya que debe “retener” paquetes para poder entregarlos a intervalos constantes. Cuánto más variación de demoras (“jitter”) exista, más grande deberá ser el buffer, y por lo tanto, mayor demora será introducida al sistema. Típicamente los jitter-buffers introducen una demora de entre 10 ms a 30 ms.

#### **4.1.6 Tamaño de los paquetes**

El “tamaño” de los paquetes influye en dos aspectos fundamentales en la transmisión de la voz sobre redes de datos: La demora y el “ancho de banda” requerido.

Para poder transmitir las muestras codificadas de voz sobre una red de datos, es necesario armar “paquetes”, según los protocolos de datos utilizados (por ejemplo, IP). Un paquete de datos puede contener varias muestras de voz. Por ello, es necesario esperar a recibir varias muestras para poder armar y enviar el paquete. Esto introduce un retardo o demora en la transmisión. Desde éste punto de vista, parece conveniente armar paquetes con la mínima cantidad de muestras de voz (por ejemplo, un paquete por cada muestra). Sin embargo, hay que tener en cuenta que cada paquete tiene una cantidad mínima de información (bytes) de control (cabezal del paquete, origen, destino, etc.). Esta información (“sobrecarga” u “overhead”), no aporta a la información real que se quiere transmitir, pero afecta al tamaño total del paquete, y por tanto al ancho de banda, como se vio en 2.5.

La duración de las “ventanas” de voz se encuentran entre 10 a 30 ms, valor que aporta a la demora total.

## **4.2 Estimación de la calidad de voz en redes de paquetes: ITU-T G.107 (E-Model)**

La industria de las telecomunicaciones ha aceptado una representación numérica de la calidad de la voz, llamada “MOS” (Mean Opinion Score), y estandarizada en la recomendación ITU-T P.800. La calidad de la voz es calificada con un número, entre 1 y 5. El valor numérico de MOS es proporcional a la calidad de la voz. 1 significa muy mala calidad y 5 significa excelente. Los valores son obtenidos mediante el promedio de las opiniones de un gran grupo de usuarios.

La ITU-T ha creado un “modelo” en la recomendación ITU-T G.107, llamado “E-Model” [32], para estimar o predecir la calidad de la voz en redes IP (VoIP)

percibida por un usuario típico, en base a parámetros medibles de la red. El resultado del E-Model es un factor escalar, llamado “R” (“Transmission Rating Factor”), que puede tomar valores entre 0 y 100.

El “E-model” toma en cuenta una gran cantidad de factores que pueden deteriorar la calidad de la voz percibida, como por ejemplo, el uso de compresión, los retardos de la red, así como también los factores “típicos” en telefonía como ser pérdida, ruido y eco. Puede ser aplicado para estimar la calidad de la voz en redes de paquetes, tanto fijas como inalámbricas [33].

El E-Model puede ser utilizado para evaluar como se verá afectada la calidad de la voz en una red en base a parámetros mensurables. El modelo parte de un puntaje “perfecto” (100) y resta diversos factores que degradan la calidad, según se puede ver en la ecuación (4.2.1).

$$R = R_0 - I_s - I_d - I_{e,eff} + A \quad (4.2.1)$$

donde

$R_0$  Representa la relación señal/ruido básica (antes de ingresar en la red) que incluye fuentes de ruido, tales como ruido ambiente. El valor inicial puede ser como máximo 100. Las fuentes de ruido independientes del sistema como ser el ruido ambiental, pueden hacer que este valor inicial sea menor a 100.

$I_s$  Es una combinación de todas las degradaciones que aparecen de forma más o menos simultánea con la señal vocal. Por ejemplo, volumen excesivo y distorsión de cuantización.

$I_d$  Representa las degradaciones producidas por el retardo y el eco

$I_{e,eff}$  “Effective equipment impairment factor”. Representa las degradaciones producidas por los códecs y por las pérdidas de paquetes de distribución aleatoria.

$A$  Factor de Mejoras de Expectativas. Muchas veces, los usuarios están dispuestos a aceptar peor calidad de voz si saben que se están utilizando tecnologías “no clásicas” (por ejemplo celulares o VoIP). Permite compensar los factores de degradación cuando existen otras ventajas de acceso para el usuario.

Los valores de R varían entre 0 y 100, correspondiendo los valores más altos a mejores calidades de voz.

Los tres tipos de degradaciones ( $I_s$ ,  $I_d$  y  $I_{e,eff}$ ) se subdividen, a su vez, en la combinación de otros factores, como se detalla a continuación.

### Cálculo de $I_s$

$$I_s = I_{olr} + I_{st} + I_q \quad (4.2.2)$$

donde

$I_{olr}$  Representa la disminución de calidad producida por valores demasiado bajos de OLR (Overall Loudness Rating). El OLR se calcula, a su vez, como

$$OLR = SLR + RLR \quad (4.2.3)$$

Siendo

SLR (Send Loudness Rating), es la pérdida entre la boca del emisor y el micrófono del aparato telefónico

RLR (Receive Loudness Rating), es la pérdida entre el parlante del aparato telefónico y el oído del receptor

$I_{st}$  Representa la degradación producida por efectos locales no óptimos, y depende esencialmente del factor STMR (Side Tone Masking Rating). Parte de la señal recibida por el micrófono es transmitida, dentro del mismo teléfono, al parlante, generando un “efecto local” que hace que la persona que habla se escuche por el oído en el que tiene el tubo o microteléfono. La atenuación de la señal que pasa del micrófono al parlante del mismo aparato se conoce como STMR. Si este valor no está dentro de los parámetros adecuados, genera una sensación de “eco”, o de “línea muerta”, según el caso, bajando la calidad de la comunicación.

$I_q$  Representa la degradación producida por la distorsión de cuantificación. Se calcula en base a “unidades qdu” . 1 qdu se define como el “ruido” de cuantización” que resulta de una codificación y decodificación completas en Ley A o Ley  $\mu$

La fórmula de cálculo detallada de los parámetros ( $I_{olr}$ ,  $I_{st}$ ,  $I_q$ ) puede verse en la recomendación G.107 [32].

### Cálculo de $I_d$

$$I_d = I_{dte} + I_{dle} + I_{dd} \quad (4.2.4)$$

Donde

$I_{dte}$  Expresa una estimación para las degradaciones debidas al eco para el hablante. Se calcula en base al factor TELR (Talker Echo Loudness

Rating) y la demora media T de punta a punta en un sentido. El factor TELR es la medida de la atenuación del eco percibido por el hablante.

$I_{dle}$  Representa degradaciones debidas al eco para el oyente. Se calcula en base al factor WEPL (Weighted Echo Path Loss) y la demora media Tr de ida y vuela. El factor WEPL es la medida de la atenuación entre la señal “directa” recibida por el oyente, la señal retardada recibida como eco.

$I_{dd}$  Representa la degradación producida por retardos absolutos demasiado largos  $T_a$ , que se producen incluso con compensación perfecta del eco. Si  $T_a < 100$  ms, el factor  $I_{dd}$  es 0.

La fórmula de cálculo detallada de los parámetros ( $I_{dte}$ ,  $I_{dle}$ ,  $I_{dd}$ ) puede verse en la recomendación G.107.

El efecto de la demora en el valor de R se grafica en la Figura 4.4, asumiendo todos los otros factores ideales [34].

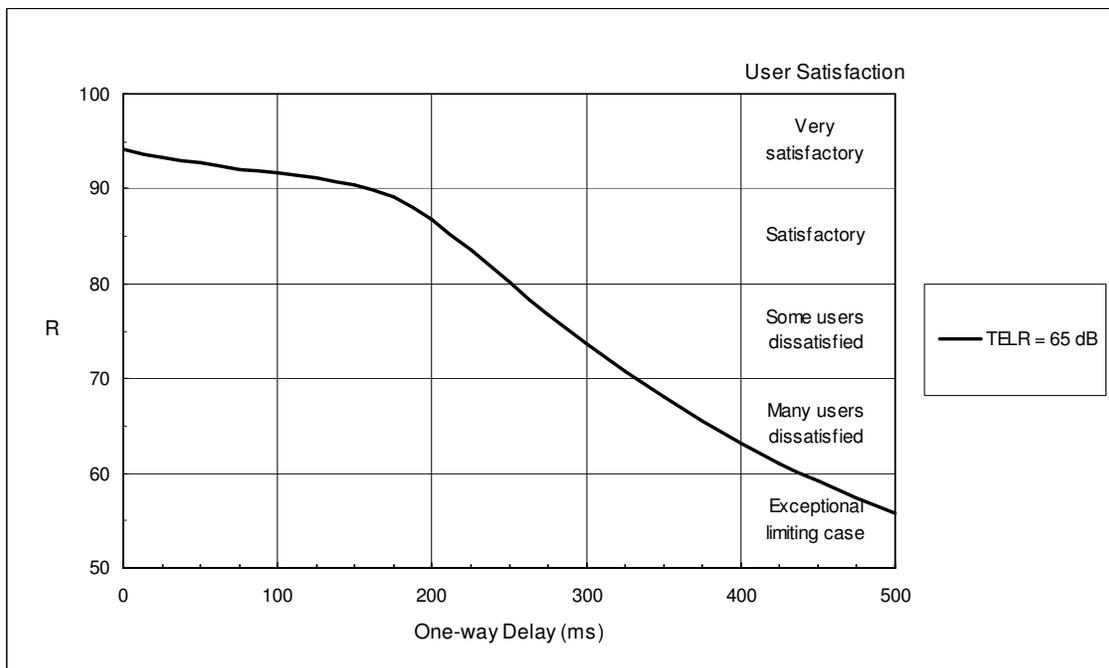


Figura 4.4

Puede verse como hasta 175 ms el valor de R es mayor que 90, y se encuentra en la zona de “Muy satisfechos”. Sin embargo, luego de los 175 ms, el efecto de las demoras degrada fuertemente la comunicación, haciéndola poco natural.

Si a la gráfica anterior se le suma el efecto del eco, el modelo E predice las curvas presentadas en la Figura 4.5.

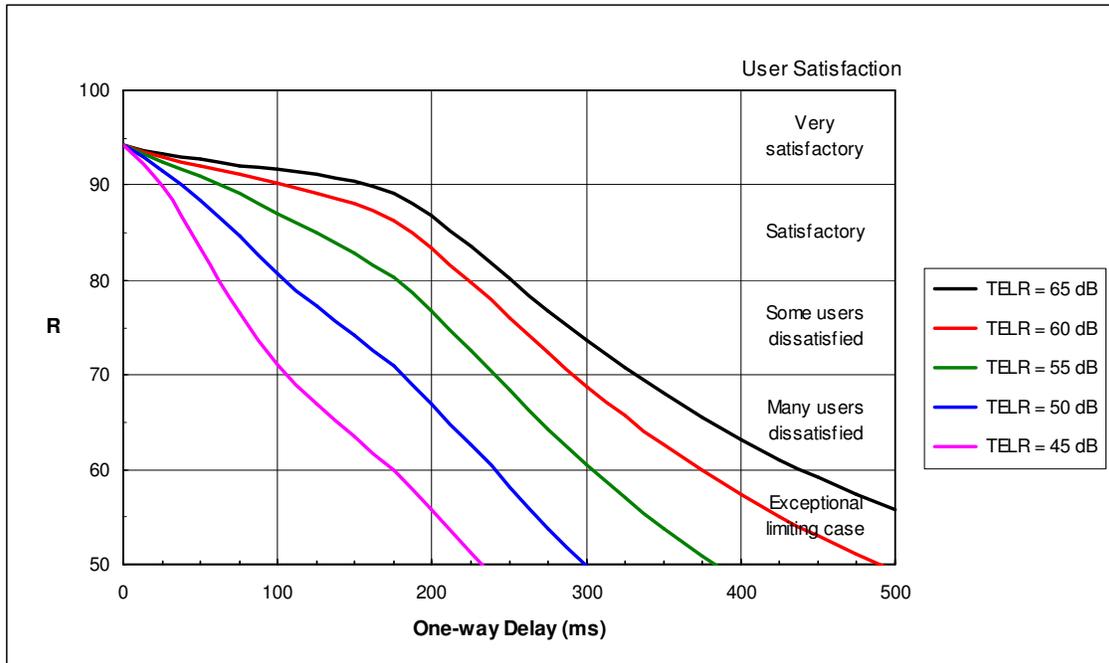


Figura 4.5

Es de hacer notar que el valor TELR es la medida de la atenuación del eco percibido por el hablante. Cuanto más atenuado el eco percibido (mayor valor en db de TELR), menor efecto tiene el eco sobre la degradación. En la medida que aumenta el eco, el valor de R decrece rápidamente con el retardo.

### Cálculo de I<sub>e-eff</sub>

I<sub>e-eff</sub> representa las degradaciones producidas por los códecs y por las pérdidas de paquetes, según la siguiente fórmula:

$$I_{e-eff} = I_e + (95 - I_e) \cdot \frac{P_{pl}}{\frac{P_{pl}}{BurstR} + B_{pl}} \quad (4.2.5)$$

Donde

I<sub>e</sub> Es un valor que depende del codec utilizado, y representa la degradación percibida producida por los diferentes algoritmos de compresión.

P<sub>pl</sub> Representa la probabilidad de pérdida de paquetes

B<sub>pl</sub> Se define como el “factor de robustez” contra pérdida de paquetes, y es un valor preestablecido para cada codec

BurstR Es la “Relación de ráfaga”, y se define como

$$BurstR = \frac{\text{Longitud media de las ráfagas observadas en una secuencia de llegada}}{\text{Longitud media de las ráfagas previstas en la red en condiciones de pérdida "arbitraria"}}$$

Si no existen pérdida de paquetes ( $P_{pl}=0$ ), el factor  $l_{e-eff}$  depende únicamente del tipo de codec utilizado

Los valores de  $l_e$  para los diferentes codecs se detallan en la Tabla 4.2:

Codec Type	Reference	Operating Rate kbit/s	$l_e$ Value
<b>Waveform Codecs</b>			
PCM	G.711	64	0
ADPCM	G.726, G.727	40	2
	G.721, G.726, G.727	32	7
	G.726, G.727	24	25
	G.726, G.727	16	50
<b>Speech Compression Codecs</b>			
LD-CELP	G.728	16	7
		12.8	20
CS-ACELP	G.729	8	10
	G.729-A + VAD	8	11
VSELP	IS-54	8	20
ACELP	IS-641	7.4	10
QCELP	IS-96a	8	21
RCELP	IS-127	8	6
VSELP	Japanese PDC	6.7	24
RPE-LTP	GSM 06.10, Full-rate	13	20
VSELP	GSM 06.20, Half-rate	5.6	23
ACELP	GSM 06.60, EFR	12.2	5
ACELP	G.723.1	5.3	19
MP-MLQ	G.723.1	6.3	15

Tabla 4.2

En una red sin pérdida de paquetes y sin eco, el valor de R dependerá de la demora y de los codecs utilizados, según se muestra en la Figura 4.6, para G.711, G.729A y G.723.1 (notar que la gráfica “negra” coincide con las gráficas anteriores)

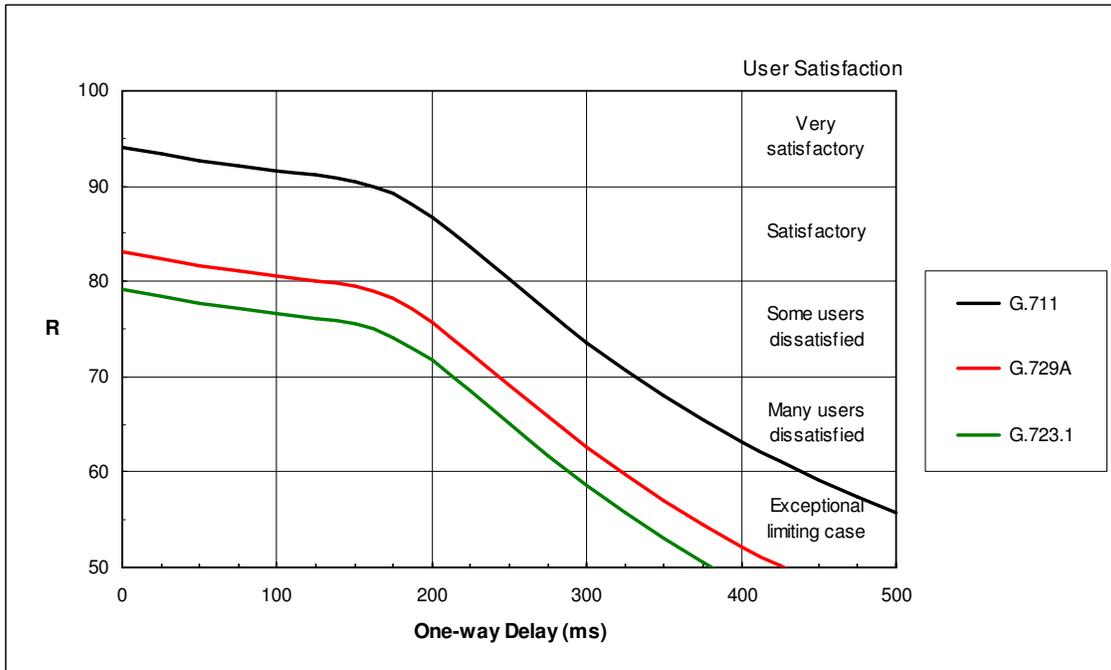
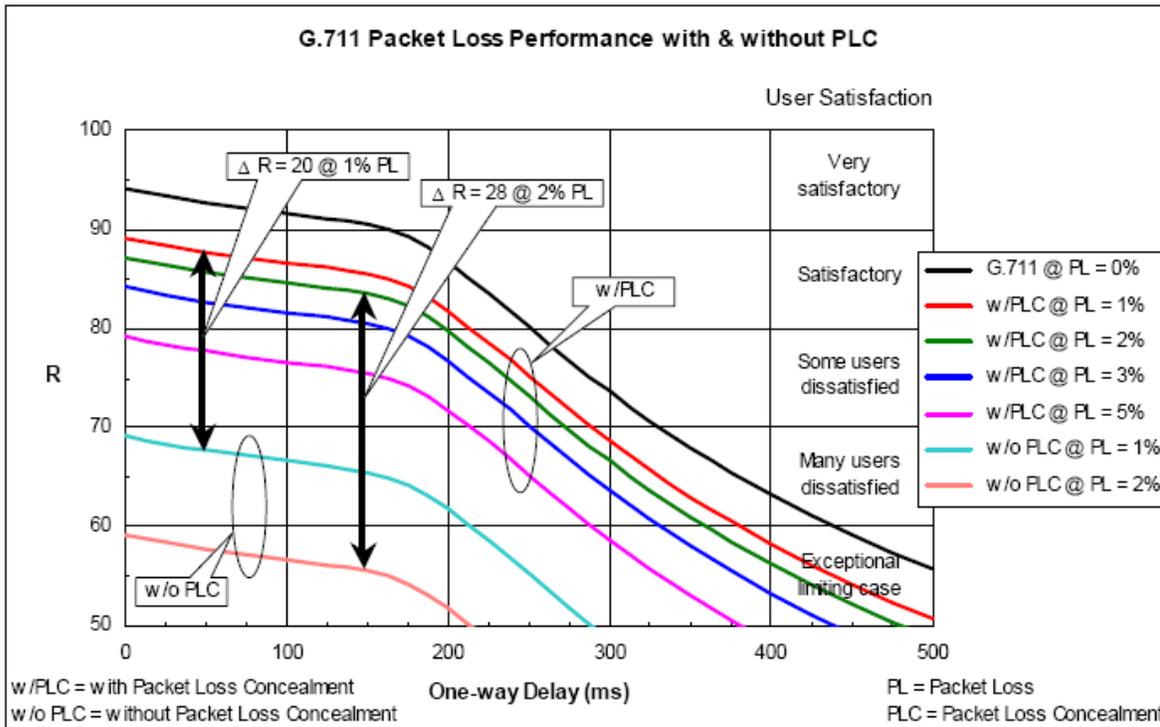


Figura 4.6

En una red con pérdida de paquetes, el valor de R depende de la utilización o no de técnicas de PLC (Packet Loss Concealment). A modo de ejemplo, para los codecs G.711 y G.729, las gráficas de la Figura 4.7 muestran el efecto conjunto de la demora y la pérdida de paquetes.



