

Resumen #6 - (r6)

Esteban Ignacio Durán Vargas - 2020388144

IC - 7602 - 2023 I Semestre

7.4 Multimedia

7.4.2 Compresión de audio

MPEG es el más popular, mp3 es el más poderoso y es la parte de audio de compresión de MPEG.

Se puede codificar por forma de onda donde se transforma la frecuencia mediante Fourier y la amplitud del componente se codifica de forma mínima, con menos bits. La otra forma es codificación perceptual, donde se aprovechan fallas del sistema auditivo humano para codificar una señal que suene igual pero se ve diferente en un osciloscopio (psicoacústica). Esta enmascara la frecuencia de modo que tiene un sonido fuerte en una banda de frecuencia para ocultar otro más suave. El oído reduce su ganancia cuando el sonido fuerte empieza y toma tiempo infinito en volver a aumentar (enmascaramiento temporal). Al mantener un registro de cuáles señales están siendo enmascaradas por señales más poderosas de bandas de frecuencia cercanas, podemos omitir más y más frecuencias en la señal codificada, lo que ahorra bits. Si una señal poderosa se detiene en una banda de frecuencia, con sus propiedades de enmascaramientos se pueden seguir omitiendo frecuencias enmascaradas por algún intervalo de tiempo mientras el oído se recupera. MP3 transforma con Fourier para obtener la potencia de cada frecuencia y transmitir sólo frecuencias no enmascaradas, usando menos bits.

La compresión de audio se realiza muestreando la forma de onda de 32,44.1 o 48 kHz. Puede ser: monofónica (1 flujo de entrada), monofónica dual, estéreo separado y estéreo unido (cada canal separado o unido).

Se elige la tasa de bits de salida. Depende de la relación señal a ruido del audio. Si se escogen más bajas se pierde aún más calidad. Luego las muestras se procesan en grupo de 1152, cada uno se pasa por 32 filtros para obtener 32 bandas de frecuencia, con la entrada en un modelo psicoacústico para determinar las frecuencias enmascaradas. Luego cada banda se transforma para tener una resolución espectral más fina. Luego, los bits disponibles se dividen entre bandas, la mayoría a aquellas con mayor potencia no enmascarada, a las de poca potencia menos bits, las enmascaradas no se les asignan bits. Finalmente, los bits se codifican mediante la codificación de Huffman, con códigos cortos a números que aparecen frecuentes y largo a los que no.

7.4.3 Compresión de audio

Escuchar por internet o música bajo demanda. El proceso general es que se establece un TCP con el server, se envía un GET con la canción, se obtiene el disco y regresa al navegador. El navegador investiga el tipo para desplegarlo, escribiéndolo en un archivo de trabajo para que el reproductor de medios obtenga y reproduzca la música bloque por bloque. El problema es que la canción completa se transmite en la red antes de que comience, mientras se descarga. Por eso existen metaarchivos, que solo nombra a la música, que se escribe en el disco, inicia la reproducción con una aplicación auxiliar y entrega el nombre. El reproductor lo lee y ve el url, que contacta al servidor y pide la canción. El servidor es especializado RTSP (protocolo de retransmisión en tiempo real). El reproductor de medios trata errores, utiliza un protocolo RTP donde los paquetes pueden perderse. A veces se intercala mediante una interpolación de las muestras anteriores y posteriores. También el

reproductor descomprime la música y elimina la fluctuación con un bufer de unos segundos antes de reproducir que continúa llenando a la tasa que el reproductor lo vacía. Esto se hace con un servidor pull siempre que haya un búfer para otro bloque para llenarlo como adicional, pero tiene muchas solicitudes innecesarias. También está el servidor push donde el reproductor manda un PLAY y el servidor continúa enviándole datos. El servidor puede tener la velocidad normal o más rápida. En ambos se almacenan datos en el búfer antes de la reproducción. A velocidad normal los datos se agregan al final del búfer y el reproductor los elimina al frente y mantiene datos constantes en el búfer. Si lo envía a más velocidad no se garantiza que el servidor se ejecute a una tasa regular y se puede reponer si queda atrás, pero el búfer se puede desbordar.

La solución es que haya una marca de agua baja y una alta en el búfer, donde solo se envían datos hasta que el búfer llegue a la marca de agua alta, donde se le indica al reproductor que haga una pausa para que el servidor se vacíe y que cuando llegue a la baja continúe. El reproductor tiene un control remoto RTSP para el servidor push.

