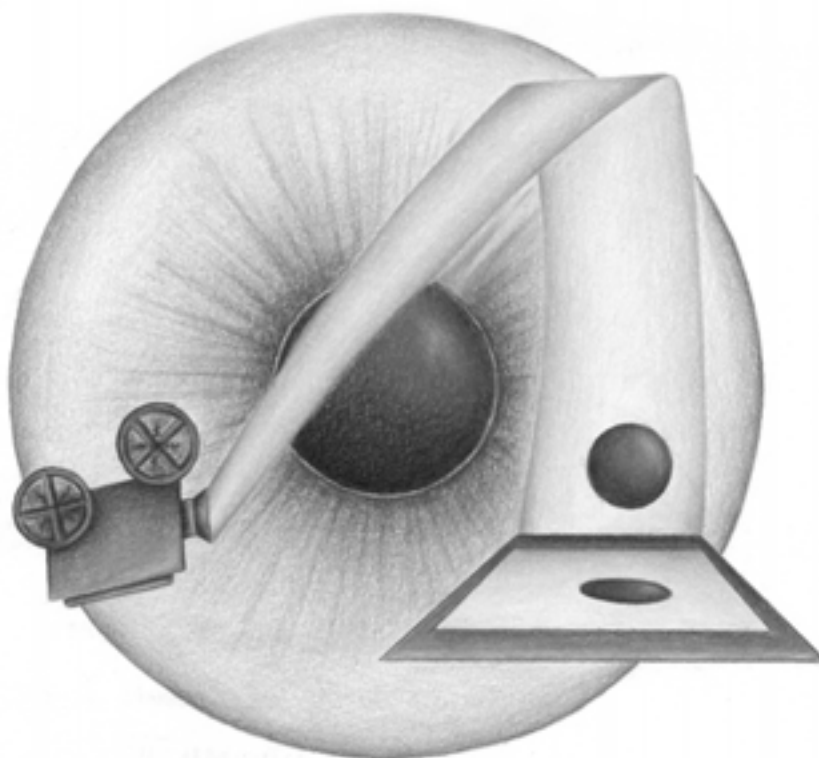


Transmisión de vídeo en redes



Índice

Etapa 1: Transmisión de vídeo en redes de datos	5
Introducción	5
Estándares de transmisión de contenidos multimedios en redes de datos	6
Codificadores de vídeo	7
Codificación de audio: G.711	8
Codificadores de audio opcionales	9
El estándar H.320	10
Protocolo de multiplexación H.221	11
Otros protocolos utilizados en el H.320	13
Estándares H.321 y H.322	15
Estándares de segunda generación	15
Protocolo de control H.245	17
Codificadores de audio y vídeo en estándares de segunda generación	18
Redes de conmutación de paquetes. El estándar H.323	18
Inicio de llamada y descomposición en paquetes: H.225.0	19
Fraccionamiento de streams multimedia en paquetes	20
Limitaciones del H.323	21
Redes de conmutación de circuitos. El estándar H.324	22
ATM y RDSI de banda ancha. El estándar H.310	23
Interoperabilidad entre los estándares de la serie H.3xx	24
Etapa 2: Descarga y reproducción de vídeo en Internet	25
Introducción	25
Primeras arquitecturas de streaming	26
VivoActive 1.0. La primera aplicación de <i>streaming</i>	26
El protocolo de transporte TCP y el <i>streaming</i> multimedia	26
<i>Buffering</i> y <i>rebuffering</i>	26
Formatos genéricos y específicos de <i>streaming</i> por medio de TCP	27
El formato ASF	28
RealAudio 1.0	28
Transmisión de vídeo por medio de <i>streaming</i>	30
Servidores de streaming	32
Características básicas de un servidor de <i>streaming</i>	32
Protocolos utilizados en servidores de <i>streaming</i>	33
Protocolos de transporte: TCP frente a UDP	33
El protocolo RTP	34
El protocolo RTCP	35
Control y ajuste de la velocidad de transferencia	36
Multicast	37

Etapa 1: Transmisión de vídeo en redes de datos

Introducción

En este módulo consideraremos las tecnologías utilizadas para la distribución de señales de vídeo a través de redes de datos. Generalmente, el vídeo se transmite conjuntamente con una o varias bandas de audio y señales de control de la comunicación. Para la codificación de las componentes de vídeo se utilizan sistemas de compresión que pueden ser propietarios o estándares y que ya hemos estudiado en capítulos anteriores. De forma análoga, la señal de audio también se comprime y debe multiplexarse con la señal de vídeo para que la información de ambas pueda recibirse al mismo tiempo, sincronizarse en el receptor y reproducirse. Los estándares de transmisión de vídeo se ocupan de especificar el formato y los protocolos que se utilizarán para entrelazar las señales de vídeo y audio en un único *stream* de datos que, además, incluye diferente información de control sobre la comunicación.

Los formatos de compresión de audio y vídeo utilizados en la transmisión de estas señales digitales en redes de datos coinciden con los estándares y métodos propietarios utilizados para el almacenamiento de la información (aplicaciones DSM –*Digital Storage Media*–). Así, es habitual la transferencia y visualización de archivos de audio y vídeo por Internet utilizando algunos de los compresores propietarios estudiados en el capítulo anterior o la transmisión de señales de videoconferencia utilizando el estándar H.263 o MPEG-4, también analizados anteriormente. En redes de alta velocidad (ATM) o intranets es usual utilizar vídeo de alta calidad compatible con el formato MPEG-2.

No obstante, aunque los formatos de compresión sean los mismos, las redes de datos presentan particularidades propias que deben tenerse en cuenta, ya que pueden afectar directamente al proceso de decodificación y reproducción. Así, en las redes de conmutación de paquetes, como Internet, puede producirse la pérdida de paquetes que representan datos significativos sobre la señal de audio y vídeo o pueden recibirse los paquetes desordenados. Los protocolos y procedimientos de comunicación deben tener en cuenta todas estas peculiaridades y definir procesos que garanticen la correcta decodificación de los datos bajo las condiciones que impone la red de datos.

Las principales redes de transmisión de datos que consideraremos a lo largo de este capítulo son las siguientes:

- La red digital de servicios integrados (RDSI) de banda estrecha.
- RDSI de banda ancha. ATM.
- Redes locales.
- Redes de acceso por conmutación de circuitos (red telefónica convencional –POTS–).
- Redes de conmutación de paquetes. Internet.

También pueden considerarse como redes de transmisión de datos la difusión de señales de televisión a través de satélite, cable o enlaces terrenos. No obstante, debido a la especial importancia de este tipo de redes, se considerarán sus características por separado en el siguiente módulo.

En este primer módulo empezaremos considerando los principales estándares para la transmisión de señales de videoconferencia en todo tipo de redes de datos. Estos estándares definen varios procedimientos de control de comunicaciones de audio y vídeo en tiempo real para todo tipo de redes y establecen las bases de la transmisión de material multimedia en redes digitales para otras aplicaciones como la difusión de material por Internet o los sistemas de audio y vídeo a la carta. Posteriormente analizaremos la filosofía de funcionamiento de los servidores de *streaming* para la difusión de vídeo o audio en tiempo real a través de Internet o de redes que funcionen según el protocolo TCP/IP. Debido a que muchos de los aspectos relacionados con los protocolos de transmisión son más propios de las tecnologías de redes de datos que de los sistemas audiovisuales, sólo presentaremos los aspectos más significativos que hacen referencia a la transmisión de señales multimedia y omitiremos los detalles concretos de los protocolos de control de la comunicación.

Estándares de transmisión de contenidos multimedia en redes de datos

La elaboración de los estándares de transmisión de señales multimedia en tiempo real a través de redes de datos son responsabilidad de la ITU (International Telecommunication Union). Se trata de un organismo internacional dependiente de las Naciones Unidas que se encarga de regular las recomendaciones y normativas de todos los sistemas de telecomunicación. Los estándares a los que hace referencia la transmisión de materiales multimedia en tiempo real, incluyendo la videoconferencia, se conocen como la serie H y engloban todos las posibles redes de transmisión de datos. En los siguientes apartados veremos un resumen de las principales características de estos estándares, las funcionalidades que proporcionan y los módulos en los que se descomponen, sin entrar en detalles sobre las particularidades de su implementación.

Los estándares H pueden descomponerse en dos generaciones. La primera generación se inicia con el estándar H.320, que aparece en 1990 y que establece las recomendaciones (en la ITU, los estándares se denominan recomendaciones) para la transmisión de señales de videoconferencia a través de canales de la red digital de servicios integrados. Posteriormente, dentro de la primera generación se producen dos nuevas recomendaciones H.321 y H.322, que no son más que una revisión del H.320 para extender el uso de la videoconferencia a redes de banda ancha (ATM) y a redes locales Ethernet.

La segunda generación está formada por los estándares H.323, H.324 y H.310 que aparecen entre 1995 y 1996. Estos estándares incorporan varias mejoras en los pro-

protocolos y admiten códecs de audio y vídeo más actualizados junto con aplicaciones de control avanzadas. Por tanto, incorporan toda la experiencia previa del uso del H.320 e intentan resolver sus principales problemas. El H.323 está orientado a transmisiones en redes locales, el H.324 a transmisiones de baja velocidad, principalmente RDSI de banda estrecha y redes telefónicas convencionales y el H.310 está orientado a la transmisión de material multimedia en redes de banda ancha.

Todos los estándares tienen definidos unos modos básicos, con los que cualquier terminal o equipo tiene que ser compatible, y modos extendidos que proporcionan mejoras en diferentes aspectos de la comunicación o su control. Además, todos ellos recurren a otros estándares para definir aspectos concretos de la comunicación. Así, el H.320 recurre al estándar H.261 como el formato utilizado para la compresión de vídeo.

En la tabla siguiente se proporcionan las principales características de estos estándares y su relación con otros estándares de la ITU que iremos comentando en los sucesivos apartados.

Tabla comparativa de los entornos de aplicación de los estándares H.3xx					
Estándar	Tipo de red	Vídeo	Audio	Múltiplex	Control
H.320	RDSI	H.261	G.711	H.221	H.242
H.321	ATM/RDSI-BA	Adaptación del H.320 a redes ATM / RDS-BA			
H.322	Ethernet	Adaptación del H.320 a redes Ethernet			
H.323	LAN/Ethernet	H.261	G.711	H.225.0	H.245
H.324	POTS	H.263	G.723.1	H.223	H.245
H.310	ATM/RDSI-BA	H.262	MPEG-1	H.222	H.245

Codificadores de vídeo

Todos los estándares H.3xx presentan un conjunto de características comunes en cuanto a la definición de su entorno de aplicación, el tipo de compresión de audio y vídeo, la confección de la trama de multiplexación y las señales de control. En principio, cada estándar tiene definido un modo básico en el que cualquier equipo perteneciente al estándar debe ser capaz de operar. El modo básico es el que se define en la tabla anterior.

Los codificadores de vídeo que pueden utilizarse en el modo básico son el H.261, el H.262 y el H.263. El H.262 es la versión de la ITU del MPEG-2 (versión ISO). El documento que utiliza la ITU como estándar H.262 es una copia directa del documento original de la ISO. Notad que el único estándar de transmisión que admite el MPEG-2 como soporte de codificación de vídeo es el H.310, cuyo ámbito de aplicación son las redes ATM de alta velocidad o la RDSI-BA (red digital de servicios integrados de banda ancha).

El uso del MPEG-2 en este tipo de redes se justifica para obtener una mejor calidad de imagen. Aunque los niveles de compresión del MPEG-2 son considerablemente menores que los del H.263, este método de codificación presenta una calidad muy superior, por lo que se justifica su uso en redes de banda ancha. Probablemente, a medida que el ancho de banda que llega al usuario vaya aumentando, cada vez será más habitual utilizar estándares de compresión de mayor calidad (MPEG-2, MPEG-4) y se reducirá el uso de estándares para velocidades de transmisión muy bajas (H.263 y sucesivos).

Fijaos también en que los codificadores de vídeo proporcionados en la tabla anterior corresponden al modo básico. Cada estándar dispone de extensiones más avanzadas que permiten utilizar algoritmos de compresión adicionales procedentes de otros estándares o, incluso, algoritmos de compresión propietarios. Así, aunque el modo básico del H.324 sólo contempla el H.263, los modos extendidos contemplan el H.263+ y pronto contemplarán el H.26L.

Los equipos transmisor y receptor pueden negociar mediante los protocolos de control H.242 o H.245 los codificadores de audio y vídeo que se utilizarán durante la conexión. La opción más habitual es que durante el establecimiento de la comunicación el descodificador indica al codificador los modos y opciones con las que puede trabajar. A partir de este momento, el codificador establece la comunicación utilizando el codificador más conveniente e indicando cómo se distribuyen los datos de vídeo, audio y otros en la trama de bits.