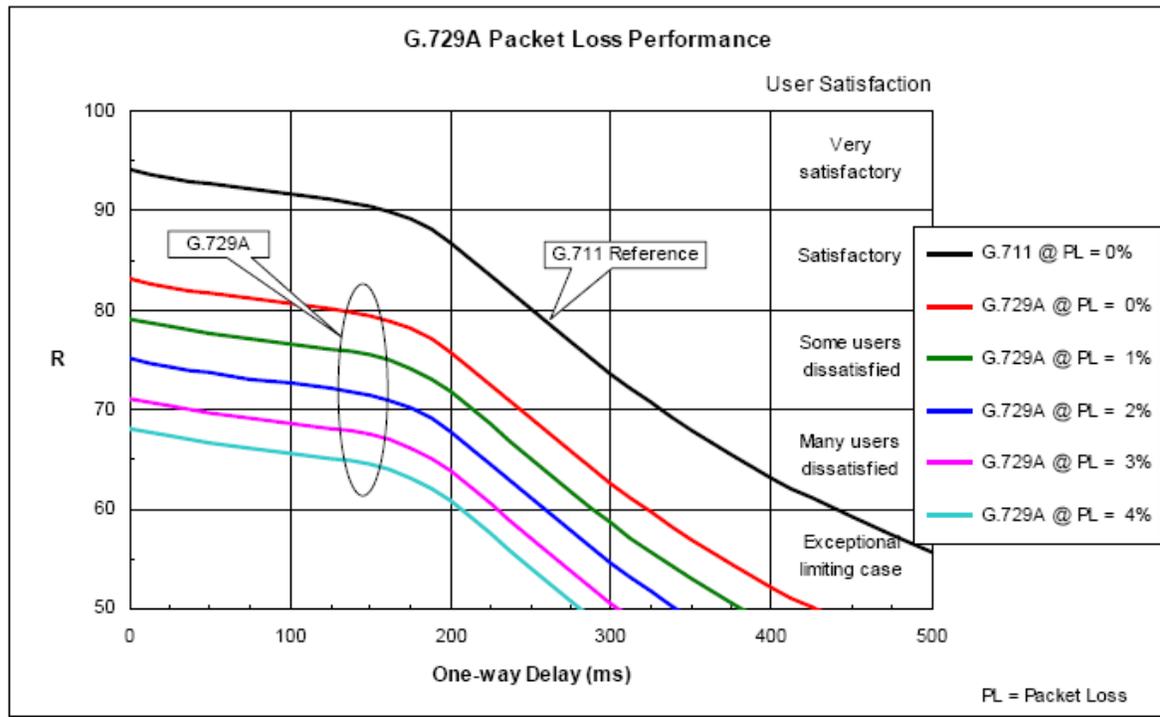


Figura 4.7

### Cálculo de A

A representa un “Factor de Mejoras de Expectativas”. Muchas veces, los usuarios están dispuestos a aceptar peor calidad de voz si saben que se están utilizando tecnologías “no clásicas” (por ejemplo celulares o VoIP). No existe, por consiguiente, ninguna relación entre A y los demás parámetros de transmisión.

El cuadro siguiente presenta los valores típicos de A para diferentes tecnologías, según la recomendación ITU-T G-113 [35].

Ejemplo de sistema de comunicación	Valor máximo de A
Convencional (alámbrico)	0
Movilidad mediante redes celulares en un edificio	5
Movilidad en una zona geográfica o en un vehículo en movimiento	10
Conexión con lugares de difícil acceso, por ejemplo, mediante conexiones de múltiples saltos por satélite	20

Tabla 4.3

**Relación de R y MOS**

El modelo relaciona el valor de “R” con el “MOS”, con un gran nivel de aproximación, según la siguiente ecuación:

Para  $R < 6.5$ :  $MOS_{CQE} = 1$

Para  $6.5 < R < 100$ :  $MOS_{CQE} = 1 + 0,035R + R(R - 60)(100 - R)7 \cdot 10^{-6}$  (4.2.6)

Para  $R > 100$ :  $MOS_{CQE} = 4,5$

La Figura 4.8y la Figura 4.9 muestran la relación entre R y MOS, según la fórmula anterior:

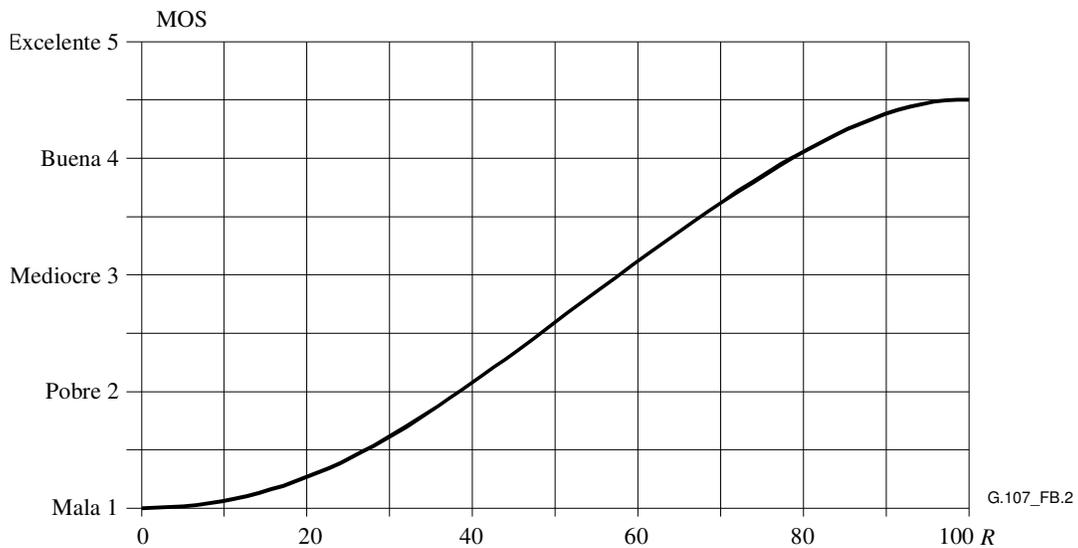


Figura 4.8

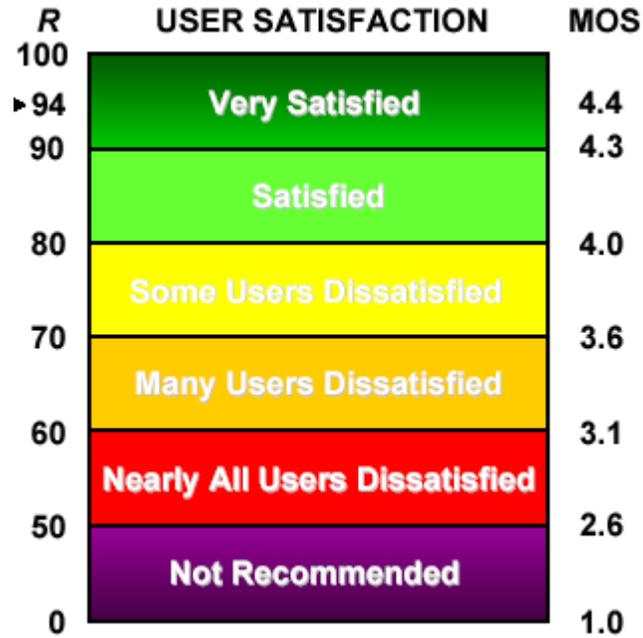


Figura 4.9

**Aplicación del E-model**

El RFC 3611 [36] define campos de “reportes extendidos” (XR, Extended Reports) en el protocolo RTCP que permiten intercambiar información acerca de la calidad de la comunicación. En este RFC se incluye la posibilidad de intercambiar información del valor de “R” entre fuentes y destinos, así como los valores percibidos de MOS-LQ (MOS listening quality) y MOS-CQ (MOS conversational quality)

En la Figura 4.10 se muestra un ejemplo de un paquete RTCP según el RFC 3611. Se puede ver el intercambio de información donde se incluye el valor de R, MOS-LQ y MOS-CQ.

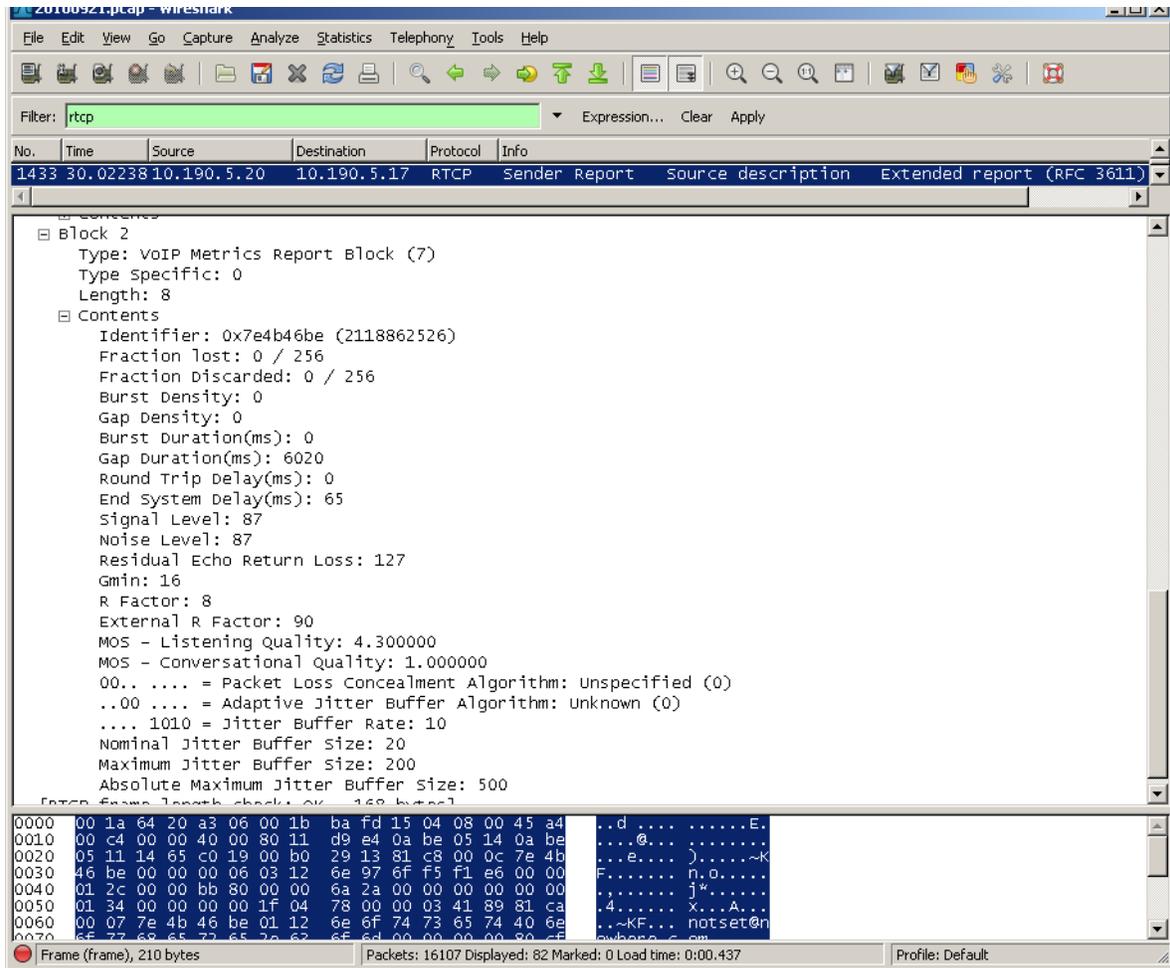


Figura 4.10

## 5 Calidad de video en redes IP

La transmisión de video y multimedia sobre redes de datos enfrenta problemáticas específicas en lo que respecta a la calidad percibida por los usuarios. Varios tipos de degradaciones suelen presentarse en las señales de video transmitidas sobre redes de paquetes. El estudio en esta área es todavía un tema de investigación [37]. Se analizarán a continuación los factores que pueden afectar la calidad de video percibida, y luego los métodos aceptados para medirla en forma subjetiva y objetiva.

### 5.1 Factores que afectan la calidad del video sobre redes de paquetes

La transmisión de video sobre redes de paquetes, y en particular, sobre la Internet, presenta características esencialmente diferentes a la de la difusión de TV por las vías clásicas (radiofrecuencia, TV-cable). Se utilizan rangos de anchos de banda variables, hay congestión y pérdida de paquetes, y típicamente, en las aplicaciones corporativas, la observación se realiza desde distancias más cortas, y generalmente en pantallas más pequeñas.

En esta sección se describirán conceptualmente los factores específicos de las redes IP que afectan a la calidad de video.

#### 5.1.1 Factor de compresión

El proceso de digitalización de video utiliza técnicas que transforman una secuencia de píxeles al dominio de la frecuencia espacial (DCT), cuantificando valores, descartando eventualmente componentes de alta frecuencia, y haciendo uso de técnicas de predicción y compensación de movimientos. Esto genera “ruido de cuantificación”, el que puede degradar la imagen original a niveles perceptibles. Es este “ruido de cuantificación” [38], el que genera las clásicas degradaciones que pueden verse en imágenes y videos con alta compresión, entre ellos, el conocido “efecto de bloques”, que hace ver a la imagen como un conjunto de bloques pequeños.

Los algoritmos de compresión utilizados actualmente en la codificación digital de video introducen varios tipos de degradaciones, las que se pueden clasificar según sus características principales [39]. Esta clasificación es útil para poder comprender las causas de las degradaciones y el impacto que tiene en la calidad percibida.

Debido al gran ancho de banda requerido en video para enviar la señal sin comprimir, el uso de codecs con compresión es sumamente habitual en video. Esto pone un límite a la calidad de imagen recibida, el que es independiente del medio de transmisión sobre el que viaje la señal.

Las principales degradaciones introducidas por el uso de codecs con compresión son las siguientes:

- Efecto de bloques (blocking)
- Efecto de imagen de base (basis image)
- Borrosidad o falta de definición (Blurring)
- Color bleeding (Corrimiento del color)
- Efecto escalera y Ringing
- Patrones de mosaicos (Mosaic Patterns)
- Contornos y bordes falsos
- Errores de Compensaciones de Movimiento (MC mismatch)
- Efecto mosquito
- Fluctuaciones en áreas estacionarias
- Errores de crominancia

### 5.1.2 Pérdida de paquetes

La pérdida de paquetes en las redes IP afecta a la calidad percibida de video. En la Figura 5.1, tomada de [40], se muestra como la pérdida de un paquete puede propagarse, afectando no solo a la información de video contenida en dicho paquete, sino a otras partes del mismo o diferentes cuadros. Típicamente, y dado que la codificación se realiza en forma diferencial, la pérdida de un paquete afectará a todos los bloques siguientes en la misma fila (“slice”). Si el paquete perdido corresponde además a un cuadro de referencia (I), también se verán afectados los cuadros predictivos, posteriores o anteriores, propagándose el error en el tiempo.

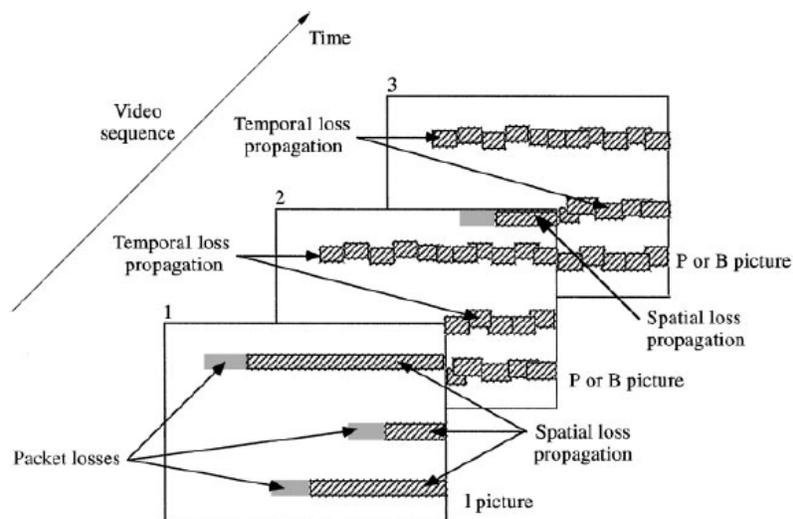


Figura 5.1

Existen técnicas de cancelación de paquetes perdidos, las que tratan de reconstruir la información perdida en base a información disponible. Por ejemplo, reemplazando los píxeles perdidos por los mismos valores de cuadros anteriores.

Varios trabajos se han realizado, estudiando la manera en que la calidad de video se ve afectada por la pérdida de paquetes. Algunos [41] proponen estimar el MSE del video, en base a diferentes técnicas, que tienen en cuenta la pérdida de paquetes. Sin embargo, estos son estimadores no se corresponden con la calidad “perceptual”, al basarse en la predicción del MSE o PSNR, los que no tienen relación directa con el MOS.

En [42] se muestra que no basta con conocer el porcentaje de pérdida de paquetes para estimar como se ve afectada la calidad de video percibida. Dependiendo de diversos factores, la pérdida de un paquete determinado de video puede o no afectar la calidad percibida. Por ejemplo, en imágenes casi estáticas, el video perdido puede ser reconstruido en base a imágenes anteriores, casi sin pérdida de calidad, lo que lleva a que la pérdida de un paquete sea prácticamente imperceptible. Algo similar sucede cuando la pérdida solo afecta a un cuadro. De esta manera, se propone un “clasificador de paquetes perdidos”, que, en base a un algoritmo, decide si el paquete perdido afectará o no a la calidad percibida del video. El algoritmo toma en cuenta cuantos cuadros se verán afectados por la pérdida del paquete, la movilidad de la imagen y su varianza y el error introducido medido con MSE, entre otros factores. Se define el parámetro VPLR (Visible Packet Loss Rate), en lugar del clásico PLR (Packet Loss Rate), para ser utilizado como entrada a los algoritmos de predicción de calidad en base a la pérdida de paquetes.

La idea de detectar como afecta la posible pérdida de cada paquete en la calidad perceptual es explorada en [43], donde se propone un método que garantice una calidad perceptual constante en la recepción de un flujo de video. La idea en este caso es marcar solo ciertos paquetes específicos con mayor prioridad (asumiendo una red con soporte para Diff Serv), de manera de “garantizar” su llegada a destino, sin inundar a la red con todos los paquetes de video marcados como prioritarios, sino solo con los paquetes cuya pérdida afecten especialmente la calidad percibida. El algoritmo se basa únicamente en la calidad deseada (PSNR) y la tasa de pérdida de paquetes (PLR).

Una idea similar es presentada en [44], donde se presenta una métrica para priorizar ciertos paquetes de video, en base a la estimación de la distorsión percibida en caso de la pérdida o llegada fuera de tiempo de cada paquete.

Dado que las pérdidas de paquetes se dan generalmente en ráfagas, la calidad puede verse fuertemente degradada por la pérdida de varios paquetes consecutivos, correspondientes a cuadros consecutivos. En [45] se propone una técnica que consiste en reagrupar los cuadros que se envían, generando un buffer en el codificador de 3 GOPs, y reagrupando el orden en el que se envía la información. Con esto se logra difundir los paquetes perdidos entre varios cuadros

separados en el tiempo, y de esta manera, según los autores, mejorar la calidad percibida.

Si bien varios trabajos se han realizado acerca de la degradación del video debida a la pérdida de paquetes, el tema está aún abierto, y no hay aún estándares ni trabajos sistemáticos de comparación de diferentes modelos.

### **5.1.3 Demora / Jitter**

El receptor debe recibir los paquetes a decodificar a intervalos constantes, para poder regenerar de forma adecuada la señal original. Dado que el jitter es inevitable en las redes de paquetes, los receptores disponen de un buffer de entrada, con el objetivo de “suavizar” el efecto de la variación de las demoras. Este buffer recibe los paquetes a intervalos variables, y los entrega a intervalos constantes.

Es de hacer notar que este buffer agrega una demora adicional al sistema, ya que debe retener paquetes para poder entregarlos a intervalos constantes. Cuánto más variación de demoras (jitter) exista, más grande deberá ser el buffer, y por lo tanto, mayor demora será introducida al sistema. Las demoras son indeseables, y tienen impacto directo en la experiencia del usuario, sobre todo en contenidos de tiempo real (por ejemplo, distribución de eventos deportivos en línea) y en aplicaciones conversacionales (por ejemplo video telefonía, o video conferencias). Se hace necesario disponer de mecanismos que minimicen el tamaño de los jitter-buffers, pero que a su vez, no comprometan la calidad debida a la perdida de paquetes que no han llegado a tiempo.

En [46] se propone un método dinámico al que llaman AMP (Adaptive Media Playout), en el que se cambia la velocidad de reproducción del medio (video / audio) dependiendo de la condición del canal de transmisión, y logrando de esta manera reducir el tamaño del jitter-buffer. Se indica que aumentar o disminuir la velocidad de ciertas partes del contenido hasta en un 25% es subjetivamente mejor que aumentar las demoras totales o tener interrupciones.

En el proyecto Multimedia del VQEG se proponen realizar pruebas con demoras entre 2 milisegundos y 5 segundos, y pérdidas de paquetes impulsivas en el rango de 0 a 50%.

## **5.2 Estimación de la calidad de video sobre redes de paquetes: ITU-T G.1070**

ITU-T ha publicado un modelo de predicción de la calidad de video, para aplicaciones de video telefonía, en base a parámetros medibles de una red IP. Este modelo es similar al E-Model visto anteriormente para audio. Ha sido

desarrollador por NTT (de Japón) y estandarizado en la recomendación ITU-T G.1070 “Opinion model for video-telephony applications” [47].

La Recomendación G.1070 propone un modelo que estima la calidad percibida en el uso de aplicaciones de “video telefonía”, que puede ser utilizado al momento de planificar una red de datos que transmita este tipo de servicios a través de IP

El modelo consiste en tres funciones, una para la estimación de la calidad del video ( $V_q$ ), otra para la estimación de la calidad del audio ( $S_q$ ), y finalmente una para la estimación de la calidad multimedia ( $MM_q$ ). En la Figura 5.2 se esquematiza el funcionamiento general del modelo.

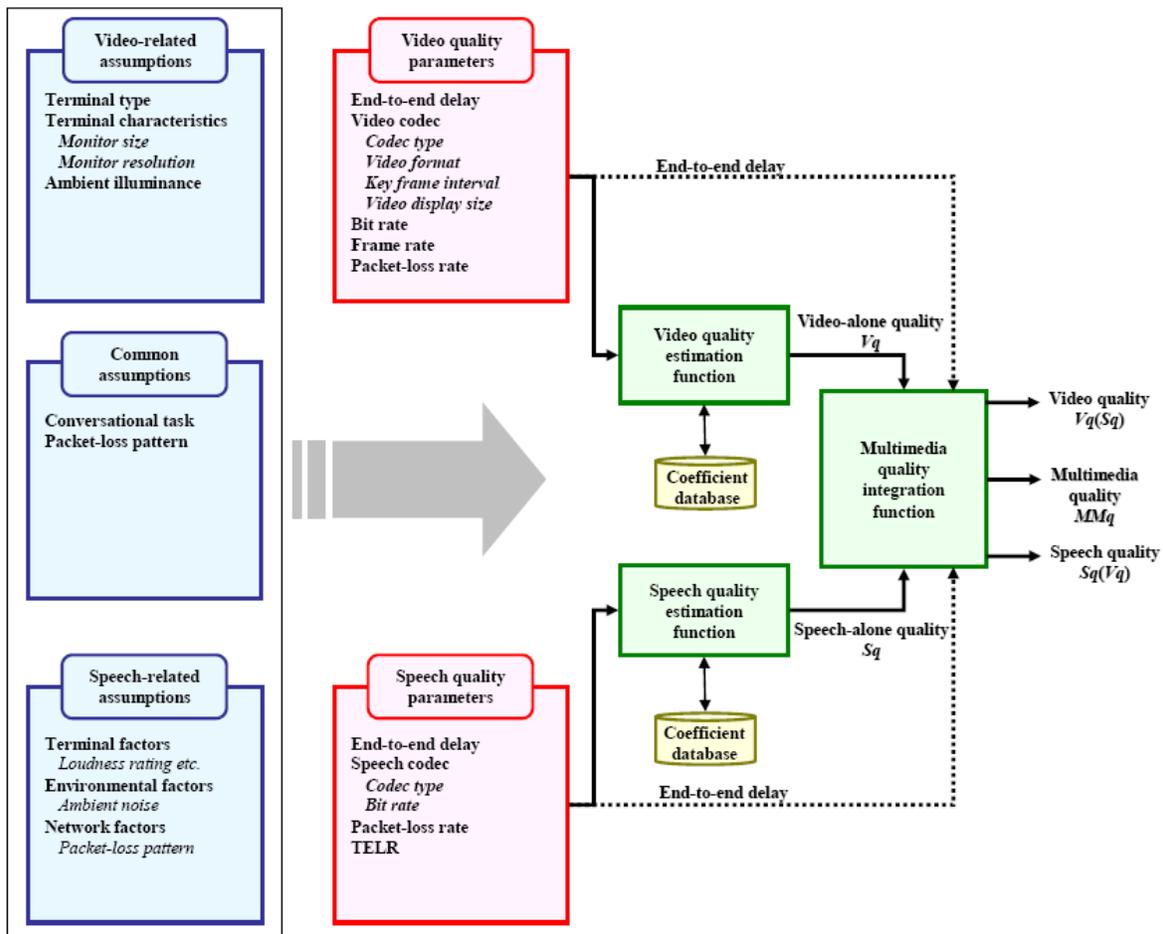


Figura 5.2

La estimación de la calidad de la voz, básicamente, se reduce el E-Model, simplificado.

$$Q = R_o - I_{dte} - I_{e,eff} \quad (5.4.2.1)$$

donde  $Q$  es el equivalente a  $R$  en el E-Model. Los parámetros  $I_{dte}$  y  $I_{e,eff}$  mantienen sus definiciones del E-Model. Las degradaciones introducidas por las demoras son quitadas de la estimación de la calidad de voz, e incluidas en la calidad "multimedia".

La relación entre  $Q$  y  $S_q$  es similar a la relación entre MOS y  $R$  en el E-Model:

$$\text{Para } Q < 0: \quad S_q = 1$$

$$\text{Para } 0 < Q < 100: \quad S_q = 1 + 0,035Q + Q(Q - 60)(100 - Q)7 \cdot 10^{-6} \quad (5.4.2.2)$$

$$\text{Para } Q > 100: \quad S_q = 4,5$$

La estimación de la calidad del video se basa en la siguiente fórmula

$$V_q = 1 + I_c e^{-\frac{P_{plv}}{D_{Pplv}}} \quad (5.4.2.3)$$

Donde  $I_c$  representa la calidad del video dada únicamente por las condiciones de codificación,  $P_{plv}$  es el porcentaje de pérdida de paquetes y  $D_{Pplv}$  representa el grado de robustez respecto a la pérdida de paquetes. Tanto  $I_c$  como  $D_{Pplv}$  dependen del codec utilizado, el bit rate y el frame rate.

La estimación de la calidad multimedia se expresa en la siguiente fórmula

$$MM_q = m_1 MM_{SV} + m_2 MM_T + m_3 MM_{SV} MM_T + m_4 \quad (5.4.2.4)$$

Donde  $MM_{SV}$  representa la calidad audiovisual, y es función de  $(V_q, S_q)$  y  $MM_T$  contiene los factores de calidad asociados a las demoras, tanto del audio como del video, y tiene en cuenta las degradaciones producidas por la falta de sincronismo entre ambos medios.

## 6 Calidad de Servicio en redes de datos

La mayoría de las redes de datos han sido diseñadas como sistemas del tipo “mejor esfuerzo” y brindan tratamientos similares a cualquier tipo de tráfico. El tráfico de paquetes experimenta en este tipo de redes diversos tipos de degradaciones, incluyendo pérdida de paquetes y demoras fijas y variables. Se ha visto, en secciones anteriores, como estos factores introducen degradaciones en la calidad percibida de los servicios multimedia (de voz y video).

En esta sección se describen algunas de las técnicas empleadas en redes IP para implementar estrategias de manejo de “calidad de servicio” (QoS Quality of Service), que permitan tratar en forma diferenciada a cada tipo de tráfico [48]. Estas técnicas de QoS intentan minimizar las degradaciones introducidas en los servicios multimedia, de manera de mantener una calidad percibida (QoE Quality of Experience) aceptable por parte de los usuarios.

### 6.1 QoS en Capa 2

Las recomendaciones IEEE 802.1q [49] y IEEE 802.1p [50] incorporan 4 bytes adicionales a las tramas Ethernet, donde se puede incluir información acerca de VLANs y etiquetas que identifican la “prioridad” de la trama. La Figura 6.1 muestra una trama Ethernet “normal” y una trama Ethernet 802.1q.

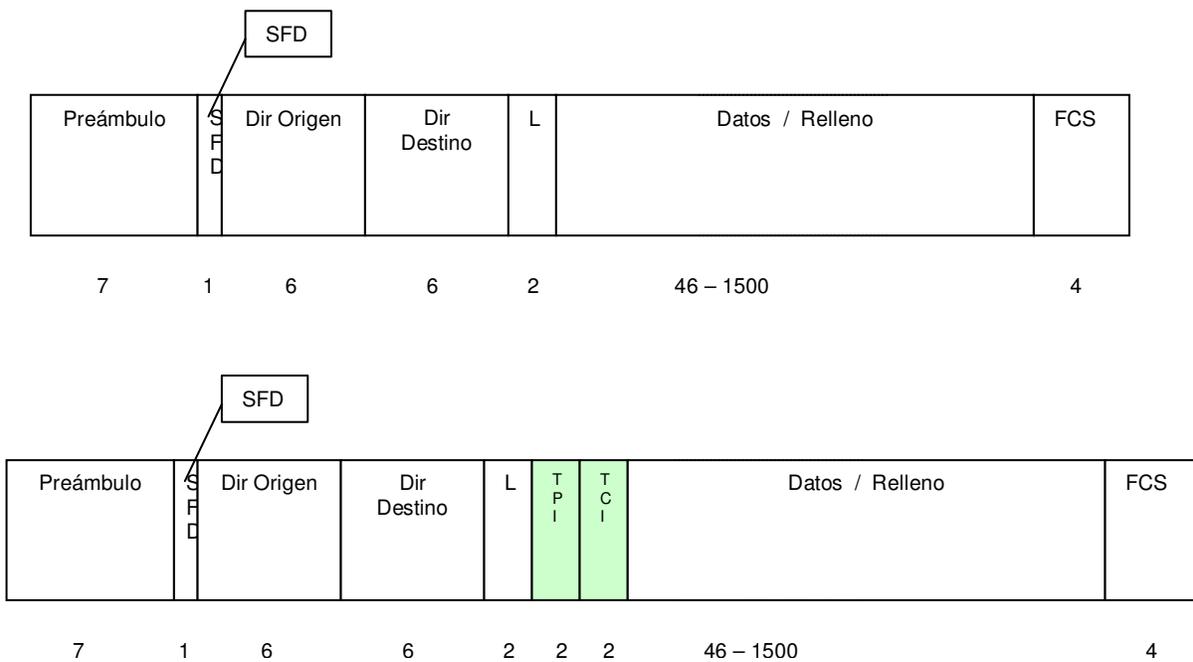


Figura 6.1

Como puede observarse, se agregan 4 bytes: los primeros 2, llamados "TPI" (Tag Protocol Identifier), son fijos (en el valor 81 00) e identifican a la trama como una trama 802.1q. Los segundos 2 bytes, llamados "TCI" (Tag Control Information) se interpretan como 3 conjuntos de bits, de longitud 3 bits, 1 bit y 12 bits respectivamente, como se muestra en la Figura 6.2.

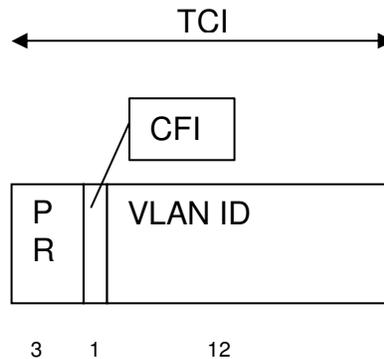


Figura 6.2

Los primeros 3 bits del "TCI" indican la "prioridad" de la trama, de acuerdo a la recomendación IEEE 802.1p. Esto permite tener hasta  $2^3 = 8$  tipos de tráfico. El cuarto bit, llamado CFI (Canonical Format Indicator), indica el orden de los siguientes bits (en formato canónico o no canónico). Los últimos 12 bits indican la VLAN a la cual pertenece la trama. Estos 12 bits permiten tener, por lo tanto hasta  $2^{12} = 4096$  VLANs.

Es de hacer notar que el cabezal de 802.1q contiene la marca de priorización 802.1p, por lo que es necesario disponer de 802.1q para interpretar 802.1p.

Un switch que soporta "calidad de servicio" puede leer el valor del campo "Prioridad" (802.1p), y de acuerdo al valor de dicho campo, transmitir la trama por el puerto de salida con la prioridad apropiada. Para ello implementa 8 colas de salida, una para cada posible valor del campo "prioridad". Las tramas que se encuentran en las colas de mayor prioridad son sacadas, en la boca de salida del switch, antes que las tramas en colas de menor prioridad (ver Figura 6.3).

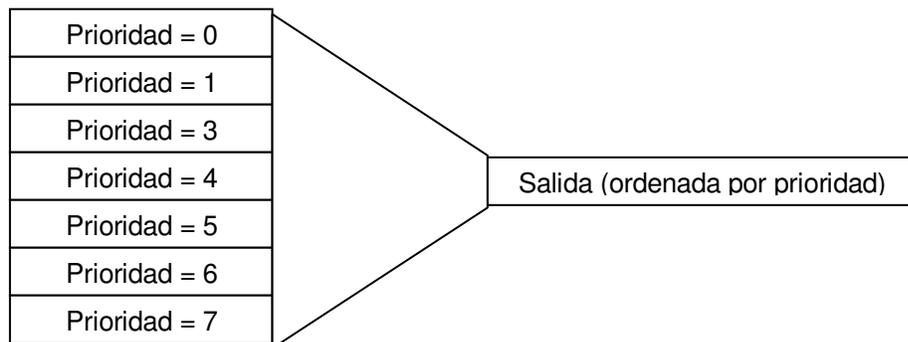


Figura 6.3

Existen diversas estrategias de encolamiento, las que pueden estar total o parcialmente soportadas por los switches o routers de la red:

- **FIFO (First In, First Out):** El primer paquete que haya ingresado en una cola, es el primero en salir.
- **PQ (Priority Queuing):** La salida de los paquetes se realiza según el orden estricto de prioridad, y dentro de cada prioridad, según el orden de llegada. Este tipo de encolamiento puede hacer que, si existe siempre tráfico de alta prioridad, el tráfico de baja prioridad nunca sea enviado.
- **FQ (Fair Queuing):** Es un esquema en el que cada cola se accede en forma circular, asegurando una distribución uniforme de ancho de banda entre todas las colas.
- **WRR (Weighted Round Robin):** Permite asignar diferentes anchos de banda a cada cola.
- **WFQ (Weighted Fair Queuing):** Es una combinación de PQ y FQ, garantizando que aplicaciones de alto tráfico no monopolicen el enlace.

También puede implementarse, a nivel de capa 2, priorización basado en VLANs. Las “VLANs” (Virtual LANs, o redes LAN virtuales) permiten utilizar los mismos medios físicos para formar varias redes independientes, a nivel de la capa 2. Un mismo conjunto de switches pueden implementar, utilizando VLANs, varias redes LAN independientes.

Los criterios para formar las VLAN pueden ser varios. Entre los más comunes se encuentran:

- **VLAN por puertos:** Los puertos de los switches se agrupan en VLANs. De esta manera, las máquinas conectadas a un puerto únicamente “ven” a las máquinas que están conectadas a puertos de la misma VLAN
- **VLAN por direcciones MAC:** Las direcciones MAC se agrupan en VLAN. De esta manera, se puede restringir la red únicamente a ciertas direcciones MAC, independientemente de en que puerto de los switches se conecten.
- **VLAN por protocolo:** Algunos switches que soportan VLAN pueden inspeccionar datos de la capa 3, como el protocolo utilizado, y formar redes independientes según estos protocolos
- **VLAN por direcciones IP:** Las direcciones IP (de capa 3) pueden ser leídas por los switches, y pueden formarse redes independientes con ciertos conjuntos de direcciones IP

Muchos switches de datos permiten implementar cierta priorización del tráfico basado en VLANs. De esta forma, se puede poner a todos los dispositivos de VoIP en la misma VLAN, y darle prioridad frente al tráfico de otras VLANs, dedicadas a aplicaciones de datos. Adicionalmente, en este caso el tráfico de voz no se ve

afectado por el de datos, ya que, a nivel de capa 2, se encontrarían en redes independientes.

## 6.2 QoS en Capa 3

A nivel de capa 3 (protocolo IP en este caso), el mecanismo conocido DiffServ (Differentiated Services) es comúnmente utilizado para gestionar prioridad en los paquetes. DiffServ es un método efectivo para mantener la calidad de servicio (QoS, Quality of Service) en la red, cuando existen equipos de capa 3 (routers, switches de capa 3, etc.). La información de priorización se encuentra en el cabezal del paquete IP, en un campo llamado TOS (Type Of Service), como se muestra en la Figura 6.4. Este campo contiene, a su vez, dos informaciones: DSCP (Differentiated Services Code Point) y ECN (Explicit Congestion Notification).

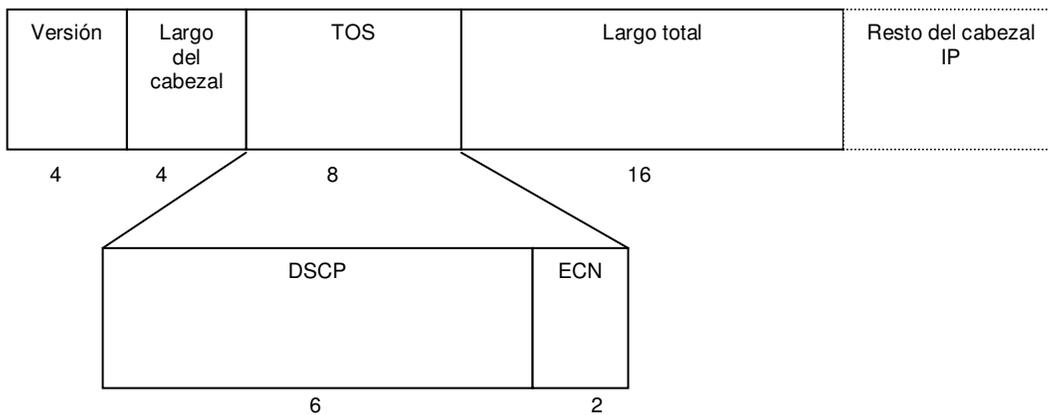


Figura 6.4

En el campo DSCP es posible codificar hasta  $2^6 = 64$  posibles prioridades. De éstas, 32 están reservadas para usos experimentales y 32 pueden ser utilizadas, de las cuales, a su vez, 21 están estandarizadas por el IETF. Las prioridades estandarizadas se dividen en 3 grupos:

- **DE (Default):** Se asume el comportamiento por defecto, utilizando por tanto técnicas de encolamiento de “mejor esfuerzo”. El valor típico de DSCP para este tipo de tráfico es 000000.
- **AF (Assured Forwarding):** Estandarizado en el RFC 2597, donde se definen 4 clases de prioridades dentro de este tipo de priorización.
- **EF (Expedited Forwarding):** Estandarizado en el RFC 2598, establece las máximas prioridades para el tráfico marcado con este identificador. El valor típico de DSCP utilizado es 101110.

El campo ECN permite conocer el estado de congestión del destino. Es utilizado para que el destino pueda indicarle a la fuente, aún antes de perder paquetes, que

existe cierto estado de congestión, de manera que la fuente pueda tomar los recaudos apropiados, por ejemplo, disminuyendo el ancho de banda utilizado. Un valor de ECN = 11 indica que existe congestión. Los valores 10 y 01 indican que no existe congestión. El valor 00 indica que el extremo distante no soporta la función de notificación de congestión.

Otro de los mecanismos utilizados para asegurar cierta calidad de servicio es la reserva de ancho de banda. El protocolo RSVP (Resource Reservation Protocol) establece los mecanismos para reservar cierto ancho de banda en la comunicación entre dispositivos que pasen a través de routers.

El tráfico también puede ser priorizado en base a la dirección IP de origen o destino. Esto puede ser implementado cuando se utilizan direcciones IP estáticas.

### **6.3 QoS en Capa 4 y superiores**

Los paquetes de datos pueden ser priorizados en base a los puertos TCP o UDP. Sin embargo, diferentes aplicaciones podrían utilizar los mismos puertos, por lo que este tipo de priorizaciones debe ser evaluada en cada caso.

Es posible también tener prioridades según el protocolo de capas superiores. Por ejemplo, puede ser priorizado el tráfico RTP respecto a otros, y asignarlo a las colas de alta prioridad.

## 7 Voz y Video sobre redes inalámbricas

La adopción de tecnologías de voz sobre redes de datos inalámbricas, conocidas como VoWLAN (Voice over Wireless LAN) o VoWi-Fi (Voz sobre Wi-Fi) está comenzando a incrementarse en el mercado corporativo. Sin embargo, este tipo de tecnologías presentan desafíos adicionales para obtener calidades de voz aceptables [51].

Una introducción a las redes de datos inalámbricas puede verse en [2]. Diversos aspectos de estas redes inalámbricas (WLAN) deben ser revisadas, y en algunos casos re-diseñadas para soportar estas aplicaciones, como se describe a continuación

### 7.1 Cobertura

Las redes WLAN han sido típicamente diseñadas para aplicaciones de datos. Por tanto, la cobertura muchas veces se limita a las áreas donde se conectan los usuarios (salas de reuniones compartidas, recepción, etc.) Bajas señales de radio frecuencia son soportadas por las aplicaciones típicas de datos (correo electrónico, navegación en Internet, etc.), aún con tasas de errores elevadas.