Codificación de audio: G.711

El codificador de audio básico contemplado por los estándares de primera generación es el G.711. Este estándar data de 1977 y es el primer formato de audio digital para aplicaciones de telefonía. Se trata de un formato no comprimido que realiza el muestreo de la señal de audio con una frecuencia de 8.000 Hz tomando 8 bits por cada muestra. El número total de bits por segundo nominales es, pues, de 64.000 bps (64 kbps). En efecto:

$$8.000 \text{ muestras/seg} \times 8 \text{ bits/muestra} = 64.000 \text{ bits/seg} = 64 \text{ kbps}$$

El G.711 es el formato utilizado directamente para la transmisión de voz en RDSI. También es el formato utilizado entre centrales telefónicas para la transmisión digital de canales de voz. La calidad que proporciona es equivalente a un ancho de banda de 3 KHz para señal de voz que resulta excelente para aplicaciones de telefonía.

El formato G.711 utiliza convertidores análogo-digitales con leyes de cuantificación no lineales. Existen dos clases diferentes de leyes de cuantificación, denominadas ley mu (se utiliza principalmente en Estados Unidos) y la ley A (que se utiliza en Europa). Cuando se establece una comunicación por medio del estándar H.320 utilizando el G.711 como estándar de audio, es necesario negociar el tipo de ley que se intercambiará a través de los protocolos de comunicación.

El estándar G.711 representa una carga computacional prácticamente nula (sólo se realiza el muestreo y no se requiere compresión) y un retardo en la comunicación inferior a 1 ms.

A partir del estándar de 64 kbps por segundo, pueden obtenerse versiones de 56 kbps y de 48 kbps. Estas versiones proceden de descartar el bit menos significativo o los dos bits menos significativos (obteniendo un total de seis o siete bits por muestra). La calidad, evidentemente, se reduce un tanto. La versión de 7 bits se utiliza en el H.320, con lo que es posible aprovechar el bit restante para la señalización de protocolos de control entre los terminales.

Codificadores de audio opcionales

El G.711 se eligió como estándar dentro del modo básico del H.320 debido a su reducida complejidad computacional y su compatibilidad con la telefonía digital convencional. No obstante, es bastante ineficiente cuando se compara con otros estándares de codificación de audio que se admiten como modos opcionales. Los bits adicionales que no se utilizan para la transmisión de audio pueden aprovecharse para el vídeo de forma que mejore considerablemente su calidad. En la tabla siguiente se muestran los estándares de audio opcionales contemplados en los diferentes formatos de videoconferencia de las recomendaciones H.3xx. Observad que existen dos calidades o anchos de banda claramente diferentes. En los primeros, el ancho de banda está limitado a 3 KHz, mientras que en los segundos la limitación es a 7 KHz, con lo que se obtiene una calidad de audio superior. La tabla también proporciona información orientativa sobre el retardo medio introducido por cada codificador-descodificador y la complejidad computacional asociada. Debemos considerar que al aumentar el factor de compresión aumenta la complejidad del algoritmo y el retardo de la señal de audio. Sin embargo, estos retardos no son excesivamente importantes debido a que el retardo que experimenta la señal de vídeo es aún superior. El H.320 (y todos los estándares posteriores) incorporan protocolos especiales para retardar la señal de audio y sincronizarla con el vídeo de forma que el movimiento de los labios coincida con la señal de vídeo. Este retardo debe contemplarse en todos los modos de codificación de audio, siendo más importante cuanto menor sea el retardo asociado al propio sistema de codificación de audio.

Los codificadores de audio opcionales que se utilizan con mayor frecuencia en el H.320 son el G.728 y el G.722. El primero tiene un ancho de banda equivalente de 3 KHz y utiliza 16 kbps, mientras que el segundo permite trabajar a 56 kbps con una calidad superior (ancho de banda equivalente de 7 KHz). Observad que en aplicaciones de videoconferencia la calidad de la componente de audio es prioritaria con respecto a la calidad del vídeo. El flujo de muestras debe poder reproducirse sin interrupción y el mensaje debe poderse interpretar claramente (inteligibilidad alta). En cuanto al vídeo, no es tan importante que se pierdan algunos fotogramas o que aparezcan errores esporádicos en algunos bloques.

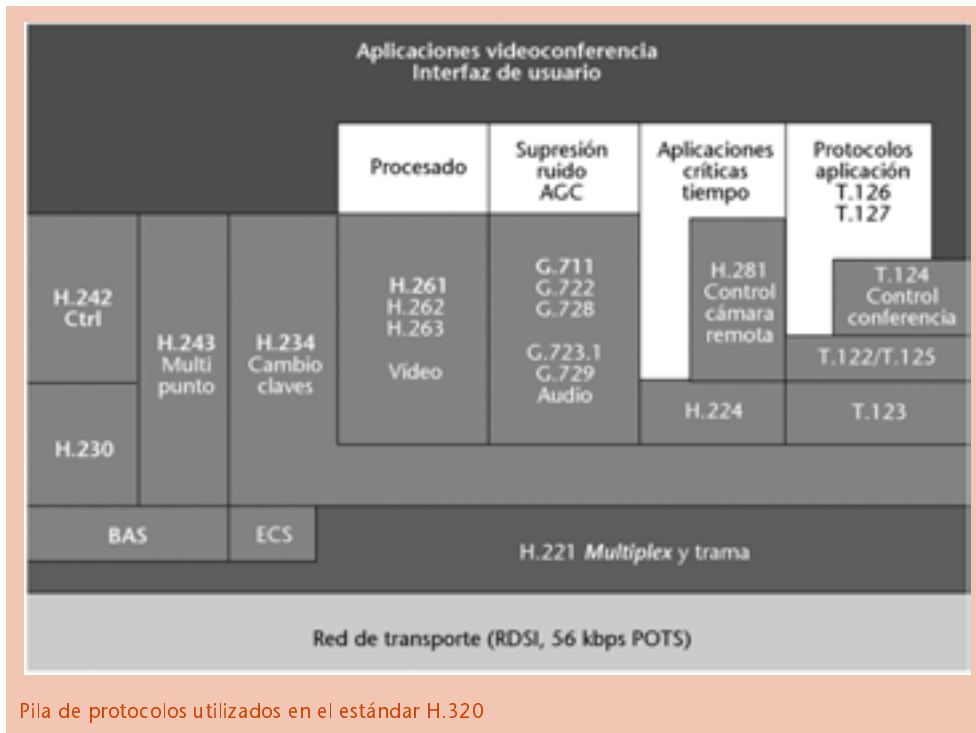
Otros codificadores de altas prestaciones son el G.723.1, que puede transmitir a 5,3 ó 6,4 kbps y el G.729 de 8 kbps, que aunque no son habituales en los modos opcionales del H.320 se utilizan en los estándares de segunda generación.

Codificadores de altas prestaciones					
Estándar	Velocidad transmisión	Ancho de banda	Complejidad	Tamaño marco	Retardo
G.711 (1977)	48 kbps 56 kbps 64 kbps	3 KHz	Aprox. cero	125 μ s	<< 1 ms
G.728 (1992)	16 kbps	3 KHz	35-40 MIPS	0,625 ms	< 2 ms
G.723.1 (1995)	5,3 kbps 6,4 kbps	3 KHz	18-20 MIPS	30 ms	97,5 ms
G.729 (1995)	8 kbps	3 KHz	18 MIPS	10 ms	35 ms
G.729a (1996)	8 kbps	3 KHz	11 MIPS	10 ms	35 ms
G.722 (1988)	48 kbps 56 kbps 64 kbps	7 KHz	10 MIPS	125 μ s	<2 ms
G.16K (1999)	16 kbps 24 kbps 32 kbps	7 KHz	< 15 MIPS	< 20 ms	z60 ms

El estándar H.320

Este estándar contempla la transmisión de señales de videoconferencia a través de RDSI de banda estrecha a diferentes velocidades de transmisión. Las posibles velocidades de transmisión admitidas son múltiplos de 64 kbps. La selección de la velocidad de transmisión para establecer la videoconferencia depende del número de canales básicos que estén disponibles para realizar la comunicación entre el codificador y el decodificador, y puede ir desde los 64 kbps de un canal básico B, los más habituales 128 kbps de dos canales B o los 2 Mbps de un acceso primario a RDSI.

En la siguiente figura se muestran a modo de resumen los diferentes protocolos que se utilizan dentro del estándar H.320 para gestionar los diferentes componentes que intervienen en la videoconferencia. Algunos de estos protocolos (los marcados en negrita) son obligatorios y representan los requisitos mínimos del sistema. Otros son opcionales y pueden estar incorporados o no en los terminales dependiendo de las funcionalidades de los mismos. En los siguientes subapartados comentaremos las características y funcionalidades de los protocolos más representativos.



La red de transporte es generalmente RDSI y utiliza varios canales de 64 kbps para gestionar la videoconferencia entre los dos terminales (o varios terminales cuando se utiliza la opción de multiconferencia). Los datos que se envían a la red de transporte tienen una estructura totalmente compatible con la RDSI, utilizando los mismos periodos de trama y la misma velocidad de señalización. No obstante, en este caso, la interpretación de los bits y su contenido son distintos a como son en el caso de la telefonía convencional.

Para multiplexar la información de audio, vídeo, datos de usuario y control de la comunicación se utiliza el protocolo H.221 (normativo). Este protocolo se encarga de multiplexar los datos y contiene información de control de la comunicación. Los dos canales de vídeo y audio normativos (un canal en cada dirección de la comunicación) pueden estar asociados a varios canales de 64 kbps. El control de la comunicación se incluye también dentro de la trama múltiple y los protocolos de nivel jerárquico superior deberán ser los encargados de interpretar correctamente esta información. La trama está formada por señales de audio y vídeo, el protocolo de control H.242 más protocolos adicionales que son opcionales y que pueden encargarse del encriptado de los datos, el control remoto de la cámara de vídeo, el control de la conferencia, etc. Comentaremos alguna de estas funcionalidades más adelante.

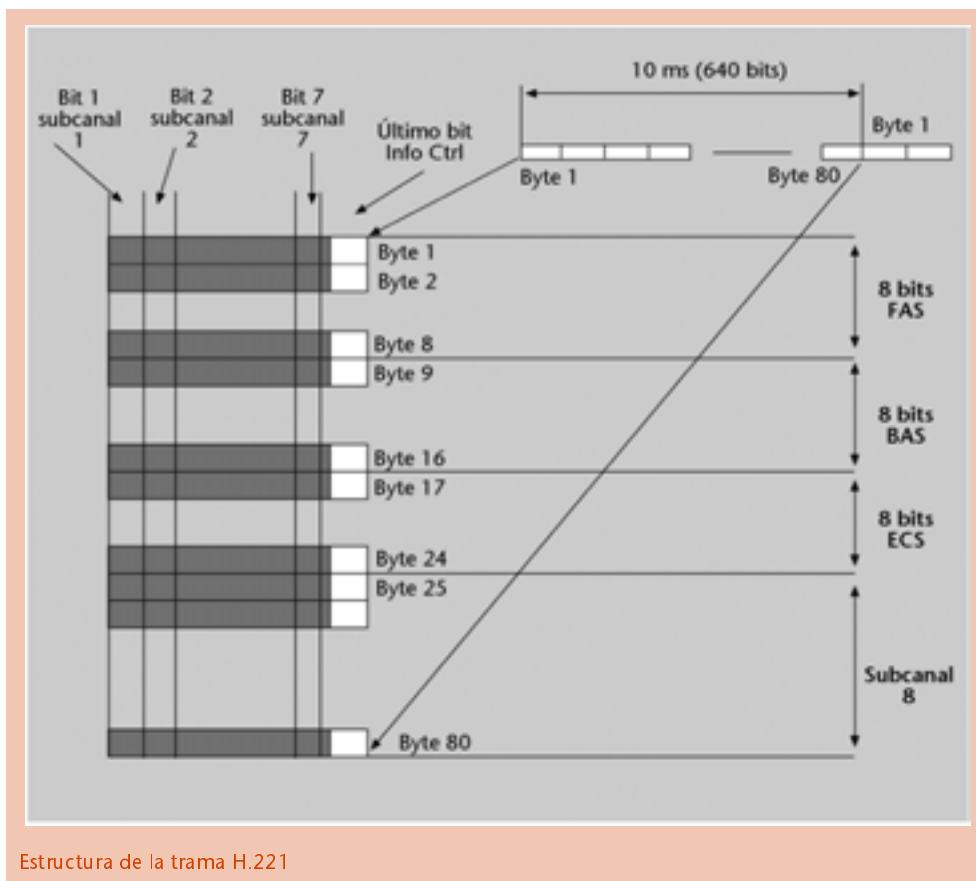
Protocolo de multiplexación H.221

El H.221 se encarga de multiplexar la información de audio, vídeo, datos de usuario e información de control. La trama básica del H.221 tiene una velocidad de transmisión de 64 kbps que es el canal básico de acceso a la red digital de servicios integrados, que se conoce como canal B. Un acceso básico de usuario que contrata RDSI garantiza dos canales del tipo B (128 kbps). Los usuarios pueden contratar más cana-

les. Un ejemplo típico es una configuración de seis canales tipo B que forman una conexión denominada H_0 , de uso habitual entre empresas, y que permite un flujo de comunicación de 384 kbps. Esta configuración permite el establecimiento de videoconferencia con cierta calidad. Otra posible configuración es la de un acceso primario que permite utilizar un ancho de banda de hasta 2 Mbps.

Normalmente, el establecimiento de una videoconferencia exige un mínimo de dos canales tipo B, aunque en la mayoría de los casos, si se desea una calidad aceptable, deben ser más. Los protocolos de control de comunicación, situados en niveles jerárquicos superiores, se encargan de descomponer los distintos tipos de información (*streams* de audio, vídeo y datos) que se envían en cada uno de los canales de 64 kbps de forma multiplexada en el tiempo.

La trama básica del H.221 está formada por 80 *bytes* (640 bits) y tiene una duración de 10 ms. Los 640 bits contienen información de audio, vídeo y canales de control de acuerdo con la distribución que se indica en la siguiente figura. Si los *bytes* se ordenan en una memoria en el orden de transmisión o recepción, se obtiene el diagrama de bits que se representa en la figura. Cada uno de los siete primeros bits de información contiene un subcanal (pueden admitirse configuraciones con dos canales de audio, dos de vídeo y tres canales de usuario). El último bit contiene información de control diversa.



Estructura de la trama H.221

La información de control que se transmite con la trama indica:

- *FAS (Frame Alignment Signal)*. Son ocho bits que proporcionan información sobre el alineamiento de trama y permiten la sincronización de los equipos. Se transmite una palabra clave conocida de antemano en las tramas impares. Las tramas pares se utilizan para señalar eventos entre el transmisor y el receptor como la correcta sincronización de la trama y para sincronizar los distintos canales de 64 kbps que forman una conexión multicanal entre los dos terminales.
- *BAS (Bit rate Allocation Signal)*. Estos ocho bits definidos en cada una de las tramas se utilizan para especificar la posición de los canales de vídeo, audio y datos en la trama de manera que se proporcionen las velocidades de transmisión deseadas para cada canal. También se utiliza para intercambiar mensajes de control e indicación. Ésta es una de las limitaciones más importantes de H.320 debido a que se transmiten pocos bits por trama, de forma que la señalización entre los terminales es muy lenta. El establecimiento de las comunicaciones es muy tedioso debido a que los terminales deben negociar los protocolos y codificadores de audio y vídeo que utilizarán a través de este canal. También se negocian los cambios en tiempo real que se realizarán durante la comunicación. Una de las ventajas de incorporar la señalización BAS en la misma trama que los canales de audio y vídeo es que permite sincronizar el intercambio de mensajes con los contenidos audiovisuales de forma muy simple.
- *ECS (Encryption Control Signal)*. Estos ocho bits se utilizan para enviar información de intercambio de códigos para la encriptación de la información de audio y vídeo en el canal de datos. La encriptación es opcional y viene gestionada por el protocolo H.233. Cuando no se usa el encriptado, estos bits quedan libres para su uso como información de audio o vídeo.

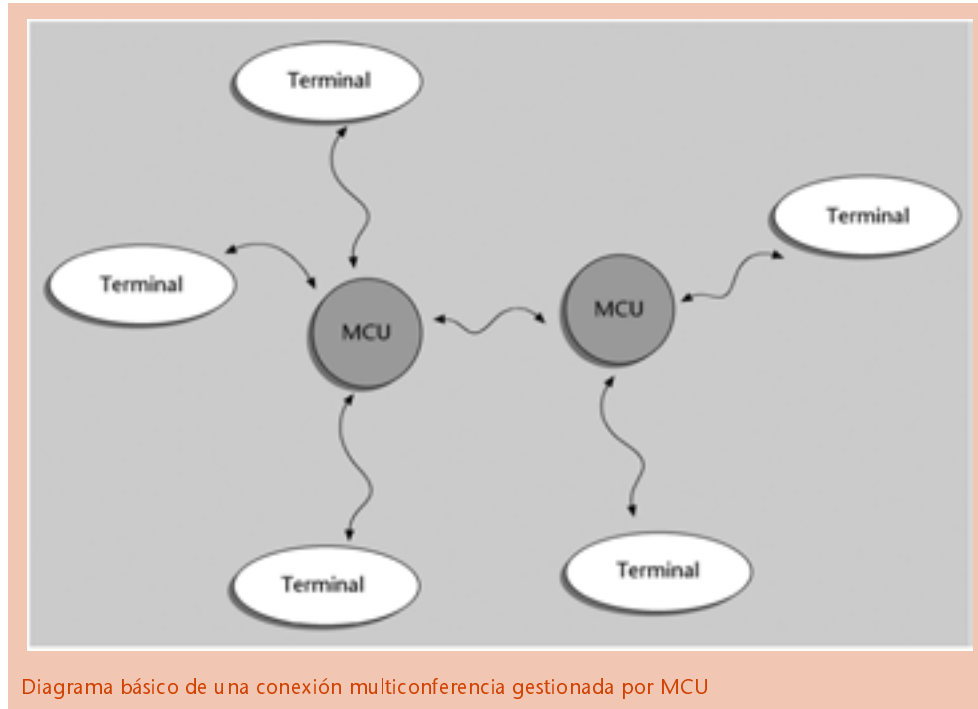
Otros protocolos utilizados en el H.320

El protocolo H.230 es normativo y se utiliza para enviar un conjunto de mensajes que incorporan información diversa como cuando se envían datos de audio y vídeo válidos, solicitudes para congelar imágenes que se enviarán como intra (por ejemplo para enviar y visualizar documentos con mayor precisión que con el modo de vídeo convencional), etc.

El protocolo de control H.242 es el encargado de cubrir los procedimientos básicos de establecimiento de la llamada y de la operatividad de la comunicación. Es también un protocolo obligatorio.

Las conexiones multiconferencia están implementadas a partir de los estándares H.231 y H.243. Las conexiones RDSI son inherentemente punto a punto, por lo que las conexiones de varios usuarios compartiendo canales deben ser gestionada por un dispositivo que actúa de puente en la propia central. Este dispositivo se conoce como MCU (*Multipoint Control Unit*). Las conexiones de cada usuario con la central son

punto a punto. El MCU recibe las señales de audio y de vídeo de todos los terminales conectados y selecciona el audio y vídeo que debe enviarse a cada uno de los terminales. La siguiente figura muestra un esquema simplificado de varios interlocutores conectados por medio de una MCU.



La MCU tiene mucha flexibilidad para seleccionar el contenido que debe enviarse a cada uno de los usuarios, pero el algoritmo más usual que se utiliza es el de combinar las dos o tres señales de audio de mayor potencia, sumarlas y enviarlas a todos los usuarios. La gestión de la señal de vídeo se realiza detectando al usuario que está hablando en cada momento (la señal de audio de mayor potencia) y enviando esta información de vídeo a todos los usuarios menos al que está hablando. El usuario que habla recibe la señal de vídeo correspondiente al usuario que había estado hablando anteriormente y recibe las señales de audio de los interlocutores con mayor potencia, exceptuando la suya. El H.243 permite un modo de vídeo en el que se combinan las imágenes de todos los interlocutores con un tamaño reducido.

La opción de encriptado puede realizarse de forma opcional por medio de los estándares H.233 y H.234. El estándar H.233 cubre los procedimientos de encriptado y la selección de los algoritmos. Pueden utilizarse algoritmos normalizados por ISO o algoritmos no estándar. El H.234 cubre los aspectos de intercambios de clave y autenticación. Cuando varios usuarios se conectan en modo multiconferencia, la gestión de la encriptación se realiza desde la MCU, que utiliza claves diferentes para cada usuario.

También están definidos protocolos de control de cámaras remotas con los estándares H.224 y H.281. Permiten controlar la posición de la cámara, que debe estar situada sobre un soporte motorizado de modo que se puedan gobernar a distancia las opera-

ciones de Pan, Tilt y Zoom (motor de panorámica, motor de inclinación y motor de la distancia focal).

Finalmente, está definido un estándar opcional denominado H.331 *Broadcast* que permite modificar el H.320 para realizar transmisiones de *broadcast*, es decir, un único interlocutor que transmite simultáneamente a varios usuarios que están conectados. Se trata de un servicio parecido a la televisión digital, con menor calidad, pero útil en entornos corporativos, docentes o políticos para la difusión de mensajes o discursos. El H.331 establece cómo deben negociar los terminales del estándar H.320 en las situaciones en que no existe un canal de retorno. En este caso, todos los usuarios decodifican las señales de audio y vídeo que se transmiten en un único sentido sin necesidad de que intervenga ninguna MCU en la central.

Estándares H.321 y H.322

Estos estándares amplían el uso del H.320 a redes distintas de la RDSI. También ofrecen una intercomunicación e interoperabilidad simple para establecer comunicaciones con usuarios del H.320 a través de RDSI.

El H.321 extiende el uso del estándar H.320 a redes ATM o RDSI de banda ancha, mientras que el H.322 lo extiende al uso en redes IsoEthernet (según el estándar IEEE 802.9 isochronous Ethernet, ISLAN-16T, IsoEthernet), que es la única red de área local que garantiza calidad de servicio y es compatible con canales virtuales RDSI-B. La adaptación del estándar a este tipo de red es prácticamente directa y el propio documento del estándar sólo tiene una extensión de dos páginas (hecho poco habitual en los estándares y recomendaciones de la ISO o la ITU).

Estándares de segunda generación

El impacto tecnológico del estándar H.320 a principios de los noventa genera la aparición de diferentes modos extendidos y protocolos para incorporar nuevas funcionalidades para aplicaciones de transmisión y recepción de señales multimedia en redes de datos. No obstante, la constante evolución en tecnologías digitales de compresión de señales audiovisuales y el acceso a diferentes topologías de redes evidencia las limitaciones del H.320 para expandirse a otras aplicaciones para las que no ha sido diseñado.

Entre 1995 y 1996, la ITU introduce nuevos estándares para comunicaciones multimedia en redes de datos cuya intención inicial es sustituir progresivamente al H.320. Actualmente estos estándares contemplan distintas topologías y características de redes de comunicación y se utilizan de forma habitual en un gran número de aplicaciones basadas, sobre todo, en el uso de ordenadores personales de carácter genérico. No obstante, a pesar de las limitaciones originales, el H.320 sigue teniendo cierto im-

pacto y aún no ha sido sustituido completamente por los estándares de nueva generación en algunas instalaciones.

Los nuevos estándares son el H.323 y el H.324, que siguen una numeración secuencial con los de la primera generación, y el H.310, destinado a aplicaciones de banda ancha en redes ATM o RDSI de banda ancha. Los estándares H.323 y H.324 consideran topologías de redes basadas en la conmutación de paquetes (el H.323) y comunicaciones por medio de redes conmutadas de baja velocidad (el H.324).

Los tres estándares de nueva generación pretenden introducir varias mejoras en las principales problemáticas que se habían detectado en el H.320. Estas mejoras pueden resumirse en los siguientes puntos:

- Acelerar el proceso de establecimiento de la comunicación. Los procedimientos de control utilizados para la negociación de los protocolos de transmisión y recepción en el H.320 son lentos e impiden establecer la comunicación en poco tiempo.
- Soporte para múltiples canales de audio y vídeo. El H.320 está limitado a dos canales de audio y dos de vídeo (canales de ida y vuelta). La multiconferencia sólo puede obtenerse mediante las aplicaciones MCU proporcionadas por el operador de la red.
- Asignación de ancho de banda con mayor flexibilidad entre los diferentes canales de la comunicación.
- Mejora de la codificación y compresión de las señales de audio y vídeo.
- Aumento del número de modos, tamaños y frecuencias de refresco de las señales de vídeo.
- Facilitar los mecanismos para incorporar nuevas utilidades y modos de transmisión a los estándares.