En contraste, las aplicaciones de telefonía móvil requieren una cobertura extendida, en escaleras, pasillos, áreas de descanso, y diversos sectores donde típicamente no eran áreas de trabajo para conexión de laptops. Asimismo, los puntos de acceso (AP, "Access Points") deben ser ubicados de tal forma que sus áreas de cobertura se solapen lo suficiente como para que las funciones de "roaming", cuando un móvil pasa del área de cobertura de un AP a otro, no produzcan cortes o interrupciones en la comunicación. En aplicaciones de VoWLAN, en algunos casos, se requiere de mayor cantidad de APs, ubicados más cerca entre sí, pero trabajando con menores potencias que en aplicaciones únicamente de datos. Esto requiere de planificar con mayor cuidado las posibles interferencias y el control de potencia de los APs. Para aplicaciones de VoWLAN el límite de las celdas típicamente se establece en -67 dBm [52].

### 7.2 Movilidad

Cuando los usuarios se mueven por el área de cobertura, pueden cambiar su conexión de un AP a otro, efecto llamado "roaming" o "hand off". Si ambos AP pertenecen a la misma red, el cambio se produce a nivel de capa 2, y la dirección IP del móvil se mantiene. Si los AP pertenecen a redes diferentes, el cambio se produce a nivel de capa 3, y el móvil debe solicitar una nueva dirección IP a un servidor DHCP.

Este proceso inevitablemente introduce demoras. Si las demoras son mayores a 100 ms, comienzan a ser percibidas en la conversación. Los procesos de roaming

o hand off pueden llevar un tiempo considerablemente mayor a éste, lo que es inaceptable para aplicaciones de VoIP.

A nivel de capa 2, las demoras introducidas por el roaming son debidas a tres procesos: la búsqueda de un nuevo AP, la “re-asociación” y la “re-autenticación”. Estos procesos pueden tardar de algunos cientos de milisegundos a varios segundos, siendo la re-autenticación el proceso que mas demora. Este proceso típicamente se realiza de acuerdo a la recomendación IEEE 802.11x [53]. A los efectos de reducir estos tiempos, varios fabricantes han implementado mecanismos propietarios de autenticación para los sistemas de VoWLAN.

El IEEE ha estandarizado en julio de 2008 la recomendación 802.11r [54], posibilitando un proceso de roaming rápido. Esto se logra estableciendo la autenticación entre el móvil y el nuevo AP previo o durante el proceso de re-asociación, evitando las demoras que este proceso causa cuando se realiza luego de la transición.

Si el roaming se produce a nivel de capa 3, el móvil debe renovar su dirección IP, generalmente a través de servidores de DHCP, lo que lleva varios segundos. También para evitar estas demoras, varios fabricantes han desarrollado alternativas propietarias.

### **7.3 Calidad de Servicio**

La transmisión de datos en redes inalámbricas utilizan estrategias de acceso al medio muy diferentes a las que se utilizan en redes cableadas. Estas estrategias se basan en evitar las colisiones (CA, “Collision Avoidance”). Cuando la red inalámbrica se comparte entre aplicaciones de voz y de datos, la calidad de la voz puede verse fuertemente afectada, debido a que los paquetes de datos pueden ser excesivamente largos, a velocidades de transmisión relativamente bajas, generando por tanto demoras y jitter mayores a lo que se produce en redes cableadas (cabe recordar que el tamaño de las tramas de capa física en WLAN son mayores a sus correspondientes en Ethernet cableada, y a su vez, las velocidades de transmisión pueden ser considerablemente más bajas).

IEEE ha estandarizado la recomendación IEEE 802.11e [55] en setiembre de 2005, la que establece dos nuevas estrategias de acceso al medio, para asegurar la calidad de servicio: EDCA (Enhanced Distributed Control Access) y HCCA (Hybrid Controlled Channel Access).

EDCA garantiza al tráfico de alta prioridad un acceso al medio estadísticamente más rápido, y establece 4 categorías de acceso: voz, video, mejor esfuerzo y “background”. La categoría “voz” es la que tienen la máxima prioridad. El default es “mejor esfuerzo”.

HCCA consiste en un sistema centralizado de control que permite a las aplicaciones reservar recursos de red basados en sus características de tráfico. Los requerimientos son enviados desde los clientes a los AP. HCCA provee un control parametrizado de la calidad de servicio, controlando la latencia y el ancho de banda de cada cliente. Esto lleva a poder tener mayor cantidad de comunicaciones de voz por AP respecto a EDCA.

La “Wi-Fi Alliance” comenzó a realizar, en 2004, las certificaciones WMM (Wi-Fi Multi Media), básicamente asociadas al protocolo EDCA de la IEEE 802.11e. La siguiente tabla muestra los valores por defecto de la asociación de prioridades WMM (IEEE 802.11e) con los tags de IEEE 802.1d [56].

Access Category	Description	802.1d Tags
WMM Voice Priority	Highest priority Allows multiple concurrent VoIP calls, with low latency and toll voice quality	7, 6
WMM Video Priority	Prioritize video traffic above other data traffic One 802.11g or 802.11a channel can support 3-4 SDTV streams or 1 HDTV streams	5, 4
WMM Best Effort Priority	Traffic from legacy devices, or traffic from applications or devices that lack QoS capabilities Traffic less sensitive to latency, but affected by long delays, such as Internet surfing	0, 3
WMM Background Priority	Low priority traffic (file downloads, print jobs) that does not have strict latency and throughput requirements	2, 1

La técnica para priorizar el tráfico se basa esencialmente en reducir los tiempos de inicio de transmisión para las colas de alto tráfico. Estos tiempos de inicio de transmisión, en una red inalámbrica, están dadas por 2 factores:

- El tiempo mínimo entre frames (AIFSN = Arbitrary Inter-Frame Space Number)
- Un tiempo de “contención” (CW = Contention Window), aleatorio.

Ambos tiempos son menores para los tráficos de mayor tráfico, como se muestra en la Figura 7.1. Para cada cola, según su prioridad, el tiempo de espera para transmitir se calcula como el AIFSN más un número aleatorio, entro 0 y el máximo CW para esa prioridad. Si al intentar transmitir el canal está ocupado, o existe una colisión, este máximo CW se duplica, hasta un valor máximo, que también depende de la prioridad. Luego de una transmisión exitosa, el valor del CW se reinicia a su valor inicial. Este mecanismo hace que, estadísticamente, las tramas de alta prioridad tengan mayores oportunidades de ser enviados antes que las tramas de baja prioridad.

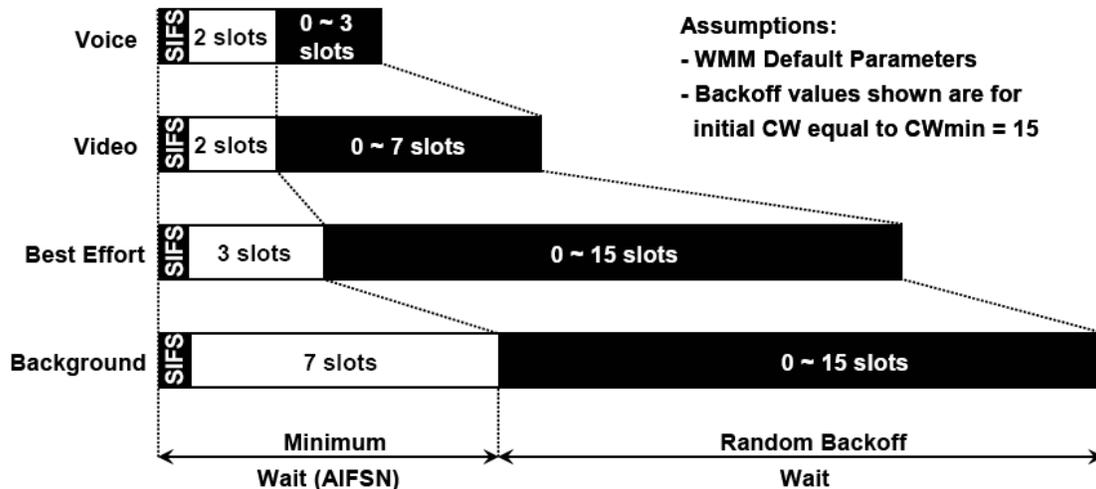


Figura 7.1

A los efectos de garantizar la calidad de servicio de punta a punta, diferentes protocolos deben ser “mapeados” entre las redes inalámbricas y las redes cableadas. Los AP deben mapear las categorías de tráfico EDCA (o WMM) en los protocolos de red de capa 2 (por ejemplo 802.1p) y/o de capa 3 (por ejemplo DiffServ)

#### 7.4 Capacidad

La capacidad (cantidad de usuarios móviles soportados), es un aspecto fundamental a considerar durante la etapa de diseño de un sistema inalámbrico que soporte voz. Esta capacidad está determinada por diversos factores. La cantidad máxima de llamadas en determinada área será función de la cantidad de usuarios esperados en dicha área, y de las reglas de tráfico habituales en telefonía (ver [1]). La capacidad de las redes WLAN está esencialmente determinada por la cantidad de canales de RF no solapados y la densidad de APs instalados. Agregando APs que utilicen canales de RF que no interfieran (que no se “solapen” en la frecuencia) se incrementa la capacidad en un área determinada. Si los canales se solapan, agregar más APs genera interferencia de RF, lo que termina disminuyendo la capacidad, debido a las colisiones y re-transmisiones que se generan.

802.11b y 802.11g son mayoritariamente soportados por los dispositivos inalámbricos. Operan en el rango de los 2.4 GHz y tienen únicamente 3 canales de RF que no se solapan, es decir, que no interfieren entre sí. 802.11a opera en el rango de los 5 GHz y dispone de 8 a 20 canales de RF que no se solapan (ver [2] por una descripción más detallada de las recomendaciones 802.11a, 802.11b y 802.11g). Es por esta razón que 802.11a soporta mayores densidades de AP, y puede proveer por tanto mayor capacidad de llamadas concurrentes en un área determinada.

## 8 Protocolos de señalización en VoIP

### 8.1 H.323

H.323 (*"AUDIOVISUAL AND MULTIMEDIA SYSTEMS: Infrastructure of audiovisual services – Systems and terminal equipment for audiovisual services"* [57]) es una recomendación de la ITU-T que describe los terminales y demás dispositivos que proveen servicios de comunicaciones multimedia (video, voz y datos) sobre redes de paquetes que no garantizan calidad de servicio (por ejemplo Ethernet con protocolos TCP/IP).

La primera versión de H.323 fue aprobada en 1996 por la ITU-T. La versión 2 fue aprobada en enero de 1998, la versión 3 en 1999, la versión 4 en 2000, la versión 5 en 2003 y la versión 6 en junio de 2006. H.323 es parte de las recomendaciones H.32x (como por ejemplo H.320 para ISDN y H.324 para la PSTN)

H.323 es aplicable a cualquier red conmutada de paquetes, con independencia de los protocolos utilizados en la "capa física". La red debe proveer protocolos de entrega "confiables" (como por ejemplo TCP - Transmission Control Protocol) y protocolos de entrega "no confiables" (como por ejemplo UDP - User Datagram Protocol). Los protocolos "confiables" proveen mecanismos de confirmación de recepción de paquetes, y retransmisiones, de ser necesarias, para asegurar la correcta recepción de los paquetes enviados. Los protocolos "no confiables" son del tipo "mejor esfuerzo" en la entrega de paquetes, pero no sobrecargan a la red con paquetes de confirmación y eventuales retransmisiones, lo que los hace a su vez más "rápidos".

#### 8.1.1 Arquitectura H.323

La recomendación H.323 define una arquitectura en la que se diferencian los siguientes componentes, como se esquematiza en la Figura 8.1:

- Terminales H.323
- Pasarelas ("Gateways") para interconectar el "mundo" H.323 con otros sistemas de telecomunicaciones (típicamente la red telefónica pública analógica y digital)
- Controlador H.323 ("Gatekeeper")
- Unidades de control multipunto ("MCU - Multipoint Control Units")

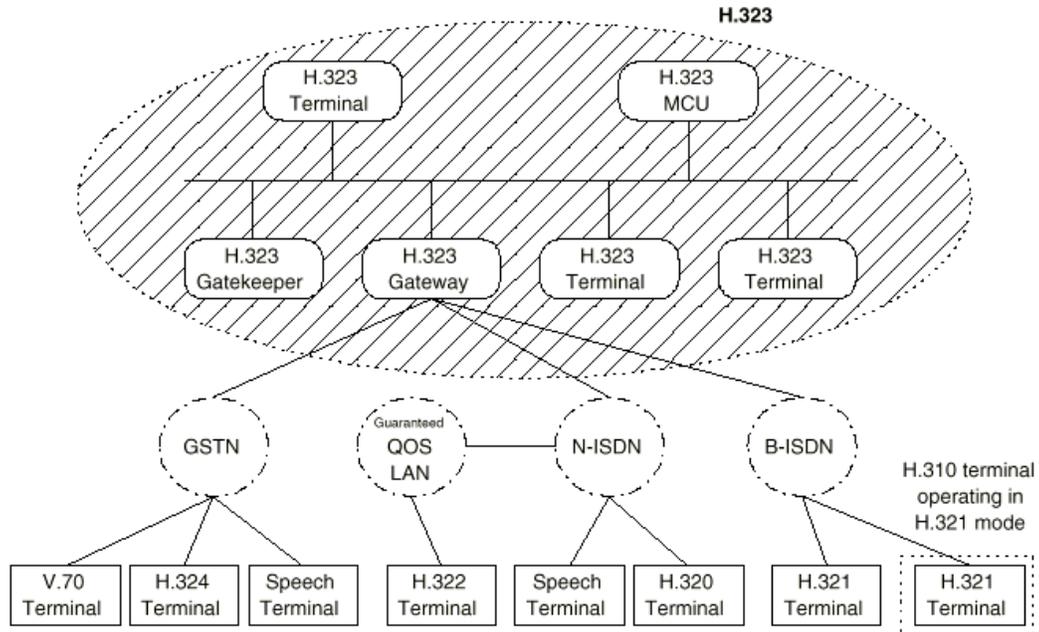


Figura 8.1

### 8.1.1.1 Terminales

Los terminales H.323 son los “teléfonos multimedia IP”. Estos “teléfonos” pueden ser aplicaciones informáticas, que utilizan las capacidades multimedia del PC (parlantes y micrófono), o terminales físicos de similar aspecto a cualquier teléfono o videoteléfono.

La recomendación establece que esos terminales deben soportar obligatoriamente comunicaciones de voz y opcionalmente comunicaciones de datos y video. La recomendación también establece los protocolos utilizados en la señalización de las llamadas, los mensajes de control, la manera de realizar la multiplexación de estos mensajes, los codecs de audio y video y los protocolos utilizados para el intercambio de datos entre los terminales. Un esquema de un terminal H.323 se muestra en la Figura 8.2.

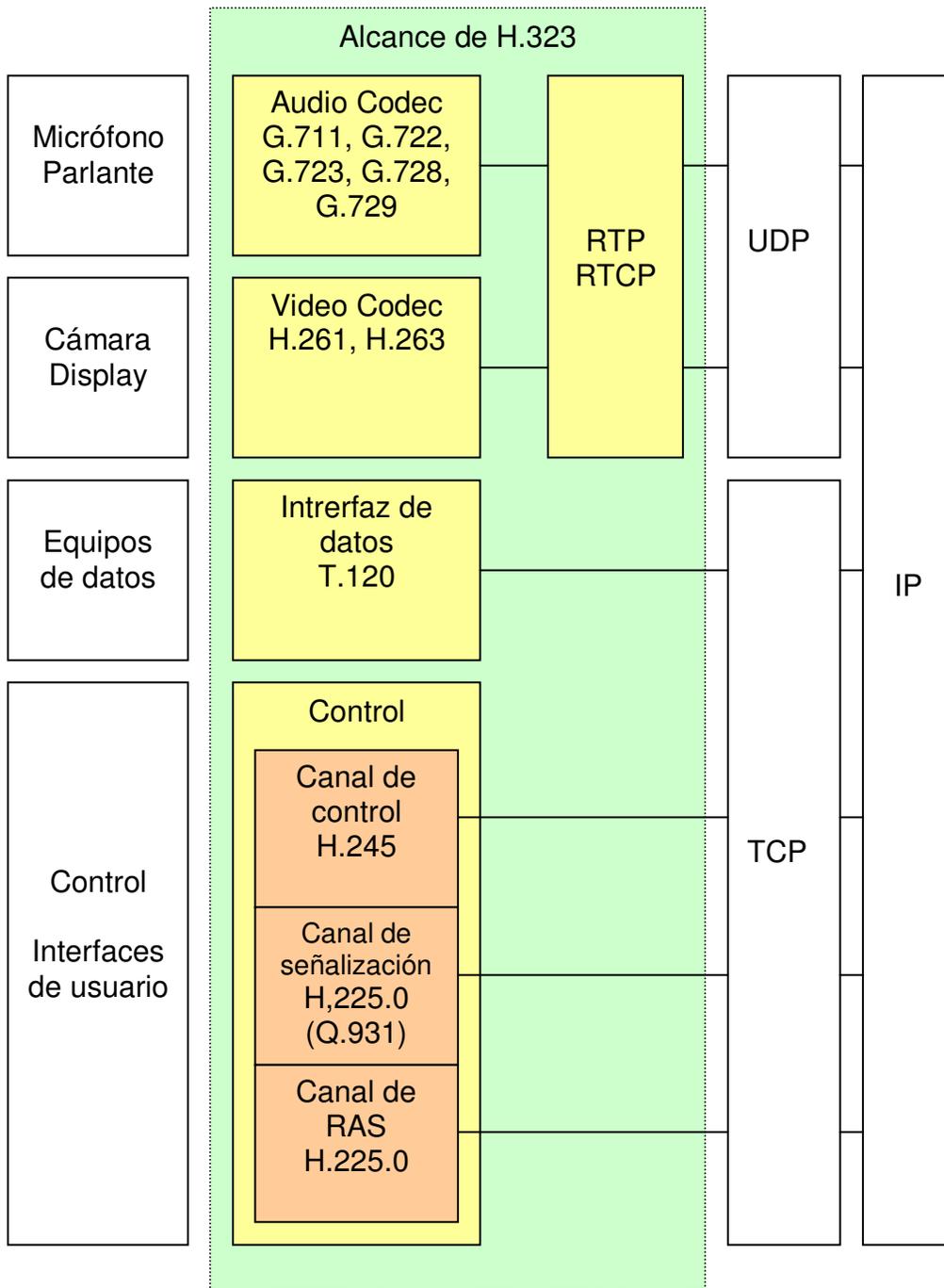


Figura 8.2

### Audio Codec

La recomendación H.323 admite los siguientes tipos de codificación de audio (ver capítulo 2.1):

- G.711 (64 kb/s)

- G.722 (7 kHz speed at 48, 56 and 64K bit/s (hi-fi voice)
- G.728 (16 kb/s)
- G.723.1 (Dual Rate Speed at 6.4 and 5.3K bit/s)
- G.729 (8 kb/s)

Todo terminal H.323 debe obligatoriamente disponer de un codec de audio. Este codec de audio debe soportar como mínimo la codificación G.711 (en “Ley A” y “Ley  $\mu$ ”), y opcionalmente las otras admitidas por la recomendación H.323. El tipo de codec a utilizar al establecer una comunicación de audio con otro terminal es negociado según la recomendación H.245, por el “canal de control de llamadas”. En una misma comunicación, el terminal H.323 debe ser capaz de utilizar un codec en la recepción y otro diferente en la transmisión.

Si el terminal soporta G.723.1, debe ser capaz de codificar a 5.3 kb/s y a 6.3 kb/s.

El audio generado es “formateado” según la recomendación H.225.0, utilizando los estándares RTP (Real Time Protocol) y RTCP (Real Time Control Protocol).

### **Video Codec**

La codificación de video es opcional en H.323, por lo que pueden existir terminales H.323 que no dispongan de facilidades de video. Sin embargo, si el terminal H.323 dispone de facilidades de comunicaciones de video, debe adecuarse a los siguientes codificadores:

- H.261 (n x 64 kb/s)
- H.263 (< 64 kb/s)

H.261 QCIF es obligatorio si el terminal dispone de facilidades de video. Otros modos de H.261 y H.263 son opcionales.

El tipo de codec a utilizar al establecer una comunicación de video con otro terminal, la velocidad de transmisión, el formato de la imagen y las opciones de los algoritmos utilizados, son negociados según la recomendación H.245, por el “canal de control de llamadas”.

Un mismo terminal puede soportar a la vez varios canales de video, tanto en la transmisión como en la recepción. Asimismo, en una misma comunicación, el terminal H.323 debe ser capaz de utilizar un codec en la recepción y otro diferente en la transmisión.

El video generado es “formateado” según la recomendación H.225.0, utilizando los estándares RTP (Real Time Protocol) y RTCP (Real Time Control Protocol).

## Interfaz de datos

Los terminales H.323 pueden establecer comunicaciones de datos con otros terminales H.323 (por ejemplo, para compartir documentos). Para ello, pueden abrir “canales de datos”, los que pueden ser bidireccionales o unidireccionales.

La recomendación T.120 provee un estándar de interoperabilidad para el intercambio de datos entre terminales H.323 y otro tipo de terminales (por ejemplo, terminales H.324, H.320 y H.310).

## RTP (Real-Time Transport Protocol)

El protocolo RTP, basado en el RFC 3550 (ver 2.3), establece los principios de un protocolo de transporte sobre redes que no garantizan calidad de servicio para datos “de tiempo real”, como por ejemplo voz y video.

El protocolo establece la manera de generar paquetes que incluyen, además de los propios datos de “tiempo real” a transmitir, números de secuencia, marcas de tiempo, y monitoreo de entrega. Las aplicaciones típicamente utilizan RTP sobre protocolos de red “no confiables”, como UDP. Los “bytes” obtenidos de cada conjunto de muestras de voz o video son encapsulados en paquetes RTP, y cada paquete RTP es a su vez encapsulado en segmentos UDP.

RTP soporta transferencia de datos a destinos múltiples, usando facilidades de “multicast”, si esto es provisto por la red.

## RTCP (RTP Control Protocol)

El RFC 3550 también detalla el protocolo de control RTCP. Los paquetes RTCP son enviados periódicamente y contienen indicadores de la calidad del enlace, y otros datos acerca de la fuente y destino de la comunicación. Estos indicadores incluyen la cantidad de paquetes enviados, la cantidad de paquetes recibidos perdidos y el “jitter” en el receptor. El RFC 3550 no establece que deben hacer las aplicaciones (terminales H.323 en este caso) con los indicadores recibidos, sino que esto queda librado a cada implementación. Los usos típicos están relacionados con el cambio de codecs, y el reporte de problemas, locales, remotos o globales.

## Canal de control H.245

Los terminales H.323 utilizan uno o varios “canales de control” para enviar y recibir mensajes desde y hacia otros terminales y dispositivos H.323 (gateways, gatekeepers, etc.). El protocolo utilizado en estos canales de control está determinado en la recomendación H.245 (“*Line transmisión of non-telephone signals – Control protocol for multimedia communication*” (ver [58])) de la ITU-T.

Los mensajes definidos en H.245 incluyen el intercambio de las capacidades de cada terminal, la apertura y cierre de canales lógicos, mensajes de control de flujo y comandos e indicadores generales.

Los terminales deben mantener un canal de control H.245 por cada llamada en la que el terminal esté participando. Dado que un terminal puede estar participando en forma simultánea en varias llamadas, puede tener también varios canales de control H.245 abiertos.

### **Canal RAS H.225.0 (Registration, Admission and Status)**

El canal de "RAS" es utilizado entre los terminales y el Gatekeeper (Ver 8.1.1.3). A través de éste canal, el terminal realiza las funciones de registro, admisión, solicitud de ancho de banda, status, etc. Este canal de señalización es independiente del canal de control H.245 y del canal de señalización de llamadas. En ambientes de red dónde no se dispone de Gatekeepers (recordar que los gatekeepers son opcionales en una red H.323), el canal de RAS no es utilizado por los terminales. En ambientes de red dónde se dispone de un Gatekeeper, el canal de RAS debe ser abierto entre el terminal y el gatekeeper, antes de ser abierto cualquier otro canal entre terminales.

El protocolo utilizado en el canal RAS está descrito en la recomendación de la ITU-T H.225.0 "*Call signalling protocols and media stream packetization for packet-based multimedia communication systems*" [59]

### **Canal de señalización H.225.0**

El canal de señalización del terminal H.323 utiliza funciones de señalización del protocolo H.225.0 para establecer conexiones con otro terminal H.323. Este canal de señalización es independiente del canal de RAS y del canal de control H.245. El canal de señalización es abierto por el terminal antes de establecer el canal de control H.245.

En ambientes de red en los que no hay Gatekeeper, el canal de señalización es establecido directamente entre dos terminales. En ambientes de red en los que se dispone de un Gatekeeper, el canal de señalización es abierto entre el propio terminal y el Gatekeeper, o entre el propio terminal y otro terminal, de acuerdo a lo indicado por el Gatekeeper.

La recomendación H.225.0 utiliza los mensajes descritos en la recomendación Q.931 [60], utilizada en ISDN

### 8.1.1.2 Gateways (Pasarelas)

Los gateways o pasarelas, realizan la función de interconexión entre las redes H.323 y otras redes de comunicaciones, como la red pública conmutada (PSTN – Public Switched Telephony Network) analógica y digital, o redes “SIP”. El esquema de funcionamiento se muestra en la Figura 8.3.

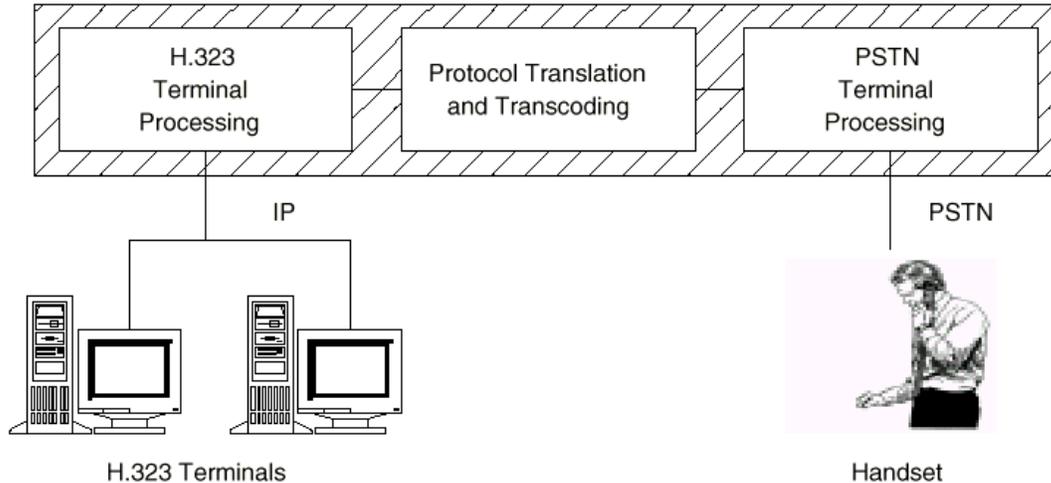


Figura 8.3

Los gateways son responsables de adaptar el audio, video y los datos, así como también la señalización, entre los formatos propios de H.323 y otras redes de telecomunicación, de manera transparente para los usuarios.

Los terminales H.323 pueden comunicarse con otros terminales H.323 de la misma red en forma directa, sin utilizar gateways. En redes dónde no es necesario tener comunicación con terminales externos a la propia red, no es necesario disponer de gateways. Por ello, son elementos opcionales en la recomendación H.323

Hacia la red H.323, el gateway presenta las características de un terminal H.323 (o de un MCU – Multipoint Control Unit), y hacia la red PSTN, el de un terminal telefónico (de acuerdo al tipo de red a la que esté conectado, podrá presentar las características de un teléfono analógico, ISDN, etc.)

Los gatekeepers conocen la existencia de gateways, ya que esto es especificado en el momento en que el gateway se registra en el gatekeeper.

La recomendación no detalla la manera en que deben implementarse los gateways. Pueden ser parte del mismo equipo donde reside el gatekeeper, ser equipos independientes, etc.

### 8.1.1.3 Gatekeeper

Las redes H.323 pueden disponer de un elemento centralizador de control y servicios telefónicos, llamado en la recomendación “Gatekeeper”. Este dispositivo, en caso de existir, debe proveer, como mínimo, los siguientes servicios:

#### **Traducción de direcciones**

Una de las funciones principales del Gatekeeper es traducir un número telefónico, o un “alias” a la dirección de red apropiada (por ejemplo, la dirección IP). Para ello, el Gatekeeper debe disponer de una tabla de traducción de direcciones, que se actualiza cada vez que un dispositivo (por ejemplo un terminal H.323) se registra o de-registra en el Gatekeeper.

#### **Control de Admisión**

El Gatekeeper puede autorizar o negar el acceso (registro) a la red H.323, utilizando mensajes descritos en la recomendación H.225.0. Las reglas de decisión para autorizar o negar el acceso no son parte de la recomendación.

#### **Control de Ancho de Banda**

El Gatekeeper debe soportar la mensajería H.225.0 respecto a la asignación de ancho de banda. Mediante los protocolos adecuados, puede indicar a cada terminal el ancho de banda total disponible según el tipo de llamada, las categorías de los terminales, etc.

#### **Gerenciamiento de su “Zona”**

Un Gatekeeper define una “Zona H.323”. Los terminales, gateways y MCUs registrados en el mismo Gatekeeper pertenecen a la misma “zona”. El Gatekeeper debe brindar como mínimo los servicios descritos anteriormente para todos los dispositivos de su “Zona”.

En forma adicional a los servicios indicados, el Gatekeeper puede brindar cualquier otro tipo de servicios adicionales, como por ejemplo:

#### **Señalización para el control de llamadas**

Cuando una red H.323 dispone de un Gatekeeper, la señalización para el establecimiento y liberación de llamadas puede realizarse directamente entre dos terminales, o a través del Gatekeeper. El Gatekeeper puede asumir la función de centralizador de señalización, de manera que los terminales tengan que utilizarlo para las funciones de señalización de llamadas.

#### **Autorización de llamadas**

Mediante el uso de señalización H.225.0, el Gatekeeper puede autorizar o negar llamadas solicitadas desde los terminales. Las razones para autorizar o negar llamadas pueden incluir criterios de restricciones de ciertos terminales, horarios

del día, etc. La recomendación no establece cuales deben ser estos criterios, los que quedan librados a los fabricantes.

Una red puede tener más de un Gatekeeper, los que pueden comunicarse entre sí. Los protocolos de comunicación utilizados entre dos o más Gatekeeper no están especificados en la recomendación.

#### 8.1.1.4 MCU – Multipoint Control Unit (Unidad de control multipunto)

El “MCU” provee soporte para realizar conferencias, entre 3 o más terminales. Se compone de unidades “Controladoras Multipunto” (MC – Multipoint Controller) y “Procesadores Multipunto” (MP – Multipoint Processor)

##### **Controladoras Multipunto (MC – Multipoint Controller)**

Los controladores multipunto (MC) proveen las funciones de control necesarias para la implementación de conferencias de 3 o más terminales.

Los MC realizan el intercambio de capacidades entre los terminales de la conferencia, de manera de establecer un modo común a todos los participantes.

Como parte del establecimiento de una conferencia, los terminales se conectan a un MC utilizando el canal de control H.245.

##### **Procesadores Multipunto (MP – Multipoint Processor)**

A diferencia de los MC, que se encarga exclusivamente del control de las conferencias, los MP reciben los canales de audio, video y/o datos de los terminales, los procesan, y los redistribuyen nuevamente a los terminales.

Los MP son los encargados de realizar la conmutación o mezcla del audio y video proveniente de los terminales, y redistribuirlo hacia los terminales.

El MP puede decidir “conmutar” el audio o video, de manera de enviar hacia todos los terminales de la conferencia el audio o video proveniente de uno de los terminales, o “mezclar” el audio y video, de manera de enviar hacia todos los terminales la suma del audio y video proveniente de todos los otros. En el caso de video, la mezcla puede consistir en enviar hacia un mismo terminal varios cuadrantes, cada uno con el video proveniente desde los otros terminales.

El criterio utilizado para realizar las mezclas o la conmutación es determinado por el MC.

El MCU mínimo puede consistir de un único MC y ningún MP. En este caso, las funciones de mezcla y conmutación de audio y video deberán ser provistas por los propios terminales. Un MCU típico, que soporte conferencias centralizadas, se compone de un MC y un MP. Un MCU típico que soporte únicamente conferencias descentralizadas, se compone de un MC y un MP con soporte únicamente la recomendación T.120 (datos)

### 8.1.2 Ejemplos de señalización H.323

En la Figura 8.4 se muestra un esquema del inicio de llamada básica (“Direct Call”) H.323, entre dos terminales, con la existencia de un “Gatekeeper” en la red. El Terminal 1 solicita autorización al Gatekeeper, y luego envía un “Setup” directamente al Terminal 2. El Terminal 2 solicita a su vez autorización al Gatekeeper, hace timbrar su dispositivo, y finalmente envía la confirmación de la conexión al Terminal 1.

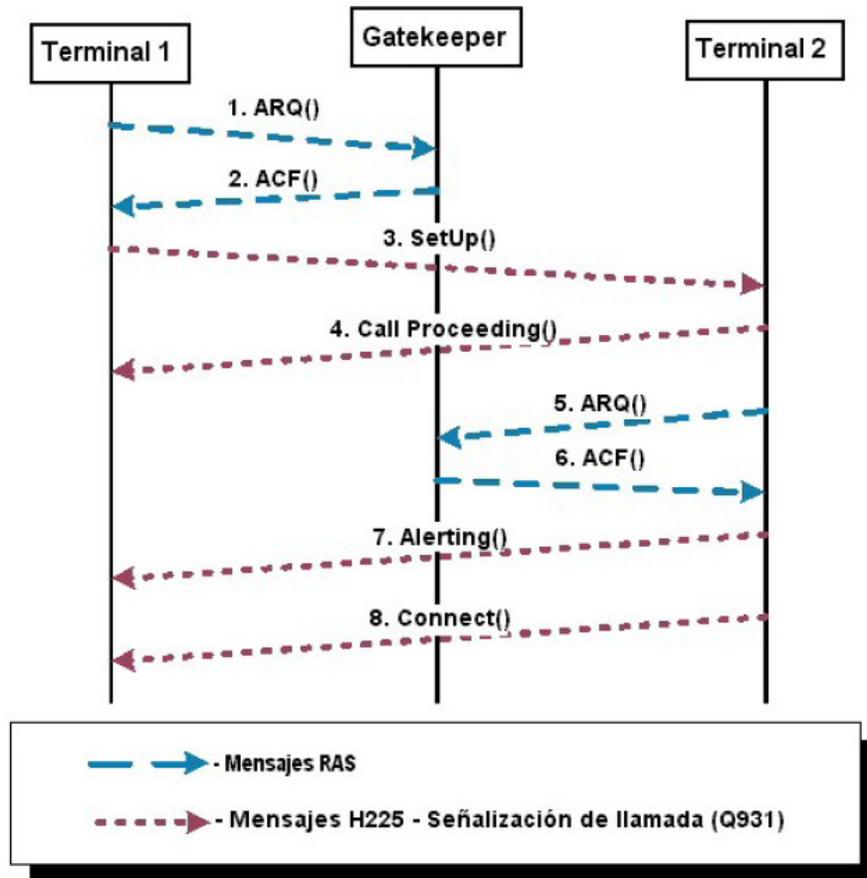


Figura 8.4

Una vez enviada la señal de “Connect”, los terminales intercambian información acerca de sus capacidades y abren los canales lógicos para el establecimiento de los flujos del medio (audio / video), como se muestra en la Figura 8.5

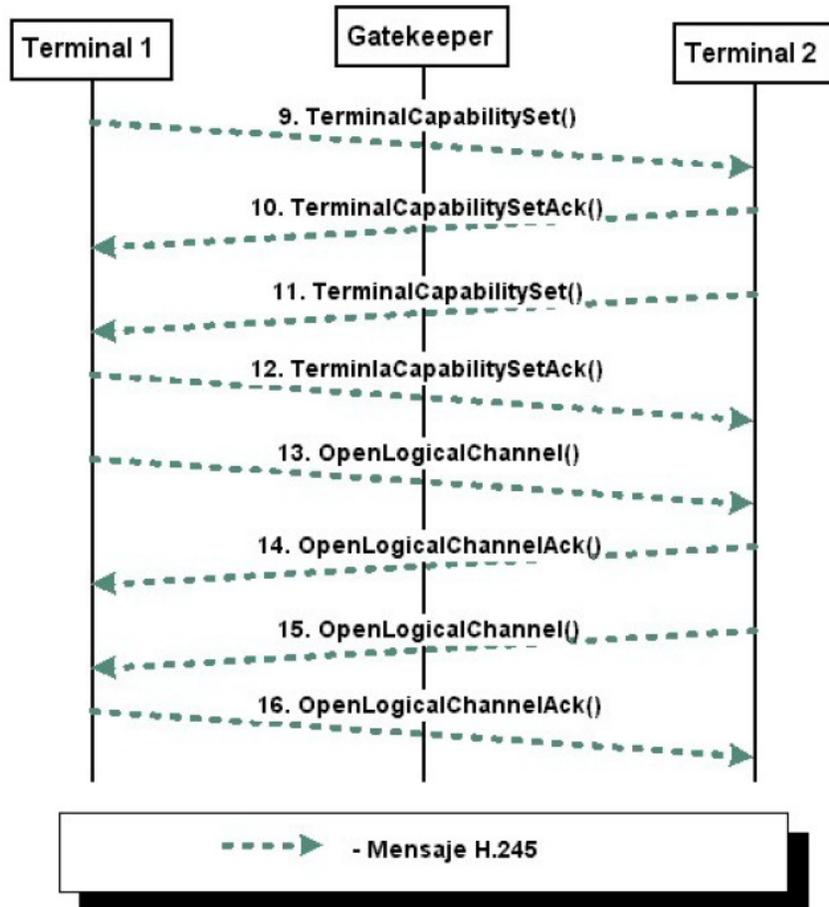


Figura 8.5

Una vez intercambiada la información capacidades (codecs soportados, etc.) y de establecer los puertos RTP (canales lógicos), se establecen los flujos del medio (audio / video), a través e mensajes RTP, como se puede ver en la Figura 8.6. Finalmente, la llamada es liberada, y los terminales se desregistran del Gatekeeper, como se muestra en la Figura 8.7.

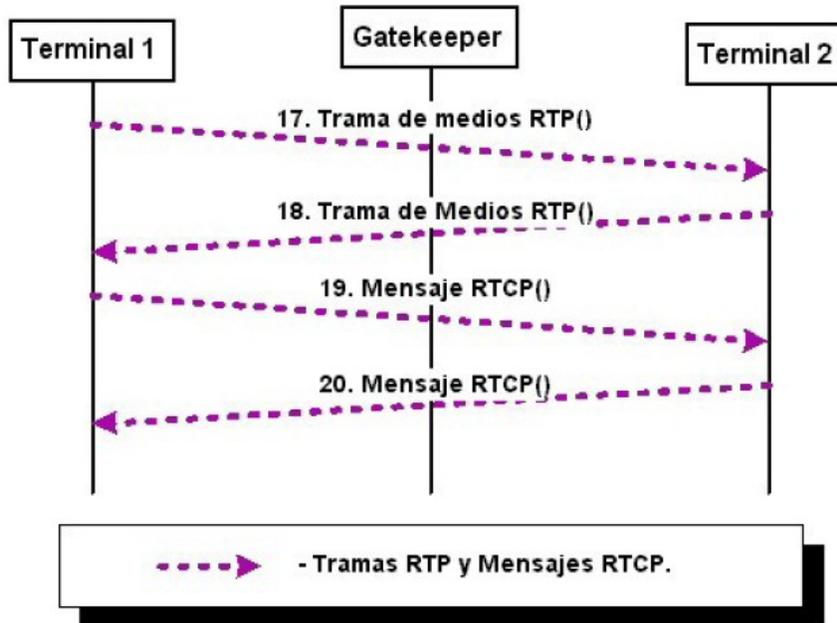


Figura 8.6

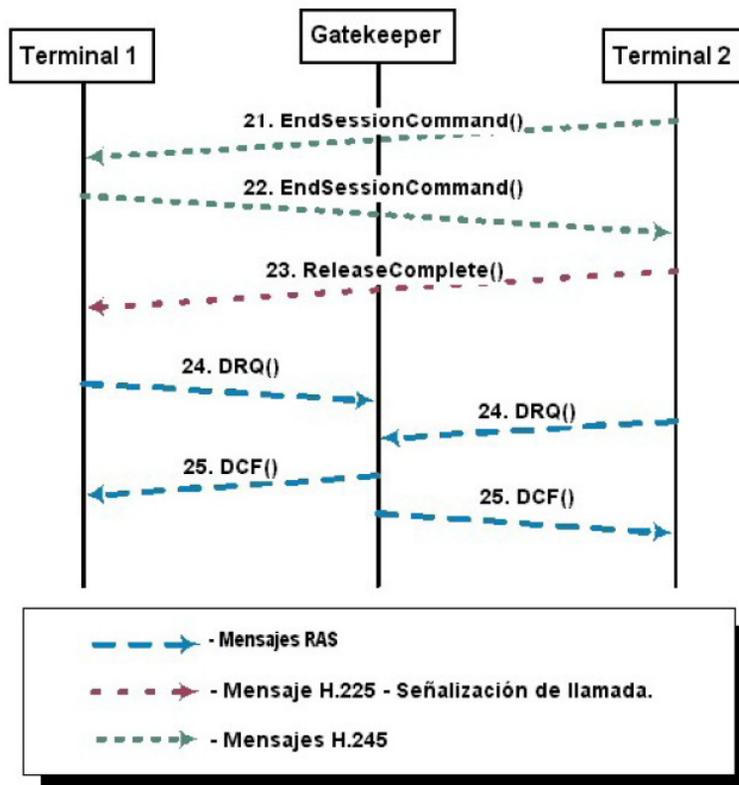


Figura 8.7

En algunas cosas, el Gatekeeper puede indicar que él mismo debe interponerse en la señalización telefónica, a los efectos de control, tarificación, o gestión. Este tipo de llamadas, conocidas como “Gatekeeper Routed”, se ilustra en la Figura 8.8. En este caso, el Gatekeeper queda en el medio de la señalización H225 (Q.931).