Los tres estándares mantienen aspectos comunes como el protocolo de control de la comunicación (protocolo H.245) que facilitan la interoperatividad en distintos tipos de redes. Además, mantienen cierto grado de compatibilidad con el H.320, de forma que pueden establecerse comunicaciones comunes entre terminales de distinta generación.

En el siguiente apartado examinamos las funcionalidades más significativas que proporciona el protocolo de control H.245, común a los tres estándares. Posteriormente describiremos las peculiaridades de cada uno.

Protocolo de control H.245

El H.245 es un protocolo de control para la transmisión de materiales multimedia de gran generalidad que supera las limitaciones del protocolo utilizado previamente en el estándar H.320. Está basado en una arquitectura formada por canales lógicos independientes y con *streams* unidireccionales que se identifican por números enteros que el propio transmisor asigna de forma arbitraria. El sistema puede operar con un máximo de 65.535 canales lógicos.

El protocolo H.245 opera en uno de estos canales lógicos (canal propio), desde el que se gobierna el contenido del resto de los canales establecidos durante la comunicación. El H.245 puede crear nuevos canales, eliminar canales existentes, aumentar o disminuir el número de canales asignados a un *stream* de audio o vídeo, generar mensajes de control del estado de la comunicación, etc.

El protocolo facilita el intercambio de información entre los terminales con el objetivo de mejorar la negociación entre los distintos modos de transmisión-recepción y los diferentes formatos en los que puede realizarse una comunicación. Al aumentar el número de modos extendidos y las posibles compatibilidades con anteriores estándares, la negociación de los protocolos utilizados en la comunicación puede ralentizarse de forma considerable. El H.245 permite que tanto los terminales expongan entre sí los modos que pueden trabajar tanto en el modo de transmisión como en el de recepción (en el H.320 sólo era el receptor el que exponía sus modos de compatibilidad).

Una vez los terminales han intercambiado sus posibilidades, cada terminal puede solicitar el modo en que desea recibir la información. Finalmente, es el transmisor el que decide, a partir de las funcionalidades posibles y los deseos solicitados por el receptor, el modo exacto en el que se realizará la transmisión. Los terminales pueden añadir o eliminar modos de transmisión y recepción durante el transcurso de una comunicación. Esta característica es de gran interés debido a que la mayoría de los sistemas de videoconferencia están implementados en ordenadores de propósito general, de manera que la capacidad de proceso de cada uno de los terminales puede alterarse en el tiempo en función de la actividad de las máquinas.

La información se transmite en canales lógicos que contienen un único *stream* de media (audio, vídeo o datos) en un sentido unidireccional. La comunicación bidireccional puede establecerse utilizando varios canales lógicos que transmiten en ambos sentidos. Cada uno de los canales lógicos aporta información sobre un solo tipo de media. La descomposición de las tramas de información en canales lógicos facilita la introducción de modos escalables en vídeo o audio, en los que las diferentes capas se transmiten en canales independientes. El estándar permite definir varios canales lógicos como canales dependientes, de forma que todos ellos contienen información parcial sobre un mismo *stream* de audio, vídeo o datos.

Codificadores de audio y vídeo en estándares de segunda generación

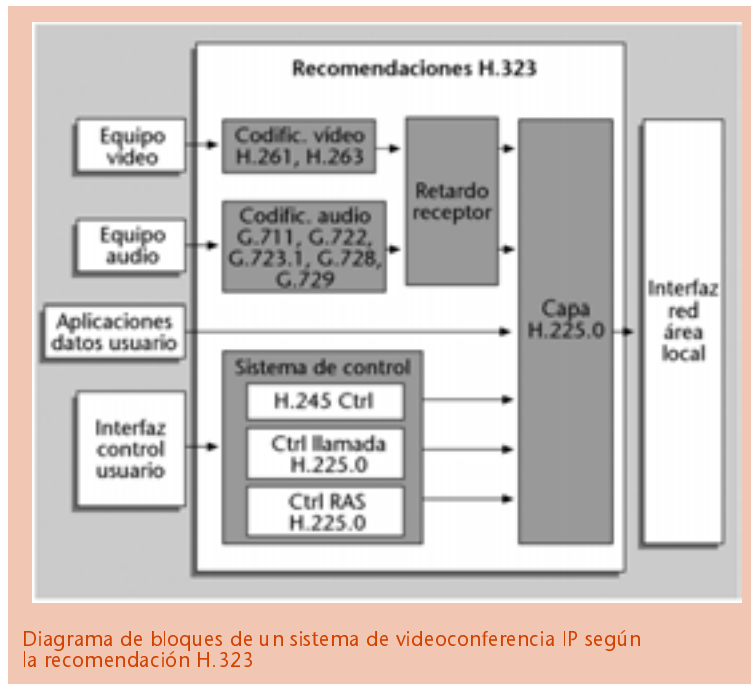
Los codificadores de audio y vídeo utilizados por los estándares H.323, H.324 y H.310 han sido descritos en los apartados anteriores. Aparte de los modos básicos por defecto, los codificadores utilizados más a menudo son el G.723.1 y el G.729 para el caso del audio y el H.263+ para el caso del vídeo.

Redes de conmutación de paquetes. El estándar H.323

Este estándar se utiliza en diferentes redes de conmutación de paquetes que incluyen la Novell IPX/SPX, la Ethernet, la Token Ring y las diferentes redes TCP/IP como intranets corporativas, o Internet. Las aplicaciones de telefonía de voz por medio de IP (telefonía IP) o aplicaciones de gestión de videoconferencia y reuniones en red como Microsoft NetMeeting están basadas en este estándar.

Las redes de conmutación de paquetes se caracterizan por el hecho de que la información se envía en tramas formadas por una cabecera y un determinado volumen de datos que puede cambiar de trama a trama. Las tramas son gestionadas por enrutadores y pueden seguir caminos distintos hacia su destino. Es posible que los paquetes lleguen al destino desordenados (no han seguido el mismo camino) o, incluso, que alguno de ellos se extravíe. Los protocolos de transmisión deben garantizar que aun en estas condiciones puedan reproducirse de forma aceptable las señales de audio y vídeo.

En la siguiente figura se representa un diagrama de bloques típico de un terminal H.323 que incluye los distintos protocolos y estándares de codificación que se utilizan en la transmisión de contenidos multimedia. Los estándares de codificación de audio y vídeo que se definen por defecto son el G.711 y el H.261 respectivamente. No obstante, se utilizan con mayor frecuencia codificadores opcionales como el H.263 o el H.263+ para el vídeo y el G.723.1 o el G.729 para el audio. El protocolo H.245, como corresponde a los estándares de segunda generación, se encarga del control general de la comunicación y el establecimiento de los canales lógicos. Cuando se abre un canal lógico, el receptor devuelve una dirección de red y un número de puerto que constituyen una dirección de red a la que se enviarán todos los paquetes asociados a este canal lógico. El estándar H.225.0 es el encargado de descomponer la información de audio y vídeo en paquetes y de gestionar el inicio de la llamada. Generalmente, el flujo de datos asociado a una comunicación de audio y vídeo se sitúa entre los 100 kbps y los 500 kbps, mientras que el resto de la red se deja disponible para otras aplicaciones. La parte de multiplexación de datos del estándar H.323 es la misma red, ya que se establecen distintos canales lógicos entre el emisor y receptor, que se transforman en forma de paquetes y se envía la información de vídeo o audio en paquetes distintos (canales lógicos distintos).



Inicio de llamada y descomposición en paquetes: H.225.0

El H.225.0 se encarga de gestionar el inicio de la llamada utilizando un subconjunto reducido de mensajes y procedimientos que ya habían sido definidos previamente para sistemas de telefonía RDSI. Los procedimientos implementan las funciones de marcaje, señal de ocupado, etc. Las llamadas pueden iniciarse con la dirección IP del usuario al que se llama o, de forma más común, con alias como las direcciones de correo electrónico que finalmente serán mapeadas a direcciones IP por la propia aplicación de videoconferencia.

Los flujos de datos de audio y vídeo son divididos en paquetes según el protocolo de transporte *Real Time Transport Protocol* (RTP/RTCP), que está especialmente diseñado para el envío de este tipo de información multimedia por medio de Internet. Este tipo de protocolo usa canales no confiables de forma que el receptor nunca solicita la retransmisión de mensajes cuando se produce una pérdida del paquete. El protocolo más usual para el transporte de paquetes en Internet es el TCP, que comprueba la recepción de todos los paquetes y solicita la retransmisión de los que han llegado con errores o los que no se han recibido. Al utilizar un protocolo no confiable para la señal de vídeo o audio, se garantiza que no se producen retardos significativos durante la transmisión debidos al reenvío de información. No obstante, los receptores deben ser capaces de tratar parte de los paquetes perdidos mediante métodos de ocultación de errores u otras técnicas. Los paquetes de audio y vídeo incorporan marcas de índice de tiempo para permitir que el receptor pueda sincronizar los datos de audio y vídeo.

Los mensajes de establecimiento de la comunicación, el canal de control (H.245) y los posibles intercambios de datos asociados a la comunicación (pizarra o documentos en NetMeeting) se descomponen en paquetes que se transmiten de forma fiable me-

dante el protocolo de transporte TCP. Con ello se garantiza la correcta recepción de todos los paquetes sin que se produzcan pérdidas o desordenamientos. Veremos algunos detalles adicionales sobre los protocolos de transporte y las técnicas de descomposición en paquetes de la información de audio y vídeo en la etapa siguiente.

Los principales problemas con los que pueden encontrarse los equipos de H.323 son que la red tiene una velocidad lenta o que los retardos ocasionados por la congestión son muy elevados.

Fracionamiento de *streams* multimedia en paquetes

Uno de los principales problemas que presentan los protocolos de transporte utilizados en redes de conmutación de paquetes es la sobrecarga de datos debido a la cabecera que incluyen todos los paquetes. La cabecera proporciona información de control para los enrutadores y terminales e identifica las direcciones de procedencia y destino de los paquetes. Por lo general, un paquete de H.323 tiene una cabecera con una longitud situada entre 40 y 54 *bytes*. Esto significa que si deseamos que el rendimiento en la transmisión de los datos sea adecuada, debemos incorporar un volumen de datos considerable para que la relación entre *bytes* asignados a la información útil y *bytes* totales del paquete (información útil + cabecera) esté próxima a la unidad. En general, esta filosofía no representa ningún problema cuando la red tiene velocidades altas (por ejemplo, 1 Mbps) debido a que los paquetes se envían de forma rápida y los retardos no son considerables.

Sin embargo, a velocidades de transmisión pequeñas (módems), los paquetes de gran tamaño pueden bloquear el envío de otros paquetes que contienen información multimedia (por ejemplo, señal de audio) y cuyo envío es urgente. En efecto, al tener una velocidad lenta, el tiempo de transmisión o recepción del paquete puede llegar a varios milisegundos, lo que puede producir atascos importantes en la salida de los paquetes. En una sesión de videoconferencia pueden transmitirse simultáneamente paquetes de vídeo, audio y datos de usuario de forma que los mensajes de audio (que son los más urgentes) pueden tener que estar esperando fracciones de segundo hasta que se envíe completamente un paquete de vídeo o de datos.

Esta situación no es especialmente importante en comunicaciones multimedia unidireccionales como la multidifusión de materiales, debido a que el receptor puede construir *buffers* de memoria que permiten recuperar flujos continuos en la salida. Sin embargo, en comunicaciones bidireccionales como la videoconferencia, los *buffers* de memoria deben tener las menores dimensiones posibles para que los interlocutores tengan la sensación de que la comunicación se realiza en tiempo real. Los retardos en la recepción de paquetes de audio pueden introducir discontinuidades y dificultar la inteligibilidad de los mensajes. Actualmente se están investigando técnicas para comprimir la información de cabecera de los paquetes, pero todavía no existe ninguna alternativa operativa.

Las aplicaciones de telefonía IP, que se basan en el estándar H.323, solucionan el problema utilizando el G.723.1 como compresor de audio. Este compresor puede trabajar a velocidades de transmisión muy lentas (6,4 kbps) con una calidad aceptable. Los paquetes se descomponen en fragmentos muy pequeños para que tengan poca latencia a través de la red. La eficiencia de los paquetes es muy baja debido a que la cabecera representa un volumen de información significativo cuando se compara con el tamaño del paquete. No obstante, estas aplicaciones limitan la transmisión exclusivamente a paquetes de audio, de forma que incluso los módems más lentos (33.600 bps, V.34) pueden soportar el exceso de datos asociado a las cabeceras de los paquetes sin que se produzcan retardos importantes.