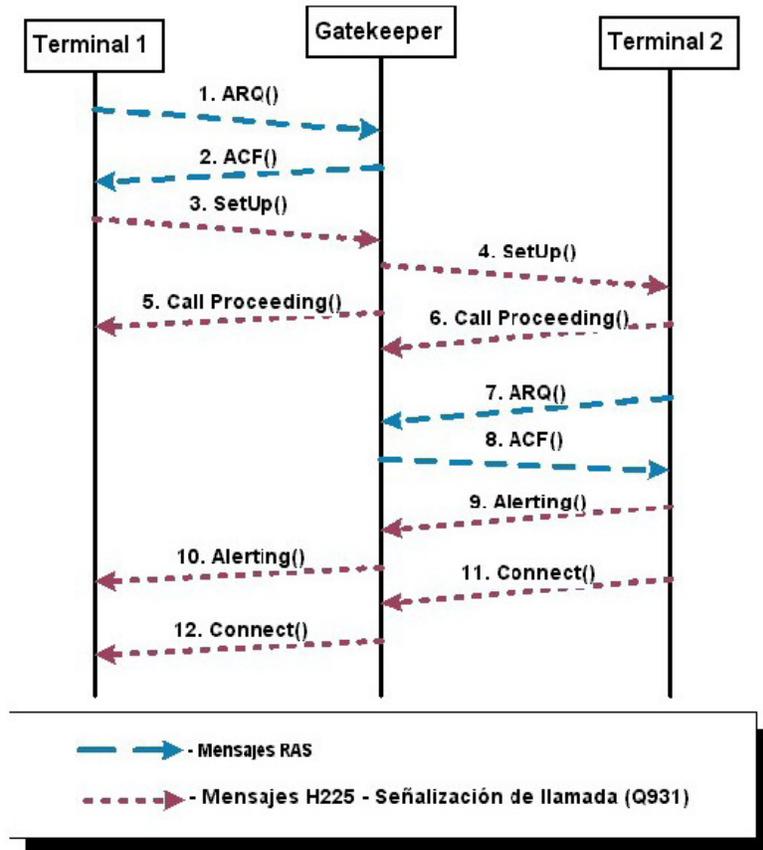


Figura 8.8

El intercambio de capacidades puede continuar siendo de Terminal a Terminal, como se muestra en la Figura 8.9, al igual que el flujo de medios (audio / video) RTP. Finalmente, para la desconexión de llamada, vuelve a intervenir el Gatekeeper, como se ve en la Figura 8.10.

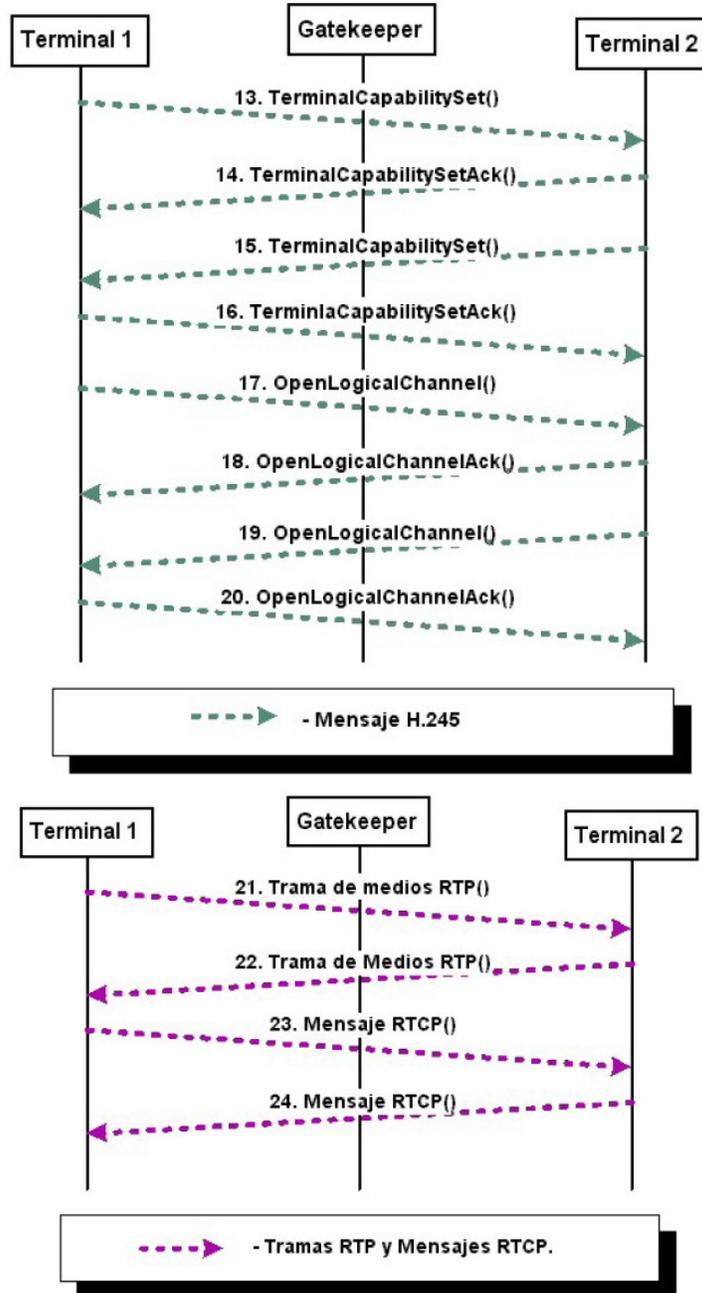


Figura 8.9

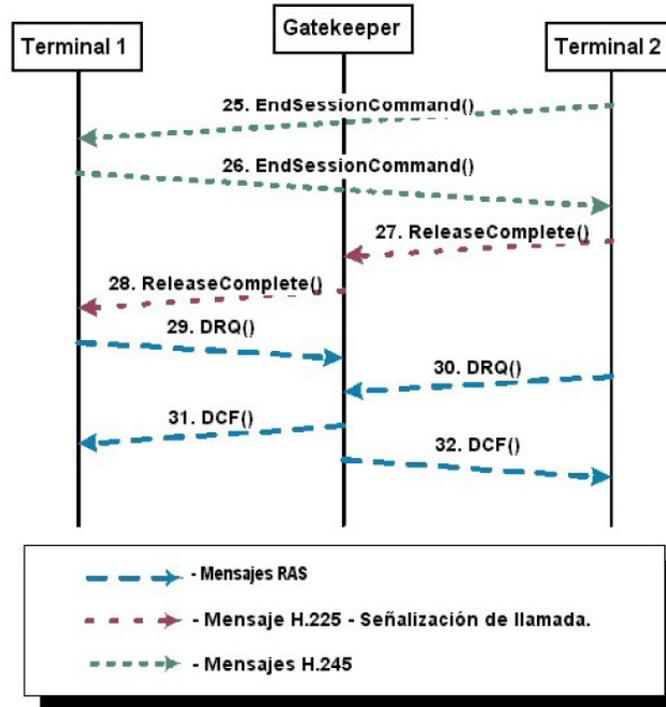


Figura 8.10

En la Figura 8.11 se puede ver una captura de una llamada real H.323 entre dos dispositivos, en este caso, sin la presencia de un Gatekeeper. En este ejemplo se muestra la señalización hasta el momento del establecimiento de la llamada (no se muestra la desconexión luego del audio).

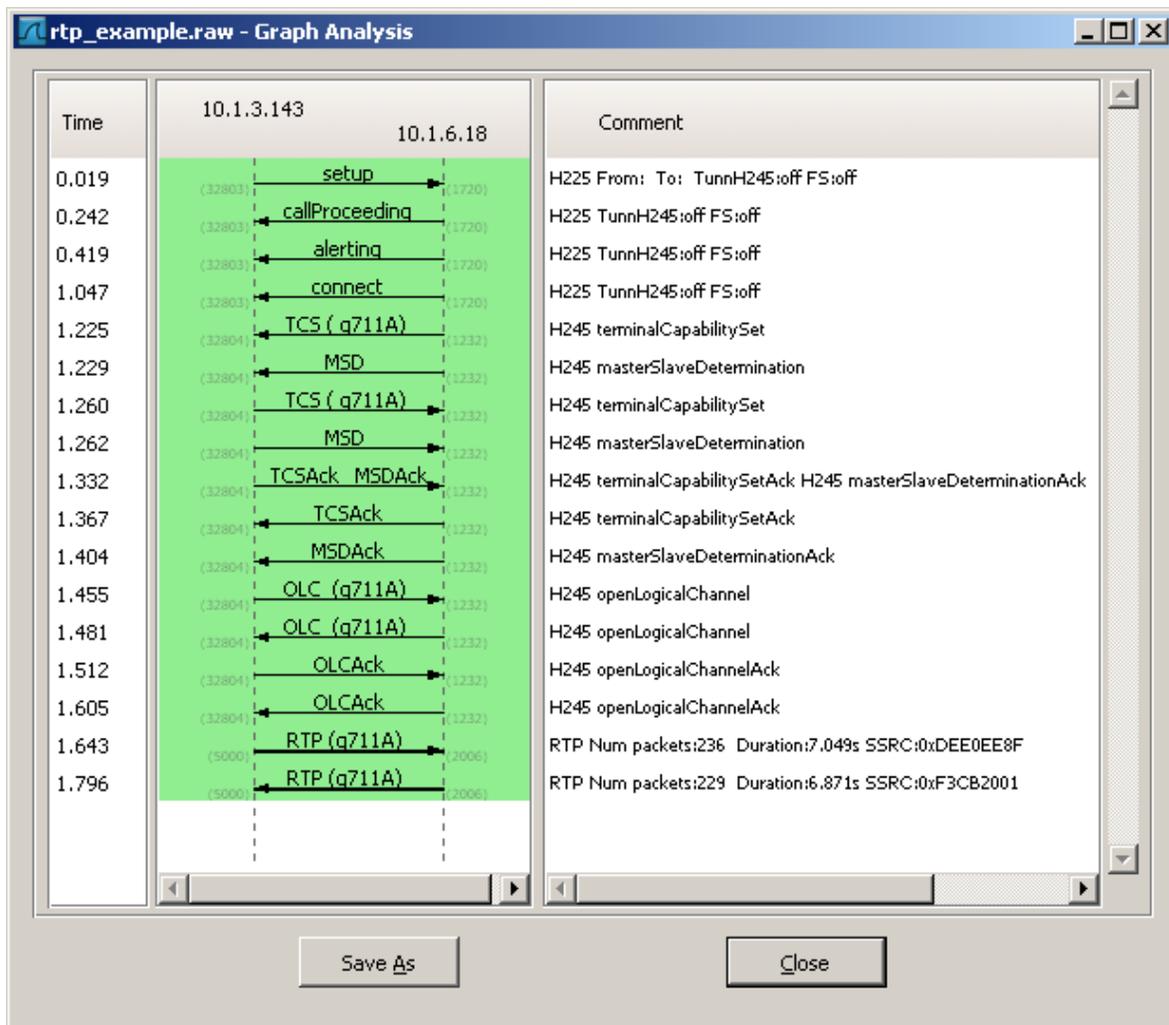


Figura 8.11

## 8.2 SIP

En marzo de 1999 es aprobado el RFC 2543, por el grupo de estudio MMUSIC del IETF, dando origen oficial al protocolo SIP (Session Initiation Protocol). SIP tiene sus orígenes a fines de 1996, como un componente del “Mbone” (Multicast Backbone), El Mbone, era una red experimental montada sobre la Internet, para la distribución de contenido multimedia, incluyendo charlas, seminarios y conferencias de la IETF. Uno de sus componentes esenciales era un mecanismo para invitar a usuarios a escuchar una sesión multimedia, futura o ya establecida. Básicamente un “protocolo de inicio de sesión” (SIP).

En junio de 2002, el RFC 2543 fue reemplazado por un conjunto de nuevas recomendaciones, entre las que se encuentran los RFC 3261 al 3266 [61][62][63][64][65][66].

Una completa introducción a SIP puede leerse en [67].

### 8.2.1 Mensajería SIP

La mensajería SIP está basada en el esquema “Request” – “Response” de http. Esto presenta ciertas ventajas, sobre todo para los familiarizados con las tecnologías HTTP. A diferencia de H.323, todos los mensajes son de texto plano, y por lo tanto fáciles de interpretar (recordar que en H.323, los mensajes eran binarios).

Para iniciar una sesión se envía un mensaje de “Request” a una contraparte de destino. El destino recibe el “Request”, y lo contesta con el correspondiente “Response”. Un ejemplo del establecimiento de una comunicación se muestra en la Figura 8.12.

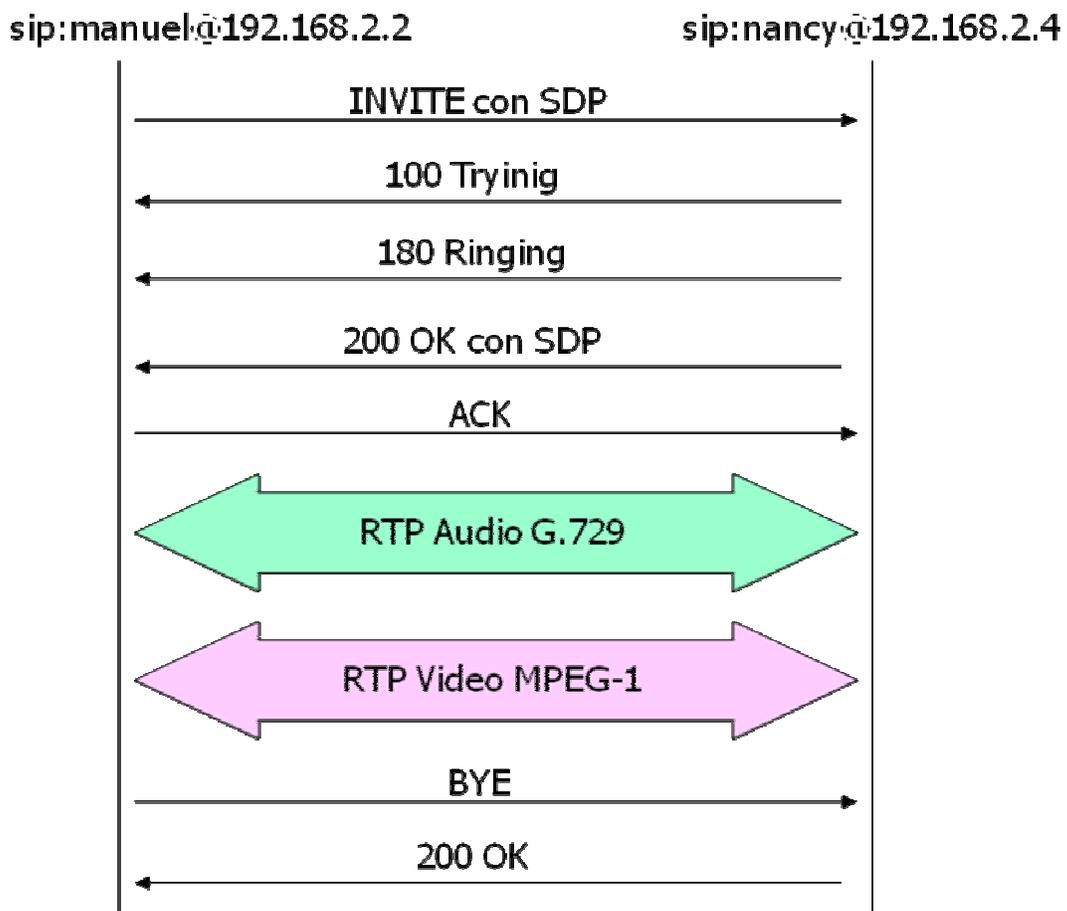


Figura 8.12

Los mensajes de “Request” tiene el formato

<Método> <URL> <SIP-Version>

Por ejemplo: INVITE sip:pepe@fing.com SIP/2.0

La siguiente tabla resume los métodos SIP

Método	Descripción
INVITE	A session is being requested to be setup using a specified media
ACK	Message from client to indicate that a successful response to an INVITE has been received
OPTIONS	A Query to a server about its capabilities
BYE	A call is being released by either party
CANCEL	Cancels any pending requests. Usually sent to a Proxy Server to cancel searches
REGISTER	Used by client to register a particular address with the SIP server
SUBSCRIBE	Used to request status or presence updates from the presence server
NOTIFY	Used to deliver information to the requestor or presence "watcher."
REFER	Used to referring the remote user agent to a web page or another URI
MESSAGE	Used to transport instant messages (IM) using SIP
UPDATE	Used to modify the state of a session without changing the state of the dialog
INFO	Used by a user agent to send call signaling information to another user agent with which it has an established media session
PRACK	Provisional ACK. Used to acknowledge receipt of reliably transported provisional responses (1xx)

Las respuestas ("Response") tienen el formato

<SIP-Version> < Status-Code> <Reason>

Por ejemplo: SIP/2.0 404 Not Found

La siguiente tabla resume las respuestas SIP

Respuesta	Descripción	Ejemplo
1xx	Informational – Request received, continuing to process request.	180 Ringing 181 Call is Being Forwarded 183 Session Progress
2xx	Success – Action was successfully received, understood and accepted.	200 OK
3xx	Redirection – Further action needs to be taken in order to complete the request.	300 Multiple Choices 302 Moved Temporarily
4xx	Client Error – Request contains bad syntax	401 Unauthorized

	or cannot be fulfilled at this server.	408 Request Timeout
5xx	Server Error – Server failed to fulfill an apparently valid request.	503 Service Unavailable 505 Version Not Suported
6xx	Global Failure – Request is invalid at any server.	600 Busy Everywhere 603 Decline

Para comprender mejor el formato de los mensajes SIP, se analizará con más detalle el método “INVITE”.

El método “INVITE” contiene varios campos que forman un “cabezal”. Este cabezal incluye atributos que proporcionan información adicional acerca del mensaje. En el “INVITE” se incluye un identificador único de la llamada, la dirección del destino, la dirección del origen, e información acerca del tipo de sesión que se quiere establecer.

Un ejemplo del cabezal de “INVITE” es el siguiente:

```
INVITE sip:pepe@fing.com SIP/2.0
Via: SIP/2.0/UDP pc33.montevideo.com;branch=z9hG4bK776asdhds
Max-Forwards: 70
To: Pepe <sip:pepe@fing.com>
From: Alicia <sip:alicia@abc.com>;tag=1928301774
Call-ID: a84b4c76e66710@pc33.montevideo.com
CSeq: 314159 INVITE
Contact: <sip:alicia@pc33.montevideo.com>
Content-Type: application/sdp
Content-Length: 142
```

La primer línea del mensaje contiene el nombre del método (“INVITE” en el ejemplo anterior). Las líneas siguientes contienen el resto de los campos del cabezal.

Los detalles de la sesión, como por ejemplo el tipo de medio, el codec o la frecuencia de muestreo, no son descriptos en el cabezal SIP. El cuerpo del mensaje SIP contiene una descripción de estos parámetros, codificados en un formato conocido como SDP (Session Description Protocol). Este mensaje SDP (no mostrado en el ejemplo anterior) es transportado en el mensaje SIP de manera similar a como se transporta un documento adjunto en un mensaje de e-mail. El formato SDP se estandariza en el RFC 2327 [68]

Un ejemplo del mensaje anterior, incluyendo el cuerpo SDP, se muestra a continuación:

```
INVITE sip:pepe@fing.com SIP/2.0
Via: SIP/2.0/UDP pc33.montevideo.com;branch=z9hG4bK776asdhds
Max-Forwards: 70
To: Pepe <sip:pepe@fing.com>
From: Alicia <sip:alicia@abc.com>;tag=1928301774
```

```
Call-ID: a84b4c76e66710@pc33.montevideo.com
CSeq: 314159 INVITE
Contact: <sip:alicia@pc33.montevideo.com>
Content-Type: application/sdp
Content-Length: 142
```

```
v=0
o=AGarcia 2890844526 2890842807 IN IP4 126.16.64.4
s=Phone Call
c=IN IP4 100.101.102.103
t=0 0
m=audio 49170 RTP/AVP 0
a=rtpmap:0 PCMU/8000
```

El formato de cada renglón de SDP es `<tipo>=<valor>`. `<tipo>` es siempre un único carácter, y se diferencian mayúsculas de minúsculas. El formato de `<valor>` depende del `<tipo>` al que corresponda. No puede haber espacios al lado del "=", ni antes ni después.

En este ejemplo, SDP contiene:

- Versión del protocolo (v)
- Origen (o)  
o=<username> <session id> <version> <network type> <address type>  
  <address>
- Nombre de la sesión (s)  
Debe haber un único campo de nombre de sesión.
- Datos de la conexión (c)  
c=<network type> <address type> <connection address>  
  <network type> = IN para Internet  
  <address type> = IP4 para IP version 4  
  <connection address> Es la dirección IP, que puede ser una dirección de multicast. Puede incluirse un valor de TTL (Time To Live), luego de la dirección IP, separada con el símbolo "/", por ejemplo:  
    c=IN IP4 224.2.1.1/127
- Temporalizadores (t)  
t=<start time> <stop time>  
Indica las horas de comienzo y fin de una sesión.
- Medios (m)  
m=<media> <port> <transport> <fmt list>  
  <media> puede ser "audio", "video", "application", "data" and "control"  
  <port> Puerto para el stream de medios

`<transport>` Para IP4, típicamente es RTP/AVP, indicando que se utiliza el transporte RTP

`<fmt list>` Es el formato del audio o video transportado (Ver Tipo de información (PT - Payload Type) )

- Atributos (a)

`a=<attribute>:<value>`

`a=<attribute>`

El campo “a” permite definir atributos, tanto a nivel de la sesión como a nivel de cada uno de los medios. Puede haber varios campos de atributos. Éstos pueden ser propiedades, del tipo `a=<flag>`, o atributos, del tipo `a=<attribute>:<value>`

## 8.2.2 Arquitectura SIP

SIP utiliza una arquitectura del tipo “Cliente-Servidor”, y tiene los siguientes componentes:

- Terminales SIP (SIP User Agents)
- Servidores SIP (Registrar, Proxy, Redirect, Location, Presence)
- Pasarelas SIP (Gateways)

### 8.2.2.1 Terminales SIP

Los terminales SIP, al igual que los H.323 son “teléfonos multimedia IP”. Estos “teléfonos” pueden ser aplicaciones informáticas, que utilizan las capacidades multimedia del PC (parlantes y micrófono), o terminales físicos de similar aspecto a cualquier teléfono o videoteléfono.

Los terminales SIP, llamados “SIP User Agents”, pueden iniciar y recibir “sesiones” SIP. Cada terminal dispone de un “User Agent Client” (UAC) y un “User Agent Server”. Los UAC son los encargados de iniciar requerimientos SIP hacia otros terminales. Los UAS son quienes escuchan y atienden los requerimientos remotos.

Así como los terminales telefónicos clásicos se identifican mediante su número de teléfono, o número de abonado, y los terminales H.323 mediante su “alias”, los terminales SIP se identifican a través de su “dirección SIP”. El direccionamiento en SIP utiliza el formato de URLs de Internet: `sip:nombre@dominio`

Los terminales SIP pueden soportar servicios de presencia, incorporando agentes de presencia (PA= Presence Agent). En estos casos son capaces de recibir solicitudes de suscripciones y generar notificaciones de cambios de estados. Un

PA soporta un paquete de eventos de presencia, el que incluye los métodos SUBSCRIBE y NOTIFY.

### 8.2.2.2 Servidores SIP

#### Registrar Server

Es un servidor de registro de usuarios SIP. Los usuarios (agentes SIP) solicitan su registro en este servidor, mediante el intercambio de mensajes SIP. Un servidor de registro acepta solamente el método REGISTER, rechazando cualquier otro método con una respuesta 501 (Not Implemented). La información de los usuarios registrados es puesta a disposición de otros servidores, como los Proxies o Redirect.

#### Proxy Server

Es un servidor que atiende las solicitudes y las redirige. Para ubicar el destino, puede consultar un servidor de ubicaciones (Location Server). La Figura 8.13 esquematiza el funcionamiento de SIP Proxy.

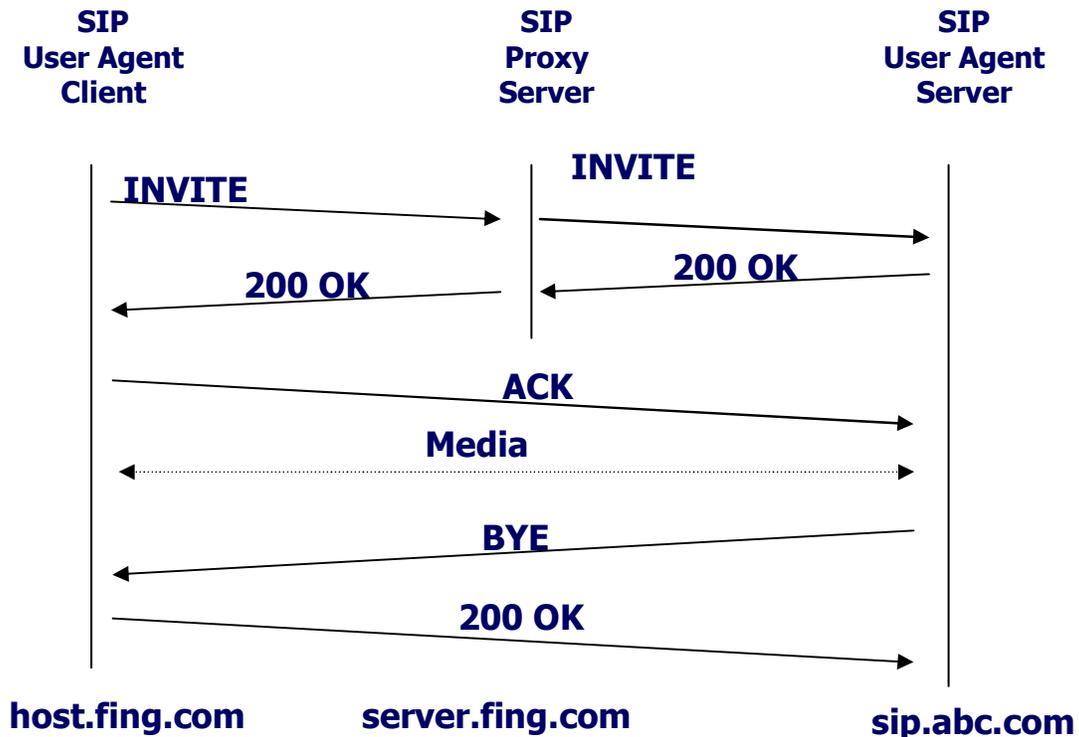


Figura 8.13

Los servidores Proxy tienen las siguientes características:

1. Un servidor Proxy no origina requerimientos (request), únicamente responde a requerimientos provenientes de agentes
2. No tiene capacidad de medios (audio, video, etc.)
3. No cambia ni interpreta los cuerpos de los mensajes. Se basa exclusivamente en los campos del cabezal del mensaje.

### Redirect Server

Es un servidor de redireccionamiento. A diferencia del "Proxy", no interviene en el establecimiento de la comunicación, sino que informa la manera de ubicar al destino final. La Figura 8.14 esquematiza el funcionamiento de SIP Redirect.

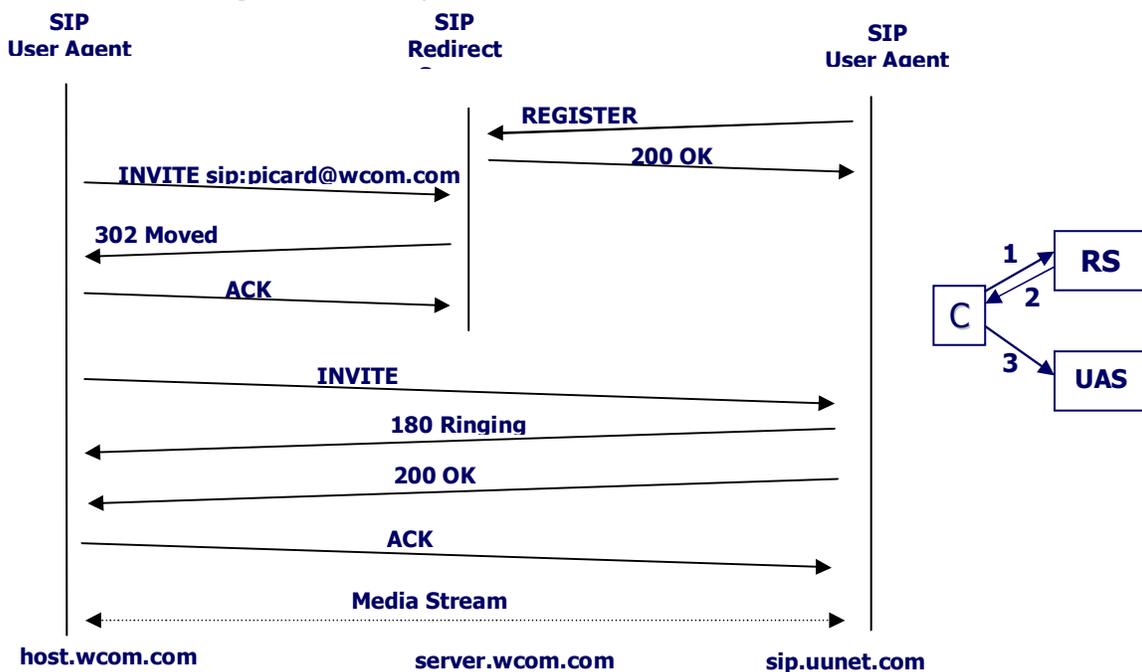


Figura 8.14

### Location Server

Es un servidor de búsqueda. Puede ser consultado para obtener la dirección final de un usuario SIP.

### Presence Server

Un servidor de presencia es un equipo que en ciertas ocasiones actúa como un agente de presencia y envía información de presencia a otros agentes, y en otras ocasiones actúa como proxy, redirigiendo las solicitudes de suscripciones a otros agentes de presencia.

Un uso típico de servidores Proxy, Redirect, Registrar y Location se ve en la Figura 8.15 [67]:

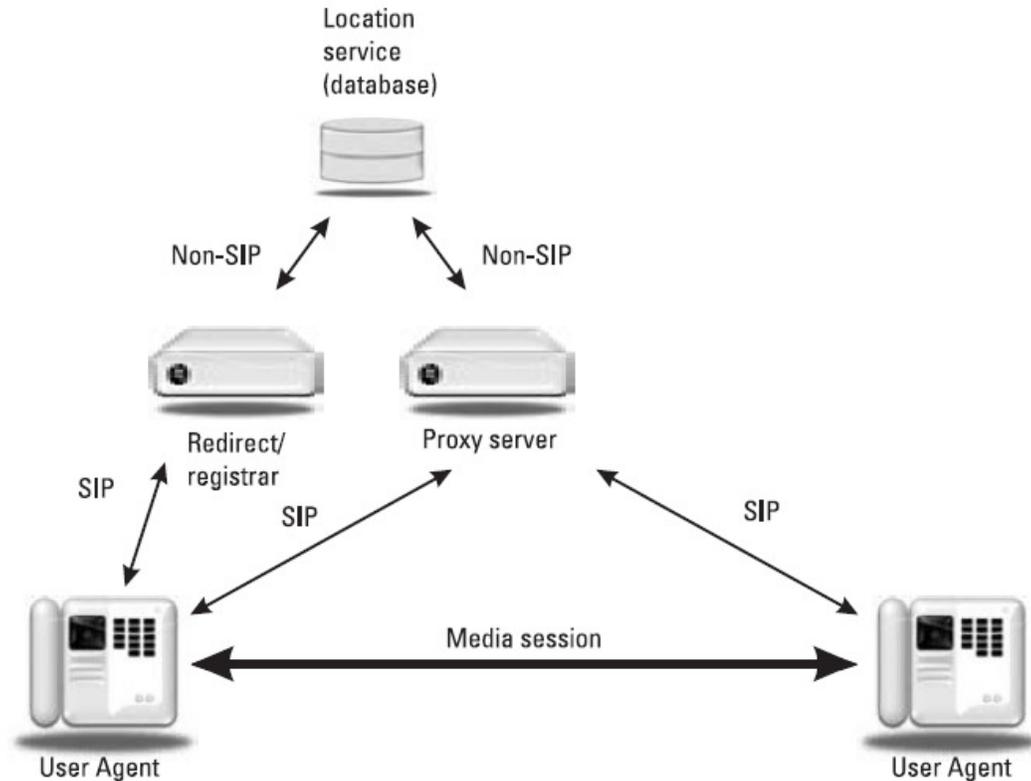


Figura 8.15

### 8.2.2.3 Gateway SIP

Al igual que en H.323, existen pasarelas SIP hacia la PSTN y también hacia H.323. Los gateways son responsables de adaptar el audio, video y los datos, así como también la señalización, entre los formatos propios de SIP y otras redes de telecomunicación, de manera transparente para los usuarios.

En redes donde no es necesario tener comunicación con terminales externos a la propia red, no es necesario disponer de gateways.

### 8.2.3 Ejemplos de señalización SIP

La figura Figura 8.16 muestra la señalización en una llamada SIP.

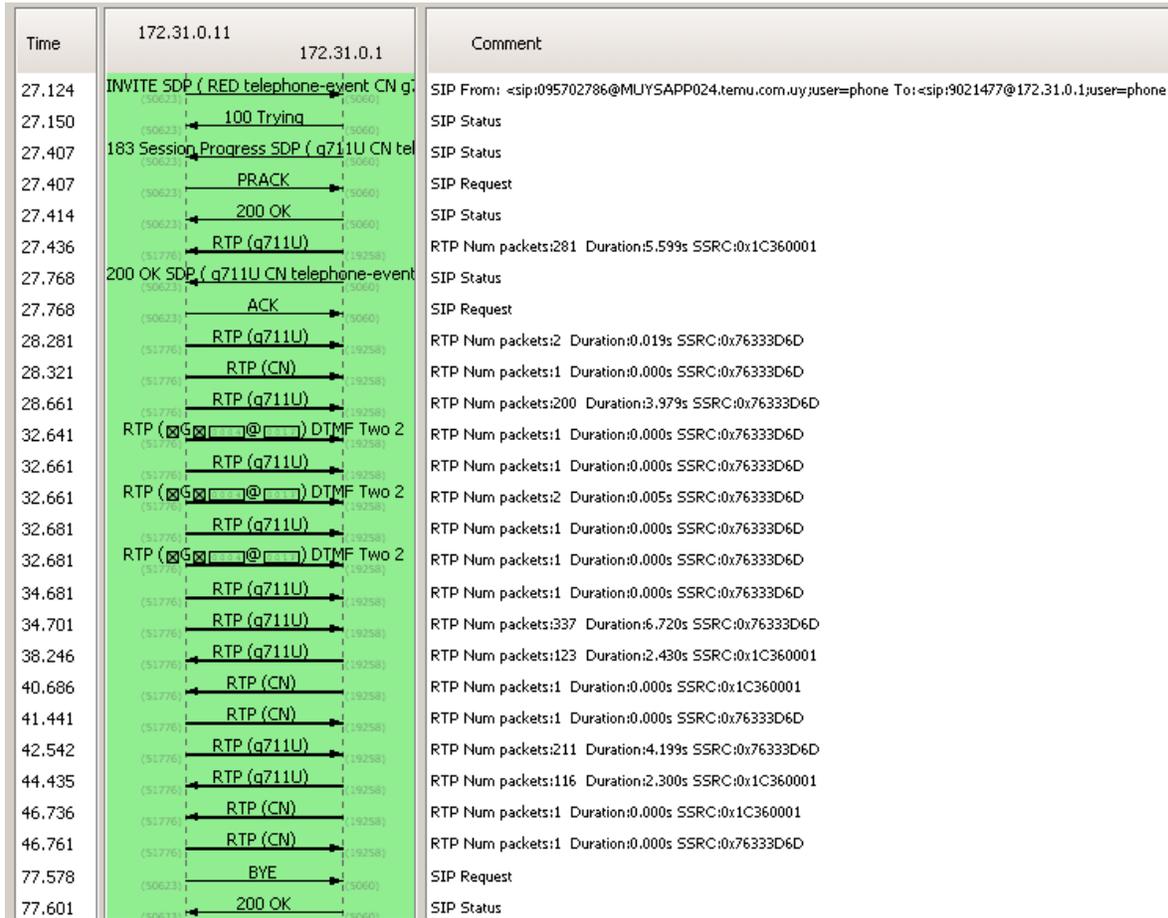


Figura 8.16

La llamada comienza con un “INVITE” del Terminal 1 (172.31.0.11 en este ejemplo) hacia el Terminal 2 (172.31.0.1 en este ejemplo). El Terminal 2 responde con un “Trying” e inmediatamente después con un “Session Progress”. Este mensaje habilita a conectar el audio desde el Terminal 2 hacia el Terminal 1, a los efectos de poder enviar la señal de “Ring Back” o constancia de llamada. El terminal 1 responde con “PRACK” o “Provisional ACK”, el Terminal 2 lo reconoce con un OK y comienza el envío de paquetes de audio RTP (por ahora en forma unidireccional).

Cuando el Terminal 2 atiende la llamada, éste envía un “OK” con cuerpo SDP, intercambian información de codecs y demás capacidades, y comienza el intercambio del medio (paquetes RTP).

En este ejemplo, se ve que como parte del protocolo RTP se utiliza el codec G.711 ley mu, con supresión de silencios y “Confort Noise”

### 8.3 Otros protocolos de señalización en VoIP

Además de los protocolos H.323 y SIP, ampliamente difundidos, existen varios protocolos propietarios, desarrollados para funcionar únicamente con ciertas aplicaciones o equipos. Entre ellos:

- **SCCP (Skinny Call Control Protocol)**  
Es un protocolo de señalización propietario de Cisco, utilizado entre su servidor de telefonía (“Call Manager”) y los teléfonos.
- **IAX2 (Inter-Asterisk eXchange protocol)**  
Es un protocolo de señalización propietario de Asterisk, utilizado para la conexión de varios servidores Asterisk, y también utilizando entre el servidor de telefonía Asterisk y los teléfonos. Está publicado en carácter informativo en el RFC 5456 de la IETF.
- **Unistim**  
Es un protocolo de señalización propietario de Avaya (comprado a Nortel), utilizado para la conexión entre teléfonos y el servidor de telefonía de Avaya “Communication Server”.
- **NOE (New Office Environment)**  
Es un protocolo de señalización propietario de Alcatel, utilizado entre su servidor de telefonía y los teléfonos.
- **Skype**  
La aplicación Skype utiliza sus propios protocolos de señalización de VoIP.

## 9 Gestión de proyectos de VoIP

Además de los desafíos propios que presenta llevar adelante cualquier tipo de proyecto de tecnología, los proyectos que involucran la implementación de tecnologías de VoIP tienen sus características y desafíos particulares. No es cierto, como a veces se dice a la ligera, que la tecnología de VoIP sea sencilla de implementar, administrar, mantener y gestionar. Sin embargo, si esta tecnología es adoptada según un proceso adecuado, sus ventajas finales compensan el esfuerzo [69].

Para ingresar en el mundo de la VoIP, o de la Telefonía IP, existen diferentes caminos. Un camino posible es pensar en una solución “completamente IP”. En este escenario, los eventuales sistemas TDM existentes son completamente reemplazados por la nueva tecnología, sin reutilizar la inversión realizada y ni la infraestructura existente. A nivel corporativo, una alternativa a este camino es arrendar los servicios de telefonía, delegando a un tercero la adquisición y gestión de los servidores. Esta modalidad, del tipo “hosting”, está comenzando a ser comercializada en diversos países.

Otro camino posible consiste en migrar sistemas TDM existentes hacia plataformas con soporte IP, generalmente en un ambiente “híbrido”. En este escenario, la migración hacia IP se realiza en forma gradual, manteniendo los servicios TDM existentes e implementado en tecnología IP las ampliaciones, o las aplicaciones que por su naturaleza requieren necesariamente de este tipo de tecnologías.

Ambos caminos son posibles, y cada uno tiene sus ventajas y desventajas. Según los datos mencionados en [69], el 69% de las Empresas se están inclinando por una solución “híbrida”, mientras que el 31% restante por una solución de reemplazo completo. Cualquiera sea la elección, los desafíos son importantes.