Limitaciones del H.323

Una de las principales limitaciones del H.323 es originada por la congestión y el retardo en la entrega de los paquetes que puede producirse en redes de conmutación de paquetes, especialmente en Internet. No obstante, no existe ningún otro estándar alternativo para materiales audiovisuales en Internet, y el problema radica en la propia filosofía de la red que no fue pensada originalmente para el intercambio de información en tiempo real.

Aparte de que algunos paquetes pueden ser eliminados de la red debido a problemas de congestión, los protocolos de transporte no garantizan retardos mínimos en la entrega, por lo que algunos paquetes también pueden llegar con retardos considerables con respecto al orden previsto.

El problema del retardo en la recepción de paquetes intenta solventarse mediante *buffers* de memoria en los receptores. Estos *buffers* permiten aceptar paquetes que han llegado con retardos significativos debido a que retrasan cierto periodo de tiempo la reproducción de los mensajes. Es habitual utilizar retardos de prácticamente un segundo para intentar capturar el máximo número de paquetes retrasados. Estos retardos en la reproducción dificultan de forma importante las comunicaciones bidireccionales en tiempo real.

Sin embargo, si el retardo de alguno de los paquetes es superior al tiempo del *buffer*, debe realizarse la reproducción sin utilizar la información de estos paquetes retardados. Este hecho puede degradar considerablemente la imagen o el audio que se reproduce. En la actualidad se están desarrollando estrategias de codificación que mantengan cierta calidad de imagen o de audio cuando se produce la pérdida o el retardo fuera de límites de paquetes. No obstante, la solución es difícil mientras la red no pueda garantizar una cierta calidad de servicio mínima.

Algunas redes locales e intranets, si disponen de cierta calidad de servicio (Ethernet), utilizan protocolos de reserva para garantizar flujos constantes en comunicaciones multimedia por medio de redes de conmutación de paquetes.

Redes de conmutación de circuitos. El estándar H.324

El H.324 fue el primer estándar de segunda generación aprobado y aborda una problemática muy distinta del H.323. En este caso, se trata de establecer comunicaciones mediante redes telefónicas analógicas convencionales (POTS) cuya principal restricción es que la tasa de transferencia de bits que puede obtenerse a través de un módem es muy baja. El H.324 también se ha extendido para considerar transmisiones multimedia por medio de redes móviles o incluso RDSI de banda estrecha (se supone que el H.324 sustituirá progresivamente al H.320 manteniendo la compatibilidad con terminales existentes).

La característica de todas estas redes es que existe una conexión punto a punto entre los dos terminales que realizan la transmisión de material audiovisual mediante una llamada. La conexión es directa y no existen algoritmos de enrutamiento, circunstancia que puede asegurar una tasa de bits fija y constante a lo largo del tiempo. La información de audio, vídeo, datos y control debe ir multiplexada en una trama de bits que identifica cómo se comparte el canal para la transmisión de todos estos elementos.

El H.324 es un estándar que proporciona un conjunto de técnicas para la compresión de audio y vídeo, la transmisión de datos y el control de la comunicación que los desarrolladores pueden elegir en función de las características de su aplicación. En la figura siguiente se proporciona un diagrama de bloques simplificado de los diferentes elementos básicos que forman parte del estándar.

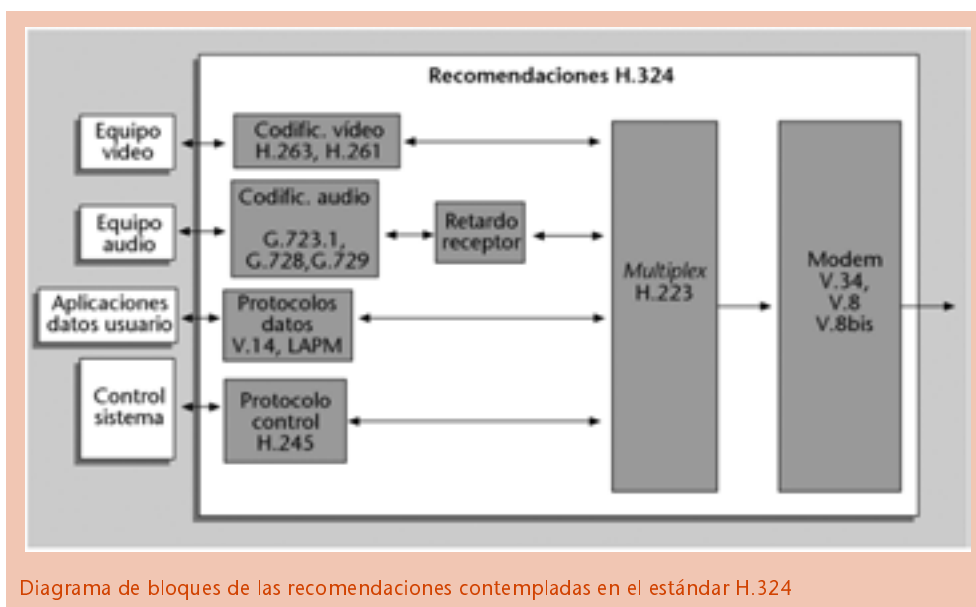


Diagrama de bloques de las recomendaciones contempladas en el estándar H.324

El sistema puede soportar la transmisión de audio, vídeo o datos combinada. Los únicos elementos impuestos por el estándar son el protocolo de control H.245, el sistema de multiplexación de la información H.223 y el módem V.34 (permite un flujo máximo de hasta 33.600 bps). La información de audio, vídeo o datos es opcional y

se permite que en una comunicación puedan enviarse varios canales de un mismo tipo. Se permiten también varios tipos de terminales siendo los más habituales los sistemas de videoconferencia gestionados por ordenadores personales. También son usuales los teléfonos con encriptación de voz, las cámaras de seguridad y los video-teléfonos, como aplicaciones que se consideran dentro del estándar.

El H.324 es el sucesor directo del H.320 y mantiene su arquitectura básica y distintos componentes ya utilizados previamente como el sistema de encriptación, los procedimientos para el control remoto de cámaras y la gestión de multiconferencias por medio de MCU gestionados directamente por la central. Las mejoras más significativas están en el sistema de multiplexación de la información H.223, el protocolo de control H.245 y los nuevos codificadores de audio y vídeo que pueden trabajar con mayores factores de compresión. El codificador de audio básico es el G.723.1 que admite velocidades de 5,3-6,4 kbps y de forma opcional, el G.729 con 8 kbps y retardos muy reducidos. La calidad en la señal de audio se puede conseguir por medio del G.16K, que requiere tasas de 16 Kbps. El compresor de vídeo es el H.263 (en sus dos versiones), que fue especialmente desarrollado para su incorporación en el H.324.

ATM y RDSI de banda ancha.

El estándar H.310

El estándar H.310 define un sistema de videoconferencia por medio de RDSI de banda ancha. La RDSI de banda ancha está basada en el modo transferencia asíncrona (ATM) y permite obtener velocidades de transmisión del orden de varios cientos de Mbps. Este ancho de banda ofrece modos de vídeo y audio de alta calidad, comparables a los sistemas de televisión. Como el resto de los estándares de segunda generación, el H.310 utiliza el H.245 como protocolo de control y se ha diseñado para mantener la compatibilidad con el resto de los estándares (con modos de baja velocidad). El H.310 incluye también un modo de compatibilidad con el H.321 para poder operar con ambos tipos de terminales.

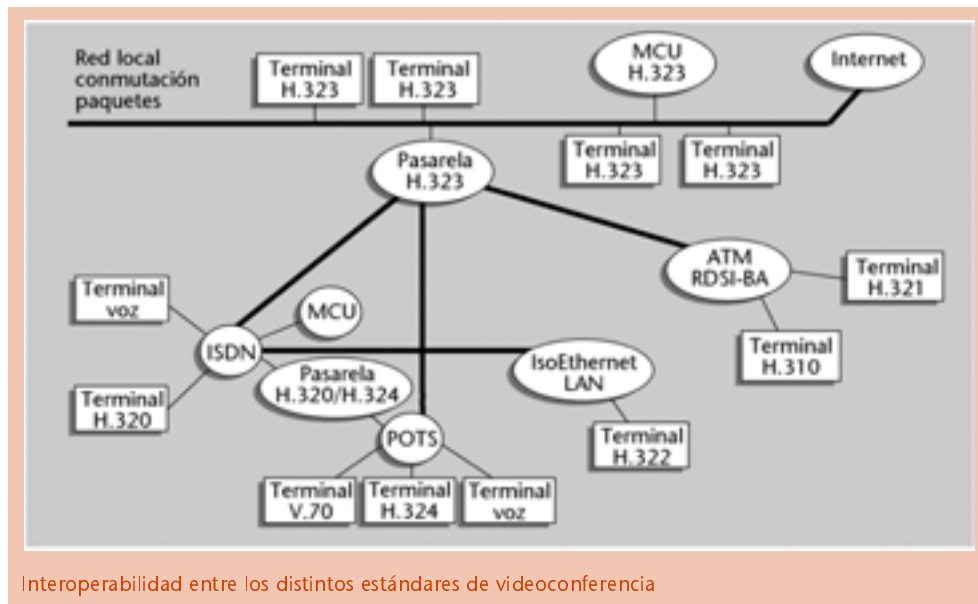
La codificación de audio se realiza utilizando técnicas de codificación perceptual extraídas del MPEG-1. Estos métodos de compresión de audio son los que se utilizan en la radiodifusión de señales de televisión digital y proporcionan una calidad comparable a los sistemas de audio de alta fidelidad. La compresión de vídeo se realiza utilizando el estándar MPEG-2 (*Main Profile & Main Level*). El estándar exige soportar las tasas de transmisión de 6,144 Mbps y 9.216 Mbps para señales de calidad media y alta. Las conexiones multiconferencia también se contemplan por medio de MCU.

En principio, la red digital de servicios integrados de banda ancha debería convertirse en la futura red de telecomunicaciones global y pública, lo que permitiría la distribución de materiales multimedia de gran calidad por medio de servicios de tarificación por volumen de datos.

Interoperabilidad entre los estándares de la serie H.3xx

Todos los estándares de la serie H.3xx han sido diseñados teniendo en cuenta que puedan establecerse comunicaciones entre terminales que trabajan en distintos tipos de redes. Para ello se definen interfaces denominadas pasarelas, para pasar de un formato a otro y mantener cierto grado de compatibilidad. Los modos básicos de operación de cada estándar están pensados para que las pasarelas puedan reducir su carga computacional y, en muchos casos, no se dé la necesidad de recomprimir las señales de audio o vídeo a nuevos formatos.

En la siguiente figura se muestra un diagrama con varios sistemas interconectados y se representan las pasarelas que permiten el paso de un estándar a otro.



Interoperabilidad entre los distintos estándares de videoconferencia

Etapa 2: Descarga y reproducción de vídeo en Internet

Introducción

Estadísticas recientes indican que la visualización de archivos de vídeo y audio desde conexiones remotas a Internet para acceder a la difusión en la red de materiales multimedia es cada vez más habitual. Se estima que a principios del año 2001 cada semana se difundían por medio de Internet unas 350.000 horas con contenidos en directo sobre acontecimientos deportivos, culturales, musicales, etc. Además, existe una gran cantidad de material multimedia (cientos de miles de horas) prerregistrado que está disponible en la red para su descarga y visualización (o audición) en tiempo real. En el futuro, sobre todo si el ancho de banda de acceso a Internet aumenta, se prevé una fuerte expansión de estas tecnologías en el sector del entretenimiento.

La introducción de materiales de vídeo y audio en los ordenadores personales se introduce a finales de la década de los ochenta con la aparición de los ficheros de audio y vídeo entrelazado (principalmente, formatos AVI y MOV) y los reproductores multimedia que eran aplicaciones especialmente diseñadas para la reproducción de estos ficheros. Posteriormente, con la aparición del formato MPEG-1, los ficheros de vídeo y audio pueden ser registrados en un CD-ROM y reproducirse con una calidad aceptable.

El impacto de Internet a principios de los noventa supone la proliferación de archivos multimedia distribuidos en multitud de servidores. Los potenciales usuarios pueden descargar estos ficheros en su ordenador y visualizarlos después mediante un reproductor multimedia convencional. No obstante, los ficheros son de gran tamaño y, a pesar de que los compresores de vídeo y audio presentan cada vez mejores prestaciones, el tiempo de descarga suele resultar excesivo. Además, dado que el material no puede visualizarse hasta su descarga completa, muchas veces la calidad final de los contenidos es decepcionante. En la práctica, debido a las bajas velocidades de Internet, sólo resulta posible descargar ficheros de corta duración y pobre calidad.

El concepto de descarga y reproducción simultánea de los archivos empieza a tomar sentido desde estos primeros inicios. El proceso se suele conocer con el nombre de *streaming* y permite ir decodificando los datos de audio y vídeo a medida que se van recibiendo. En este texto utilizaremos esta palabra anglosajona para indicar la descarga y reproducción simultánea de archivos multimedia. El *streaming* no resulta un problema trivial debido a que la estructura de Internet como red de conmutación de paquetes que utiliza protocolos confiables no está pensada para el intercambio de archivos en tiempo real. En esta sección se estudiarán los conceptos básicos del *streaming* de vídeo y audio y se revisarán las principales tecnologías de *streaming* utilizadas actualmente.

Primeras arquitecturas de streaming

VivoActive 1.0. La primera aplicación de streaming

El primer sistema de *streaming* multimedia fue propuesto por el programa VivoActive1.0 y utilizaba servidores web convencionales, basados en el protocolo de aplicación HTTP (servidores convencionales) para distribuir contenidos multimedia.

La ventaja de utilizar un servidor web convencional es evidente debido a que el proveedor del material no requiere utilizar aplicaciones especiales que gestionen el *streaming* de vídeo. Sin embargo, los servidores web están basados en el *HyperText Transfer Protocol* (HTTP) que a su vez se apoya sobre el protocolo de transporte TCP (*Transfer Control Protocol*). Este protocolo es el que mayoritariamente se utiliza en aplicaciones mediante Internet.

El protocolo de transporte TCP y el streaming multimedia

El principal problema del protocolo TCP, en el contexto de transmisión de señales audiovisuales es su confiabilidad. Esto significa que aunque los paquetes de red lleguen desordenados, el protocolo de transporte ha añadido suficiente información en la cabecera como para poder reordenarlos y entregarlos a la aplicación en el mismo orden en que han sido transmitidos. También es posible identificar paquetes que se han recibido con errores, ya que el protocolo incluye un código de detección de errores. Cuando se detecta la presencia de errores o se advierte que uno de los paquetes no se recibe, se solicita automáticamente el reenvío de estos paquetes al transmisor. Estas características en el diseño del protocolo TCP representan una gran ventaja para garantizar la transferencia de archivos y documentación a través de Internet. No obstante, desde el punto de vista de la transferencia y reproducción simultánea de archivos multimedia, el protocolo TCP representa un serio problema. En efecto, la gestión de los paquetes a través de la red implica que algunos de ellos se recibirán con un retardo considerable, sobre todo aquellos que deban ser retransmitidos.

Buffering y rebuffering

Para solucionar este problema, el VivoActive1.0 utilizaba un *buffer* temporal de memoria en el que se almacenaban los paquetes recibidos. El *buffer* podía almacenar información audiovisual equivalente a un periodo de tiempo situado entre cinco y veinte segundos. Cuando este *buffer* estaba lleno podía empezarse la reproducción. En condiciones normales, la reproducción puede realizarse sin interrupciones debido a que pueden tolerarse ciertos retardos en la recepción de los paquetes, ya que no se necesitan hasta el momento en que tengan que ser reproducidos. El tamaño del *buffer* puede ser estimado inicialmente a partir de las estadísticas de paquetes perdidos y tiempo de retransmisión que obtenga el receptor (el servidor también puede utilizar esta operación). Generalmente, estas estimaciones suelen resultar suficientes para

que el usuario pueda visionar el contenido con un flujo continuo, con el único inconveniente de que al inicio de la operación debe esperarse unos segundos. La misma estrategia puede utilizarse tanto para visionar materiales según demanda (vídeo bajo demanda) o en directo (en este caso, el material que se reproduce tiene un cierto retardo sobre el acontecimiento real).

No obstante, la arquitectura de Internet no garantiza una calidad mínima y las condiciones de tráfico pueden variar durante la recepción del programa. Esta circunstancia puede provocar que algunos paquetes se retrasen más que el tiempo predeterminado en el *buffer*, con lo que la reproducción se interrumpe. La aplicación debe esperar que el *buffer* vuelva a llenarse para volver a empezar la reproducción. Este problema se conoce con el nombre de *rebuffering*; suele resultar decepcionante y se produce con más frecuencia de lo previsto.

Formatos genéricos y específicos de streaming por medio de TCP

En principio, los ficheros multimedia de que dispone el servidor pueden estar en cualquier formato de compresión. No obstante, al transmitir la información debe garantizarse que la parte de audio y de vídeo está correctamente entrelazada para que el receptor disponga simultáneamente de ambas informaciones. Es usual incorporar marcas de tiempo a los fragmentos de audio y vídeo para que el reproductor pueda sincronizar ambas pistas. Los formatos AVI y MOV originales no garantizan que el entrelazado del fichero sea el adecuado para su reproducción en tiempo real cuando se usa una transmisión basada en una fragmentación en paquetes por medio de TCP. Simplemente garantizan que existe algún tipo de entrelazado de los canales y que se puede acceder a los datos de forma rápida y fácil cuando están almacenados en un fichero que se encuentra en el mismo ordenador en el que se realiza la reproducción.

Esta circunstancia implica que si el servidor quiere seguir utilizando los formatos de fichero convencionales AVI o MOV, deberá analizar previamente la información contenida en estos ficheros, reordenarla, introducir firmas temporales, en algunos casos recomprimirla y descomponerla en paquetes antes de transmitirla. Dado que no existe ningún estándar que especifique cómo tienen que realizarse estas operaciones, cada aplicación puede utilizar métodos propios (propietarios). El problema de esta estrategia es que para reproducir los contenidos es necesario que el cliente disponga del *software* adecuado, compatible con el procedimiento utilizado en el servidor. La compatibilidad puede obtenerse utilizando un reproductor específico (aplicación autónoma) o, como es más habitual, descargando las librerías dinámicas (*dll's*) adecuadas para que el reproductor multimedia del sistema operativo pueda interpretarlas correctamente.

Otra alternativa es que el servidor almacene los ficheros multimedia en un formato específico y propio adecuado a su transmisión por Internet. De esta forma, las operaciones de reordenación de los datos, firmas temporales, descomposición en paquetes, etc. ya han sido previamente realizadas de manera que el servidor no requiere realizar todas estas tareas cada vez que se conecta un cliente. Con ello puede reducirse la

carga computacional del servidor, permitiendo que pueda aumentar el número de clientes a los que se proporciona servicio.

Los formatos específicos más popularizados son el RealAudio (RA) de RealNetworks para la difusión en web de señales de audio, el RealVideo (RV) también de RealNetworks y el Advanced Streaming Format (ASF) de Microsoft. El principal inconveniente de utilizar estos formatos propietarios sigue siendo que requieren reproductores propios (RealNetworks) o la descarga de librerías dinámicas para que puedan ser tratados por el reproductor multimedia del sistema operativo.

La solución utilizada por VivoActive 1.0 fue una combinación entre los estándares H.263 para la codificación de vídeo y el G.723 para la de audio que estaban multiplexados mediante un protocolo propio que combinaba ambos contenidos en un único fichero.

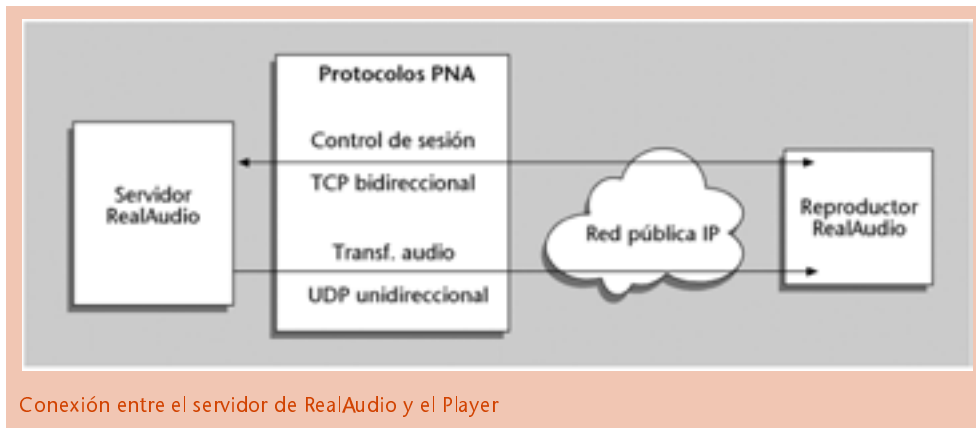
El formato ASF

El formato ASF de Microsoft tiene cierto interés debido a que permite reutilizar contenidos previamente codificados en ficheros AVI. Microsoft proporcionó una aplicación denominada AVI2ASF que permite pasar los ficheros AVI al nuevo formato ASF. También se proporcionaron librerías dinámicas que realizaban esta función. Estas librerías dinámicas han sido utilizadas por otros desarrolladores para que sus aplicaciones realicen esta transformación de formatos sin necesidad de que el usuario deba utilizar aplicaciones externas. Un ejemplo del uso de las librerías dinámicas de Microsoft para realizar estas conversiones es Adobe Premier.

El formato ASF define un multiplexado propio de la información de audio y vídeo, la descomposición en paquetes, y realiza distintas transformaciones de los datos de audio y de vídeo para adaptarlo a las condiciones de transmisión por Internet. Estas transformaciones tienen en cuenta la protección de los datos audiovisuales ante eventuales pérdidas de paquetes. La filosofía general de las estrategias de protección de los datos son parecidas a las que se utilizan en el RealAudio 1.0 y en el RealVideo 1.0, cuyo desarrollo fue anterior al ASF y que analizamos con cierto detalle en el siguiente apartado.

RealAudio 1.0

El RealAudio 1.0 puede considerarse como el primer paquete completo para *streaming* de audio e introduce distintos conceptos nuevos que merecen ser analizados con detalle. Esta aplicación tuvo un éxito rotundo en sus inicios y fue utilizada por muchas emisoras de radio en Internet. Actualmente, la *suite* completa de RealNetworks es uno de los productos más utilizados para el *streaming* de vídeo y audio en Internet. Sus competidores directos son las Windows Media Technologies de Microsoft y la *suite* de *streaming* de Apple.



En la figura se representa la arquitectura simplificada de la *suite* de RealAudio 1.0, que se conoce como la *Progressive Networks Architecture* (PNA). La idea básica es establecer dos enlaces virtuales entre el servidor de RealAudio y el reproductor. Éste último se ejecuta en el cliente y debe ser descargado e instalado previamente. Uno de los enlaces se utiliza para controlar la sesión y utiliza el protocolo TCP convencional. El tráfico a través de este enlace es reducido y se limita a aspectos básicos como el inicio de sesión, solicitudes de paquetes perdidos y breves informes sobre el proceso de transmisión-recepción.

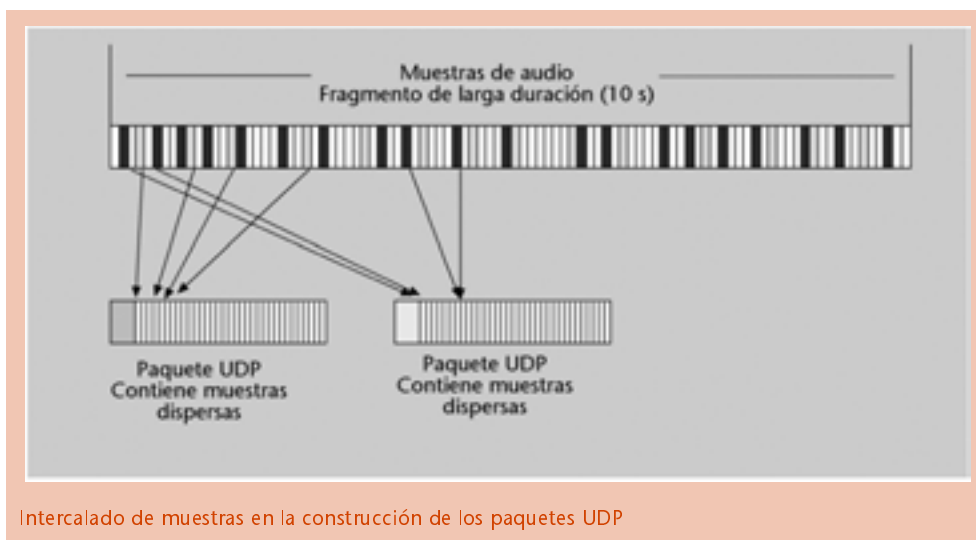
La transferencia de la información de audio se realiza mediante un enlace unidireccional (del servidor al cliente) que utiliza el protocolo de transporte UDP (*User Datagram Protocol*). Este protocolo de transporte es una alternativa a TCP y se utiliza en algunas aplicaciones. La característica básica de UDP es que es no confiable; en otras palabras, el protocolo no contempla ningún mecanismo de control para que todos los paquetes se entreguen correctamente. Igual que en el caso de TCP, los paquetes también pueden llegar desordenados e, incluso, para el UDP, no recibirse nunca. Aunque parezca contradictorio, esta estrategia facilita considerablemente la transmisión de contenidos audiovisuales. Discutiremos las ventajas del UDP frente al TCP en secciones posteriores.