Las ventajas y beneficios de la tecnología de VoIP son muchos, así como sus aplicaciones. Sin embargo, en algunos casos, estas tecnologías no presentan mejoras significativas, o las soluciones buscadas pueden ser implementadas con la tecnología existente. Es por esto que las expectativas establecidas por los interesados del proyecto (“stakeholders”) deben ser manejadas apropiadamente. Muchas personas del área gerencial no técnica, y aún en algunos casos también del área técnica, no tienen claro que es lo que la telefonía IP puede brindar y que es lo que se puede lograr mediante tecnologías ya instaladas. Es por esta razón que el asesoramiento, desde las etapas de gestación de este tipo de proyectos, es fundamental.

El análisis de ROI (Retorno de la Inversión) debe siempre ser realizado, como en todo proyecto. Cuando el motivador principal de la tecnología es el ahorro de costos, y se eliminan enlaces dedicados o llamadas de larga distancia, este ahorro puede ser fácilmente cuantizable. Sin embargo, cuando los motivadores

principales tienen que ver con el aumento de la productividad, los ahorros en costos son más difíciles de calcular. Es indudable que la integración que permite la tecnología IP redundante, potencialmente, en una mayor productividad, plasmada en mayores tiempos libres para los empleados, o en poder contactar más rápidamente a las personas buscadas. Sin embargo, resulta difícil traducir esto en un ahorro concreto de dinero.

Los problemas técnicos de implementación no deben ser subestimados. La VoIP es aún una tecnología emergente, y la experiencia general en los mercados es, en algunos aspectos, aún escasa. Problemas ya resueltos hace años, o inexistentes, en las redes TDM, aparecen con cierta frecuencia en las implementaciones de VoIP (por ejemplo, problemas de audio en un solo sentido, problemas en el discado DTMF, calidad deficiente del audio, etc.). Es de esperar que, aún siguiendo las mejores prácticas, algunos de estos problemas se presenten al momento de la implementación. Sin mencionar que, de no existir un adecuado diseño, es casi garantizado que varios de estos problemas efectivamente sucederán.

Según las prácticas reconocidas del PMI (Project Management Institute), un proyecto puede ser dividido en diversos procesos, tal como se ilustra en la Figura 9.1 [70]. Cada uno de estos procesos, que conforman el ciclo de vida de un proyecto, tiene sus particularidades, y aplican ciertas áreas de conocimiento.

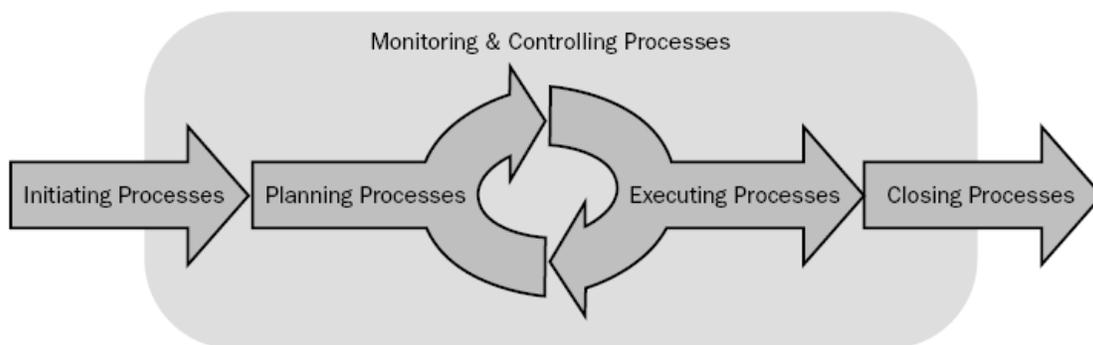


Figura 9.1

Es de notar que el comienzo del proyecto se da mucho antes de comenzar implementación, que es parte del proceso de ejecución. Las etapas previas a la implementación son tanto, o quizás más importantes, que la implementación. A continuación se presentan las actividades típicamente realizadas en cada proceso del proyecto, con foco en lo específico de las tecnologías de VoIP.

### 9.1 Iniciando un proyecto de VoIP (Procesos de Iniciación)

Durante esta fase del proyecto, se desarrollan las siguientes actividades:

- Desarrollo del “caso de negocio”

- Determinar el ROI (Retorno de la inversión)
- Establecer el alcance del proyecto, a alto nivel
- Identificar a los grupos de interés (“stakeholders”) y determinar sus necesidades y expectativas
- Identificar las restricciones conocidas
- Crear un “Project charter”, o acta de inicio del proyecto

Para los casos de proyectos de VoIP, en algunos casos, no es fácil determinar el ROI, como ya fue mencionado. Las reducciones de costos en lo que respecta a los aspectos “blandos”, como son el aumento de la productividad y la disminución de los tiempos en tomas de decisiones pueden ser estimadas.

## **9.2 Planificando un proyecto de VoIP (Procesos de Planificación)**

Durante esta fase del proyecto, se desarrollan las siguientes actividades:

- Definición de un alcance detallado
- Estimación detallada del presupuesto y asignación del presupuesto
- Creación de la Estructura de Desglose del Trabajo (WBS)
- Identificación del camino crítico
- Desarrollo de los diversos planes de gestión del proyecto
- Identificación y cuantificación de riesgos

En los proyectos de VoIP, al igual que en diversos proyectos de tecnología, es importante establecer el grado de ayuda o contrataciones externas deseado. Esta ayuda externa, típicamente tercerizada en Empresas y/o Consultores especializados, puede incluir alguna, varias, o todas las fases del proyecto (por ejemplo, desde el diseño hasta la implementación). Sin embargo, en cualquier caso, los clientes finales, quienes finalmente utilicen la tecnología, deberán ser involucrados, para poder lograr el éxito del proyecto.

Los riesgos en los proyecto de VoIP son variados, y deben ser identificados lo antes posible en la etapa de planificación. Entre los riesgos comunes se encuentran:

- Problemas de calidad de la voz
- Problemas de seguridad
- Infraestructuras de cableado que no soporten apropiadamente la nueva tecnología
- Infraestructura de red de datos que no soporte apropiadamente la nueva tecnología
- Incrementos no detectados en el tráfico
- Problemas técnicos de una tecnología emergente

Muchos de estos riesgos se pueden minimizar realizando previamente un análisis del estado de la red, llamado habitualmente “network assessment” es inglés. Este tipo de análisis permite detectar, de manera temprana, el estado de una infraestructura existente para el soporte de una nueva aplicación, por ejemplo, Voz y Video sobre IP. Conociendo a priori el tráfico esperado, es posible incluso realizar simulaciones, que permitan determinar la calidad de la voz y el impacto en la red y sus equipos asociados al implementar estas nuevas tecnologías.

Los resultados de un análisis de estado de la red sirven para realizar una adecuada planificación del proyecto, incluyendo, si corresponde, las tareas y el presupuesto necesario para mejorar la infraestructura, lo que eventualmente puede ser necesario para soportar las nuevas tecnologías. Es altamente recomendable realizarlo en todo proyecto de VoIP.

### **9.3 Ejecutando un proyecto de VoIP (Procesos de Ejecución)**

Como parte de la etapa de ejecución, se realizan generalmente las siguientes actividades:

- Determinación y asignación de el o los equipos de trabajo asignados al proyecto
- Realizar y gestionar los contratos de sub contratistas, incluyendo los contratos de hardware, software y servicios.
- Implementación, de acuerdo al alcance detallado realizado en el proceso de planificación

La etapa de ejecución es, sin dudas, sumamente importante. Sin embargo, la etapa de ejecución no supe, en ningún caso, la planificación. En caso de tener varios sub contratos, es de suma importancia mantener el control de los mismos, tener claros los límites de responsabilidades, y saber gestionar cualquier tipo de problemas que se presente entre éstos. Los proyectos de VoIP tienen impacto, generalmente, sobre diversos departamentos y grupos de trabajo. Los alcances de cada parte involucrada que debe aportar al proyecto deben estar bien establecidos y correctamente administrados para garantizar el éxito del proyecto. Entre estos grupos se incluye al departamento de IT, al de seguridad de la información, al departamento de telefonía, y a las gerencias que utilicen la tecnología resultante del proyecto, entre otros.

### **9.4 Controlando un proyecto de VoIP (Procesos de Monitoreo y Control)**

Durante las etapas de planificación y ejecución es habitual que surjan problemas, o que se materialicen los riesgos. Es por tanto muy importante en todo proyecto mantener un proceso de monitoreo y control permanente.

Como parte de este proceso, generalmente se realizan las siguientes actividades:

- Monitorear y controlar el avance general del proyecto
- Realizar la verificación y control de que se esté cumpliendo con el alcance definido
- Realizar un control de costos
- Realizar controles de calidad
- Tareas relativas a reportes de avances
- Mantener los riesgos monitoreados y controlados. En casos que corresponda, gestionar la implementación de las medidas correctivas previstas
- Administrar a los sub contratos
- Realizar un control integral de cambios

Los procesos de monitoreo y control en proyectos de VoIP deben tener especial cuidado en lo que respecta a la gestión de riesgos. Siendo ésta una tecnología emergente, es posible que se presenten inconvenientes no previstos y se deban tomar acciones correctivas o de mitigación apropiadas. Los problemas de integración o de “frontera” entre diversos sectores son frecuentes, y muchas veces difíciles de prever en la etapa de planificación. También problemas no previstos de seguridad de la información pueden presentarse.

## **9.5 Finalización del proyecto de VoIP (Procesos de Cierre)**

Los procesos de cierre generalmente incluyen las siguientes actividades:

- Obtener la aceptación de los interesados
- Finalizar los sub-contratos
- Des-asignar a los equipos de trabajo y recursos del proyecto
- Documentar las lecciones aprendidas
- Archivar la documentación para referencias futuras

El cierre de este tipo de proyectos da lugar a una fase de utilización de la tecnología, la operación y el mantenimiento. En algunos casos, la nueva tecnología trae consigo nuevas formas de operación y eventualmente la consolidación de diferentes grupos de soporte en uno solo. Es posible que esto conlleve a la realización de un nuevo proyecto, que consista en establecer los procedimientos y la manera de gestionar y operar la nueva red de VoIP. Esto puede ser o no parte del proyecto principal.

Una vez cerrado el proyecto, es recomendable realizar un análisis del éxito del mismo, las mejoras de la productividad y de costos obtenidas en la operación, y el grado de satisfacción de los usuarios. Hacer visibles estas mejoras ayudará a conseguir presupuesto para nuevas ampliaciones o futuros proyectos de tecnologías relacionadas.

Es conveniente también realizar nuevamente un análisis de la red, determinando su desempeño durante la operación y verificando que las premisas y supuestos realizados se estén cumpliendo, o implementado correcciones en caso de ser necesarias. Idealmente este tipo de estudios debe realizarse periódicamente (por ejemplo una vez por mes o cada dos meses) durante los primeros seis meses de operación de la nueva tecnología, y cada vez que se incorporan nuevos servicios o existe algún incremento en el tráfico.

## Referencias

---

- [1] Conceptos de Telefonía Corporativa, Versión 08, José Juskowicz (Julio 2009)
- [2] Redes de Datos, Versión 05, José Juskowicz (Agosto 2008)
- [3] Traffic Trading in the Competitive International Voice Market, Stefano Bregni and Maurizio Decina, IEEE Communication Magazine (August 2009)
- [4] Comunicaciones Unificadas, José Juskowicz (Setiembre 2009)
- [5] Codificación de Voz y Video, José Juskowicz (Marzo 2011)
- [6] Recommendation G.711: "Pulse Code Modulation (PCM) of voice frequencies", CCITT, 1988.
- [7] Recommendation G.723.1: "Dual Rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s", ITU-T, May 2006.
- [8] Recommendation G.728: "Coding of speech at 16 kbit/s using Low-delay code excited linear prediction", CCITT, 1992.
- [9] Recommendation G.729: "Coding of speech at 8 kbits using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)", ITU-T, Jan 2007.
- [10] Adaptive Multi-Rate (AMR) speech codec, ETSI TS 126 090 V9.0.0, 2010-01
- [11] Recommendation G.722: "7 kHz audio-coding within 64 kbit/s", CCITT, 1988.
- [12] Recommendation G.722.1: "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss", ITU-T, 05/2005.
- [13] Recommendation G.722.2: "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)", ITU-T, 07/2003.
- [14] Recommendation G.711.1: "Wideband embedded extension for G.711 pulse code modulation", ITU-T, 03/2008.
- [15] Recommendation G.729.1: "G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729", ITU-T, 05/2006.
- [16] Overview of the Microsoft RTAudio Speech Codec, Microsoft, 2006.
- [17] SILK – Super Wideband Audio Codec, <https://developer.skype.com/silk>
- [18] Recommendation G.719: "Low-complexity, full-band audio coding for high-quality, conversational applications", ITU-T, 06/2008.

- [19] RFC 3550: “RTP: A Transport Protocol for Real-Time Applications”, H. Schulzrinne et al (July 2003)
- [20] RFC 3551: “RTP Profile for Audio and Video Conferences with Minimal Control”, H. Schulzrinne et al (July 2003)
- [21] RFC 2833: “Payload for DTMF Digits, Telephony Tones and Telephony Signals”, H. Schulzrinne et al (May 2000)
- [22] RFC 2250 Payload Format for MPEG1/MPEG2 Video  
D. Hoffman et al, January 1998
- [23] WHITEPAPER – IP Streaming of MPEG-4: Native RTP vs MPEG-2  
Transport Stream  
Alex MacAulay, Boris Felts, Yuval Fisher, October 2005
- [24] DVB IP Phase 1 handbook , ETSI TS 102 034, “Digital Video Broadcasting (DVB); Transport of MPEG-2 Based DVB Services over IP Based Networks”, March 2005.
- [25] RFC 3016 RTP Payload Format for MPEG-4 Audio/Visual Streams  
Y. Kikuchi et al, November 2000
- [26] RFC 3640 RTP Payload Format for Transport of MPEG-4 Elementary Streams, J. van der Meer et al, November 2003
- [27] RFC 3984 RTP Payload Format for H.264 Video, S. Wenger et al, febrero 2005
- [28] VQEG Phase I Test Sequences. [Online]. Disponibles en:  
[ftp://vqeg.its.bldrdoc.gov/SDTV/VQEG\\_PhaseI/TestSequences/Reference/](ftp://vqeg.its.bldrdoc.gov/SDTV/VQEG_PhaseI/TestSequences/Reference/)
- [29] Video coding with H.264/AVC: Tools, Performance, and Complexity  
Jörn Ostermann, Jan Bormans, Peter List, Detlev Marpe, Matthias Narroschke, Fernando Pereira, Thomas Stockhammer, and Thomas Wedi  
IEEE Circuits and Systems Magazine, First Quarter 2004
- [30] Calidad de Voz y Video, José Joskowicz (Marzo 2011)
- [31] Recommendation ITU-T G.711 Appendix 1, A high quality low-complexity algorithm for packet loss concealment with G.711, 9/1999.

- [32] Recommendation ITU-T G.107 The E-model, a computational model for use in transmission planning, December 2011, <http://www.itu.int/rec/T-REC-G.107/e>
- [33] E-model tutorial  
<http://www.itu.int/ITU-T/studygroups/com12/emodelv1/introduction.htm>
- [34] TIA/TSB 116-A Telecommunications - IP Telephony Equipment – Voice Quality Recommendations for IP Telephony, Mar 1, 2006
- [35] Recommendation ITU-T G.113 “Degradaciones de la transmisión debido al tratamiento de las señales vocales”, 2001
- [36] RFC 3611: “RTP Control Protocol Extended Reports (RTCP XR)”, T. Friedman et al (November 2003)
- [37] Estudio de la Medida de la Calidad Perceptual de Video, José Joskowicz, Universidad de Vigo (Marzo de 2008)
- [38] DCT Quantization Noise in Compressed Images, Mark A. Robertson and Robert L. Stevenson, IEEE Transactions on Circuits and Systems For Video Technology, Vol. 15, No. 1, January 2005
- [39] Digital Video Image Quality and Perceptual Coding, H.R. Wu and K.R Rao 2006, CRC Press
- [40] User-Oriented QoS Analysis in MPEG-2 Video Delivery, O Verscheure, P Frossard, M Hamdi, Real Time Imaging 5, 1999, pp 305-314
- [41] Quality Monitoring of Video Over a Packet Network, Amy R. Reibman, Vinay A. Vaishampayan and Yegnaswamy Sermadevi, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 6, NO. 2, APRIL 2004
- [42] Visibility of individual packet losses in MPEG-2 video  
Amy R. Reibman, Sandeep Kanumuri, Vinay Vaishampayan and Pamela C. Cosman, IEEE International Conference on Image Processing) 2004, ICIP'04, Vol 1, pp 171-174
- [43] Delivery of MPEG Video Streams with constant perceptual quality of service  
Quaglia, D. De Martin, J.C. , IEEE, Proceedings International Conference on Multimedia, 2002
- [44] Estimation of packet loss effects on video quality, Bouazizi, I., IEEE, First International Symposium on Control, Communications and Signal Processing, 2004, pp 91-94.

- [45] Packet Loss Resilience for MPEG-4 Video Stream over the Internet, Jae-Young Pyun, Jae-Han Jung, Jae-Jeong Shim, IEEE Transactions on consumer electronics, Vol 48, issue 3, August 2002
- [46] Adaptive Media Playout of Low Delay Video Streaming Over Error Prone Channels , Mark Kalman, Eckehard Steinbach, Bernd Girod, IEEE Transactions on Circuits and Systems for Video Technology, Vol 14, no 6, June 2004
- [47] Recomendación ITU-T G.1070: “Opinion model for video-telephony applications” (Abril 2007)
- [48] VoIP Technologies - A comprehensive guide to Voice over Internet Protocol Nortel Press, March 2008, ISBN 978-0-9801074-0-1
- [49] IEEE 802.1q – VLAN, IEEE Standards for Local and metropolitan area networks—Virtual Bridged Local Area Networks, 2003 Edition, <http://standards.ieee.org/getieee802/802.1.html>
- [50] IEEE 802.1p - Traffic Class Expediting and Dynamic Multicast Filtering (published in 802.1D-1998)
- [51] Overcoming VoWLAN Challenges, VoWLAN v1.0906, Lucent Technologies, 2006
- [52] Enterprise Mobility 4.1 Design Guide, Cisco Validated Design I, 2008
- [53] IEEE 802.11x Port-Based Network Access Control (Diciembre 2004)
- [54] IEEE 802.11r Amendment 2: Fast Basic Service Set (BSS) Transition (Julio 2008)
- [55] IEEE 802.11e: Quality of Service Enhancements , (Noviembre 2005)
- [56] Wi-Fi CERTIFIED™ for WMM™ - Support for Multimedia Applications with Quality of Service in Wi-Fi® Networks, Wi-Fi Alliance (September 1, 2004) [http://www.wi-fi.org/knowledge\\_center\\_overview.php?docid=4480](http://www.wi-fi.org/knowledge_center_overview.php?docid=4480)
- [57] Recommendation H.323 Version 7: “Packet-based multimedia communications systems”, ITU-T (December 2009)
- [58] Recommendation H.245 Version 16 “Control protocol for multimedia communication” , ITU-T (May 2011)

- [59] Recommendation H.225.0 Version 7: “Call signalling protocols and media stream packetization for packet-based multimedia communication systems”, ITU-T (December 2009)
- [60] Recommendation Q.931: “ISDN user-network interface layer 3 specification for basic call control”, CCITT (May 1998)
- [61] RFC 3261: SIP: Session Initiation Protocol. J. Rosenberg et al. June 2002
- [62] RFC 3262: Reliability of Provisional Responses in Session Initiation Protocol (SIP) J. Rosenberg et al., June 2002
- [63] RFC 3263: Session Initiation Protocol (SIP): Locating SIP Servers. J. Rosenberg et al. June 2002
- [64] RFC 3264: An Offer/Answer Model with Session Description Protocol (SDP) J. Rosenberg et al. June 2002
- [65] RFC 3265: Session Initiation Protocol (SIP)-Specific Event Notification AB. Roach, June 2002
- [66] RFC 3266: Support for IPv6 in Session Description Protocol (SDP) S. Olson et al, June 2002
- [67] SIP: Understanding the Session Initiation Protocol. — 3ed Edition — (Artech House telecommunications library) ISBN 13: 978-1-60783-995-8, Alan B. Johnston, 2009
- [68] RFC 2327: SDP: Session Description Protocol M. Handley et al., April 1998
- [69] VoIP: Retos y recomendaciones, <http://www.networkworld.es/VoIP:-retos-y-recomendaciones/seccion-/articulo-173362>
- [70] A guide to the Project Management Body of Knowledge (PMBOK guide, 5<sup>th</sup> edition), PMI, 2013