Los paquetes UDP de audio que se envían desde el servidor al emisor se pueden extravíar. RealAudio proporciona ciertas protecciones ante este tipo de problemas.

- **Solicitud de retransmisión.** La aplicación del reproductor comprueba constantemente si existen paquetes que no se han recibido y puede solicitar su retransmisión al servidor. La solicitud de retransmisión no se realiza de forma sistemática como en el protocolo TCP, sino que se realiza en el nivel de la aplicación. El reproductor puede estimar diferentes parámetros de la comunicación con el servidor. Uno de estos parámetros es el tiempo que tarda en recibir los paquetes que solicita que sean reenviados. Cuando se observa que falta un paquete que contiene información que debe ser reproducida en un tiempo T , comprueba si éste tiene posibilidades de llegar a tiempo si se solicitase su retransmisión. En el caso en que se estime que hay tiempo suficiente, se solicitará la retransmisión. En caso contrario, deberá procederse a la reproducción de audio sin la contribución de dicho paquete. La diferencia fundamental con el protocolo TCP es que los paquetes sólo

se solicitan cuando se estima que aún pueden llegar a tiempo, lo que reduce el tráfico innecesario en la red.

- **Protección frente a pérdidas de paquetes.** La pérdida de un paquete puede resultar muy grave si no se usaran técnicas de protección ante esta eventualidad. En efecto, un paquete contiene varios *bytes* de información útil (en muchos casos, en torno a 1.000 *bytes*). El volumen de información incluido en un paquete puede corresponder aproximadamente a entre 0,25 y 0,5 segundos de audio. Si la pérdida del paquete afecta a toda esta información, se producirían discontinuidades en la señal de audio que resultarían molestas, y parte del mensaje no podría ser interpretado correctamente. La solución a este problema consiste en dispersar la información de audio que se envía en cada paquete utilizando una intercalación de muestras. La idea básica se ilustra en la figura adjunta, en la que se indica cómo un fragmento de audio de diez segundos se reparte en varios paquetes de manera que cada paquete contiene muestras de todo el fragmento de diez segundos. El receptor debe recomponer la información en el orden original una vez dispone de todos los paquetes. Si se produce la pérdida de un paquete, las muestras extraviadas estarán dispersas en todo el fragmento de diez segundos, por lo que el reproductor puede interpolar (aproximar el valor de una muestra extraviada como el promedio entre la anterior y la posterior) los datos perdidos sin que esto se aprecie excesivamente.



Transmisión de vídeo por medio de *streaming*

La generalización de los conceptos utilizados para el *streaming* de audio al *streaming* de vídeo no es directa. Existen las siguientes diferencias fundamentales que deben resolverse:

- **Tasa de bits variable.** Los compresores de audio proporcionan generalmente un tasa de codificación constante a lo largo de la trama. Esto significa que el factor de compresión se mantiene constante tanto si se considera un fragmento de audio muy corto como si se considera un fragmento largo. En cambio, la señal de

vídeo tiene generalmente factores de compresión que dependen del contenido. Cuando existe mucho movimiento en la imagen se requieren más bits que cuando existe menos. Cuando hay un cambio de plano, el nuevo fotograma requiere mucha información adicional, etc. La solución al problema consiste en suponer que el receptor dispone de un *buffer* de memoria que no puede sobrepasarse ni vaciarse por completo. Las imágenes que requieren mayor información se codifican con un mayor número de bits y las que tienen mucha redundancia pueden codificarse con menor número de bits. Esta estrategia coincide con la utilizada en la mayoría de los codificadores de vídeo (MPEG-1, MPEG-2, H.263, etc). Es importante que el valor del número de bits medio al considerar varias imágenes se mantenga constante, lo cual puede obtenerse ajustando el paso de cuantificación del compresor (detalle con el que se aproxima la imagen).

En la práctica, podemos enviar los datos al decodificador con una tasa de bits constante. La información se pondrá en el *buffer* de recepción y las imágenes se irán extrayendo a medida que vaya progresando la reproducción. Cuando se reproduce una imagen con mucha información binaria se entra mucha información del *buffer* y cuando se reproduce una imagen con poca información se extrae poca cantidad. Lo único importante es que siempre exista información en el *buffer* y que nunca se desborde.

- **Acceso aleatorio a la información de vídeo.** En aplicaciones de vídeo bajo demanda, el reproductor debe poder disponer de controles de avance rápido, avance hacia atrás, parada de imagen, volver al inicio, rebobinar, etc. Estos controles permiten obtener la misma funcionalidad que la de un reproductor de vídeo doméstico, pero representan una problemática especial para el flujo de *streaming* de la señal de vídeo. La solución a este problema está en disponer de algunos *frames* de vídeo que se codifican de forma independiente y que permiten realizar un acceso aleatorio a la información. La filosofía sigue siendo común con los aspectos ya estudiados para los estándares DSM (*Digital Storage Media*) y de *broadcasting* del MPEG. Así, cuando el usuario desea que se rebobine la cinta una determinada cantidad, la aplicación del reproductor debe determinar si dispone de la información solicitada en el propio terminal del cliente. Si no es así, solicita la retransmisión del *frame* de referencia que desea al servidor a través del canal virtual bidireccional adicional al canal de *streaming*.
- **Protección de la información de vídeo.** Las pérdidas de paquetes son más difíciles de gestionar en el caso de vídeo que en el caso del audio debido a que los *frames* de vídeo suelen requerir mucha más información. La técnica de dispersar los datos de los *frames* en varios paquetes no suele resultar eficiente, ya que un único *frame* ya requiere de varios paquetes para su transmisión. Las técnicas más utilizadas son la inserción de códigos redundantes que permitan recuperar algunos errores en la recepción de los paquetes. No obstante, si el paquete no se recibe se pierden los datos y los códigos de protección, por lo que la utilidad de esta técnica está restringida a cierto tipo de errores. Otra técnica que se está utilizando recientemente (*Source Forward Error Correction* SFEC) es enviar en un mismo pa-

quete los datos correspondientes a un grupo de bloques de imagen con buena precisión y los mismos datos correspondientes al grupo de bloques enviado anteriormente pero con una precisión baja (mala calidad). En el caso de que un paquete resulte extraviado se pueden utilizar los datos de baja calidad que vendrán incorporados en el siguiente paquete. También pueden utilizarse distintas técnicas de ocultación de errores que ya han sido definidas en estándares como el MPEG.

Servidores de *streaming*

Los aspectos analizados en el apartado anterior demuestran que para poder satisfacer las necesidades de acceso en tiempo real a materiales audiovisuales en Internet es aconsejable utilizar soluciones específicas que adecuen el tráfico de la red a las características de este tipo de contenidos. Por este motivo, las soluciones basadas en servidores web convencionales y basadas en los protocolos HTTP y TCP están cada vez más en desuso y están siendo sustituidas por los denominados servidores de *streaming* o *suites* de servicios multimedia. Los principales desarrolladores que proporcionan este tipo de aplicaciones son Microsoft, RealNetworks y Apple. Las tres soluciones tienen muchos aspectos en común desde el punto de vista tecnológico, por lo que en este apartado nos limitaremos a comentar los aspectos generales centrándonos en la implementación de las Windows Media Technologies de Microsoft. Puede obtenerse información completa del RealSystem G2 de RealNetworks o del QuickTime de Apple en las webs de estos desarrolladores.

Características básicas de un servidor de *streaming*

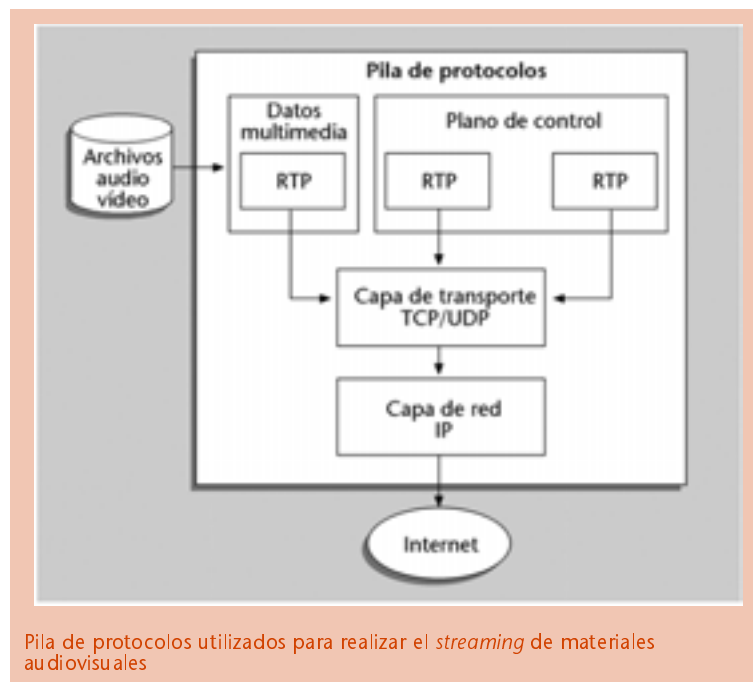
Un servidor de *streaming* está constituido por los siguientes subsistemas:

- **Comunicador.** El comunicador es el subsistema que tiene en cuenta los enlaces virtuales entre el cliente y el servidor y los protocolos que se utilizan en el nivel de aplicación y el nivel de transporte para realizar la transferencia de los archivos y el control de la comunicación. Analizaremos con algo más de detalle los protocolos utilizados en secciones posteriores.
- **Sistema operativo.** El sistema operativo sobre el que se construye el servidor de *streaming* debe permitir realizar operaciones de control de grabación y reproducción del material en tiempo real. En todas las *suites* el servidor de *streaming* puede estar instalado en el mismo ordenador que el servidor web, pero debe disponer de un núcleo sobre el ordenador que permita realizar operaciones de registro, envío de vídeo en directo, acceso aleatorio a ficheros, envío de materiales audiovisuales bajo demanda, etc., en tiempo real.
- **Sistema de almacenamiento.** El sistema de almacenamiento para los servidores de *streaming* debe permitir el registro y acceso continuo y sin interrupciones a la información y contenidos audiovisuales.

Protocolos utilizados en servidores de *streaming*

Los protocolos directamente relacionados con el problema del *streaming* de vídeo a través de Internet pueden clasificarse en las siguientes categorías:

- **Protocolos de la capa de red.** El protocolo de la capa de red es el IP (*Internet Protocol*). Éste es el protocolo mediante el cual se gestionan todos los aspectos de enrutamiento de los mensajes hacia el destino.
- **Protocolos de la capa de transporte.** Los dos protocolos principales, sobre los que ya hemos hablado, son UDP y TCP. Estos dos protocolos definen cómo se descompone la información en paquetes y se intercambian datos multimedia e información de control sobre la comunicación entre el servidor y el cliente. Sobre ellos, en aplicaciones de *streaming* es habitual que aparezcan los protocolos RTP (*Real-time Transport Protocol*) y el RTCP (*Real-Time Control Protocol*). En la figura adjunta se representa esquemáticamente cómo está construida la pila de protocolos.
- **Protocolos de la capa de sesión.** Estos protocolos se encargan de definir los mensajes y procedimientos que definen la entrega del material multimedia durante una sesión. Se trata del RTSP (*Real-Time Session Protocol*) y el SIP (*Session Initiation Protocol*). El primero se utiliza en los servidores de RealNetworks G2 y el segundo, en el Media Stream Server de Microsoft. Su incidencia sobre el proceso real de *streaming* FALTA CONTINUAR LA FRASE!!!



Protocolos de transporte: TCP frente a UDP

En los apartados anteriores ya hemos comentado las principales diferencias entre estos dos protocolos de la capa de transporte y hemos visto que aunque el UDP es un

protocolo no confiable (no garantiza la entrega de todos los paquetes), es más adecuado para la transferencia de material audiovisual.

Una característica adicional que también debemos considerar es que el protocolo TCP tiene como objetivo principal maximizar la velocidad de transferencia de datos, de modo que se asegure que la comunicación se establece en un valor próximo al máximo permitido por las condiciones de congestión de la red en cada momento. Para ello utiliza un algoritmo denominado *slow start* que consiste en que el servidor empieza a enviar los paquetes a una velocidad lenta y la va aumentando progresivamente, intentando llegar al límite que permite la red. Cuando el servidor recibe un elevado porcentaje de paquetes perdidos por el cliente reduce la velocidad de transmisión ligeramente. Después de un cierto tiempo vuelve a intentar aumentar la velocidad de transmisión de los paquetes hasta que nuevamente aumenta el porcentaje de paquetes extraviados. Con esta filosofía se puede garantizar que la comunicación es estable al máximo permitido por la red y que se va adaptando en el tiempo al tráfico real.

Esta estrategia puede parecer excelente, ya que siempre sitúa la comunicación al límite de velocidad admisible. No obstante, en aplicaciones de *streaming* no tiene mucho sentido enviar las componentes de audio y vídeo a una velocidad superior a la que se han codificado los contenidos originales. En efecto, si un material multimedia ha sido codificado a 64 kbps no tiene ningún sentido que se envíe a una velocidad superior, ya que no mejorará en ningún momento la calidad de la recepción. El uso del protocolo UDP permite que el servidor controle la tasa de transferencia de los datos enviando únicamente el flujo necesario. Esta estrategia permite que el servidor pueda atender a muchos más clientes y, además, reduce el tráfico en la red evitando que se produzcan pérdidas de paquetes que posteriormente deben retransmitirse. Otra ventaja del UDP es que al adecuar la velocidad asignada a la información audiovisual a sus necesidades reales es posible que el cliente disponga todavía de ancho de banda para conectarse a otros servicios de Internet.

El principal inconveniente del protocolo UDP es que muchos cortafuegos impiden el paso de paquetes UDP por razones de seguridad. En estas circunstancias es imprescindible que el *streaming* se realice a través del protocolo TCP. Generalmente los servidores de *streaming* intentan realizar la transmisión directamente en UDP y, en caso de no ser posible, cambian a modo TCP.

El protocolo RTP

El RTP es un protocolo estándar de Internet diseñado especialmente para soportar aplicaciones en tiempo real. En la práctica se trata de un protocolo construido por encima del protocolo UDP que permite que la aplicación disponga de información de si se han producido pérdidas de paquetes en la transferencia del archivo. Los datos de la aplicación primero son tratados por el protocolo TCP, insertando las cabeceras e identificaciones del material y posteriormente son encapsulados en paquetes UDP. Por tanto, los paquetes UDP contienen información en la que ya se ha insertado pre-

viamente el protocolo RTP. En el caso de que la transmisión del flujo multimedia se realice mediante TCP, los paquetes TCP siguen transportando información previamente procesada por RTP. Las principales funcionalidades que proporciona este protocolo son:

- Firmas de tiempo. Incluye marcas de tiempo en los paquetes para poder sincronizar los diferentes medios que componen una presentación audiovisual.
- Números de secuencia. Introduce números de secuencia en los paquetes para que la aplicación en el cliente pueda reordenar los paquetes recibidos y advertir la pérdida de algunos paquetes.
- Identificación de la carga útil. Las cabeceras del protocolo RTP contienen información sobre el tipo de media que contiene cada paquete. Existen unos identificadores predefinidos para indicar los tipos de media más habituales: MPEG-1 audio, MPEG-2 vídeo, etc.
- Identificación de la fuente (*Synchronization Source Identifier*, SSRC). Se proporciona identificación de la fuente que ha generado el material audiovisual.

El protocolo RTCP

El protocolo RTCP es un protocolo de control que trabaja de forma conjunta con el RTP. Generalmente, los parámetros de control se intercambian entre el servidor y el cliente en un enlace virtual distinto del enlace por el que se transmiten los datos (recordad el esquema de dos enlaces virtuales creado inicialmente en el RealAudio 1.0). La idea básica es que por medio de este enlace servidor y cliente puedan intercambiar información sobre el estado de la comunicación y la calidad de la misma. Los mensajes e informaciones proporcionadas por RTCP se transmiten utilizando TCP como protocolo de transporte. Los servicios que proporciona RTCP pueden resumirse en:

- Realimentación de la calidad de servicio (QoS, *Quality of Service*). Es la principal función de RTCP. La calidad de servicio puede controlarse con informes periódicos (*reports*) enviados desde el servidor al cliente y desde el cliente al servidor. Los informes contienen información sobre los siguientes parámetros: 1) Fracción de los paquetes RTP perdidos desde el último informe; 2) Número acumulado de paquetes perdidos desde el inicio de la recepción; 3) Diferencias entre tiempos de llegada de los paquetes; 4) Retardo desde la emisión del último informe. Esta información de control es útil para el servidor, el receptor y sistemas de monitorización externos. Basándose en estos informes, el servidor puede modificar la velocidad a la que se envían los paquetes para adaptarse a las características de la comunicación.
- Identificación de participantes. Se proporciona información de tipo texto (nombre de usuario, número de teléfono, *e-mail*) asociada al identificador SSRC proporcionado por el protocolo RTP.

- Control del número de paquetes de control. En conexiones en las que participan varios clientes, los mensajes de control de la calidad de recepción pueden resultar excesivos. El protocolo establece que el número de paquetes de control se mantendrá siempre por debajo de un 5% del ancho de banda total de la sesión. Un 25% de estos informes será enviado desde el servidor a los clientes y el 75%, desde los clientes al servidor. Para evitar que no se envíen mensajes de control, cada participante debe enviar un mínimo de un informe cada cinco segundos.
- Sincronización InterMedia. Se proporciona una referencia de la correspondencia entre el tiempo real y las firmas temporales introducidas por el protocolo RTP. Esta referencia puede utilizarse para sincronizar materiales procedentes de distintas fuentes.

Control y ajuste de la velocidad de transferencia

Los informes de calidad de servicio transmitidos desde los clientes al servidor puede aprovecharse para ajustar las características de la transmisión. Si un cliente reporta un elevado porcentaje de paquetes extraviados es posible que se esté saturando alguna parte de la red (probablemente el enlace entre el cliente y su proveedor de servicios de Internet –ISP–, que suele ser la conexión más lenta). En esta circunstancia, el servidor puede reducir la velocidad del flujo de datos y adaptarse a las características del enlace y mejorar la calidad del servicio. Alternativamente, si el porcentaje de paquetes perdidos es muy bajo, es probable que estemos realizando una transmisión muy por debajo de las posibilidades que ofrece la red aumentando el flujo de datos.

Existen tres alternativas básicas para adaptarse a las velocidades de transmisión que acepta el cliente:

- **Múltiples copias del material.** La estrategia más sencilla es generar múltiples copias del material audiovisual con diferentes calidades y velocidades de transmisión. Una copia puede estar adaptada a conexiones de 24 kbps, otra para 36,6 kbps, una a 64 kbps, 128 kbps, 256 kbps, etc. A partir de los informes de calidad de servicio, el servidor puede cambiar la copia que se envía a cada cliente.
- **Transcoding.** Es una estrategia que requiere cierta capacidad de proceso por parte del servidor y que consiste en eliminar *frames* o reducir la calidad de la fuente de vídeo hasta determinar el flujo de datos adecuado que puede aceptar el cliente. El material audiovisual se captura con la máxima calidad y sólo existe una única copia. Inicialmente intenta enviarse con esta calidad. Si se detecta una calidad de servicio baja, el servidor puede cambiar modificar los parámetros de la compresión. Una posibilidad es no enviar al cliente *frames* del tipo B, con lo que se reducirá la frecuencia de imagen que éste recibe. También pueden eliminarse algunos *frames* de tipo P y modificar las referencias de los vectores de movimiento a los *frames* existentes. Otra alternativa es recuantificar los *frames* con un mayor paso de cuantificación para que disminuya el flujo de bits y la calidad del vídeo. En todos estos casos, el servidor debe mantener el material original con una calidad

de codificación alta (para los clientes con ancho de banda) y calcular en tiempo real las reducciones necesarias que pueden requerir los clientes menos afortunados. La capacidad de proceso del servidor debe ser considerable. En el supuesto de que el número de clientes sea potencialmente grande, puede generar de forma automática un conjunto reducido de *transcoders* seleccionando cuál se envía a cada cliente.

- **Codificadores escalables.** Es sin duda la mejor opción y la tendencia actual de la mayor parte de los codificadores para aplicaciones de Internet. El compresor estructura la información en varias capas o niveles que proporcionan distintos grados de calidad en función de las capas que se descodifiquen. La capa base proporciona información de baja calidad (resolución y ruido) y es necesaria para poder descodificar el resto de las capas. Las capas pueden enviarse en distintos paquetes de manera que si un cliente tiene suficiente ancho de banda recibe paquetes de todas las capas, por lo que puede reproducir una calidad de vídeo excelente. En cambio, si un cliente tiene un canal reducido, sólo recibirá la capa base. El servidor tiene que realizar pocos cálculos, ya que el compresor descompone la información en varios niveles. La escalabilidad puede ser temporal (se aumenta el número de *frames* en las capas de calidad), espacial (se aumenta la resolución de la imagen) o en SNR (se reduce el paso de cuantificación, por lo que la imagen se observa con mejor calidad). También pueden usarse compresores con escalabilidad híbridas que combinan algunos de los tipos anteriores. Los estándares ISO MPEG-2 y MPEG-4 y el ITU H.263+ han desarrollado intensos trabajos para definir las sintaxis y los parámetros con los que se puede escalar la información de vídeo y audio.

Es importante observar que la calidad que obtiene el receptor cuando se reduce el flujo de vídeo o audio y no se pierden paquetes es siempre superior a la que se obtiene con un modo de calidad superior en el que se pierden algunos paquetes. En efecto, la pérdida de paquetes suele suponer la aparición de defectos visuales o auditivos muy acusados, por lo que es recomendable reducir la calidad global de la comunicación.

Multicast

Cuando un cliente se conecta para solicitar un material audiovisual bajo demanda, el servidor debe proporcionar un flujo de bits independiente para cada cliente. Esta hipótesis es aceptada en sistemas bajo demanda, aunque limita considerablemente el número de clientes potenciales. No obstante, cuando se transmite material en directo parece más adecuado que el servidor envíe un único *stream* de vídeo y audio al que se suscriben los clientes.

Cuando se utilizan compresores escalables, el servidor puede enviar todas las capas simultáneamente en canales virtuales diferentes. En este caso, será el propio receptor el que se suscriba a más o menos capas en función de la calidad de servicio que esté recibiendo. Si detecta pocos paquetes perdidos, puede suscribirse a una capa escala-

ble adicional. Si la densidad de paquetes perdidos es excesiva, deberá abandonar la suscripción.

Para resolver este problema, inherente a la filosofía original de Internet, se ha propuesto una modificación del protocolo IP denominada IP-Multicast, cuya filosofía general es que el servidor envíe una única copia de los paquetes IP que se distribuyen y replican a través de la red de enrutadores en función del árbol de clientes que están abonados al servicio. Sin embargo, a pesar de que IP-Multicast lleva más de diez años de desarrollo e investigación, su implantación actual se limita a algunas intranets con dominios limitados. Existen problemas importantes relativos a la gestión de la red, soporte para capas con funcionalidades de control de flujo, reenvío de paquetes erróneos, control de la congestión, etc. que dificultan su implantación.

Actualmente, la mayoría de los servidores de *streaming* resuelven el problema de la multidifusión en lo concerniente a la capa de aplicación. Esencialmente, la estrategia consiste en utilizar múltiples servidores intermedios que retransmiten los paquetes recibidos a sus clientes. En este caso, el servidor principal sólo tiene que enviar un flujo de datos a cada uno de estos servidores que se encargan de difundirlo a sus clientes. Uno de los primeros ejemplos de este tipo de red es el Multicast Backbone (Mbone) que se utiliza para aplicaciones de videoconferencia.

En general, los servidores de *streaming* se encargan directamente de gestionar el tipo de conexión más adecuado en cada momento. El Windows Media Server de Microsoft puede gestionar el envío de datos multimedia utilizando los protocolos UDP, IP-Multicast, TCP y HTTP-TCP. La selección de un modo u otro es totalmente transparente para el administrador del servidor y depende de las características de acceso del cliente. El servidor intenta inicialmente transmitir la información utilizando los protocolos óptimos UDP o Multicast. En caso de que no sea posible, se intenta enviar a través de TCP. Finalmente, si los cortafuegos tampoco permiten el tráfico TCP directo, se transmite como tráfico web (puerto 80) con una combinación de los protocolos HTTP-TCP.