7.4.4 Radio en Internet

Hay 2 métodos para esto, los programas se graban con anterioridad y se almacenan en disco, los escuchas se conectan a los archivos de la estación de radio y obtienen y descargan el programa para escuchar. Igual que audio de flujo continuo. También se puede almacenar cada programa justo después de transmitirlo en vivo, pero atrasado. El otro método es difundir el contenido en vivo por internet. Puede ser simultánea con la versión en vivo. Es como audio de flujo continuo pero con diferencias. Es igual almacenar en el búfer del usuario, 10-15 seg antes de empezar, y se evita la fluctuación. No transmite a una tasa mayor y la detiene con la marca de agua, sino siempre se difunde a la misma que la que se reproduce. Otra diferencia es que el audio de flujo continuo es punto a punto y radio en Internet es por TCP con la estación, pero esto detiene el flujo cuando la ventana está llena o los paquetes se pierden. Se usa porque pocos ISP soportan multidifusión, porque las estaciones no conocen RTP y para saltarse firewalls.

WMA, RealAudio y MP3 compiten para ser el formato dominante. Existen páginas web con botones para escuchar que utilizan metaarchivos. El navegador utiliza un MIME para determinar el mejor programa, se escribe el archivo en disco e inicia la reproducción. Luego visita la dirección del archivo con http y contacta al servidor para comenzar a actuar como radio. Audio solo tiene un flujo por lo que sirve con http, video necesita rtsp.

Cualquiera puede empezar una radio, con una PC ordinaria como tarjeta de sonido y un micrófono, y un software reproductor de medios para la captura de audio junto a un codec con la salida seleccionada como MP3. El flujo de audio generado por la estación se alimenta a través de Internet hacia un servidor grande y lo distribuye a una gran cantidad de conexiones TCP. Este servidor soporta varias estaciones pequeñas y tienen directorios con cada una para que los escuchas seleccionen y obtengan la alimentación TCP. Icecast maneja todas las piezas y hay servidores que manejan la distribución con un cambio de cuota.

7.4.5 Voz sobre IP

Como consecuencia de que el tráfico de datos es mayor que el tráfico de voz, los operadores de redes de conmutación se interesaron por transportar voz por datos, que requería poca banda adicional y además podía generarles más dinero

- **H.323**

El ITU emitió esta recomendación "Sistemas telefónicos visuales y equipo para redes de área local que proporcionan una calidad de servicio no garantizada". Fue la primera base de los sistemas de telefonía de Internet. Es una revisión arquitectónica. Se refiere a protocolos específicos de codificación de voz, llamadas, señalización, transporte de datos... En el centro se encuentra la puerta de enlace que conecta Internet con la red telefónica. Habla H.323 a Internet y PSTN a teléfono. Los dispositivos se llaman terminales. Una LAN puede tener un gatekeeper que controla los puntos finales en una zona.

Una red telefónica necesita protocolos. Para codificar y de condificador se usa G.711, codifica un canal de voz muestreando 8000 veces por segundo con una muestra de 8 bits y descomprime a 64 kbps. Se permiten otros de compresión como G.723.1, toma un bloque de 240 muestras y codifica predictivo para reducirlo a 24 o 20 bytes. Proporciona una tasa de salida ya sea de 6.4 o 5.4 con poca pérdida.

H.245 negocia cuáles algoritmos se usan, al igual que la tasa de bits. RTCP controla canales RTP. Se necesita un protocolo para establecer y liberar conexiones, proporcionar tonos de marcado y emitirlos, etc; ese es Q.931. Para hablar con el gatekeeper se utiliza H.225. Del PC al gatekeeper se conoce como canal RAS (Registro/admisión/estado). Permite terminales para unirse y dejar la zona, solicitar y regresar ancho de banda, entre otras cosas. RTP se usa para transmitir.

Primero se le da a la PC un UDP para buscar al gatekeeper, esta se aprende la IP y envía un RAS para registrarse. Cuando se le proporciona ancho de banda puede hacer llamadas; pide antes para evitar sobrescribir la línea saliente para proporcionar calidad. La PC establece TCP con el gatekeeper para una llamada con protocolos de red telefónicos. La PC puede enviar un mensaje SETUP de Q.931 para identificar al teléfono; el gatekeeper responde con un CALL PROCEEDING de Q.931 para confirmar la recepción de la solicitud, luego envía un mensaje setup a la puerta de enlace (mitad computador, mitad conmutador de teléfono) realiza una llamada al teléfono ordinario. La oficina central hace que suene el teléfono y regresa un ALERT (Q.931) para indicar a la PC que ya se emitió el sonido. Cuando la persona en el otro extremo levanta el teléfono la oficina manda CONNECT (Q.931) para indicar a la PC que hay una conexión. Los paquetes ignoran al gatekeeper y van a la dirección IP de la puerta de enlace. Solo es una conexión de capa física para mover bits. Se usa H.245 para negociar los parámetros de la llamada y usa su canal que siempre está abierto. Cada lado anuncia sus capacidades, se establecen canales unidireccionales y a cada uno se le asigna un codec, que pueden ser diferentes. Cuando terminan las negociaciones empieza el flujo con RTP. RTCP controla el congestionamiento, maneja sincronización audio/vídeo. Cuando alguna parte cuelga se utiliza Q.931 para terminar la conexión. La PC contacta al gatekeeper para liberar el ancho de banda.

El QoS es bueno si la conexión es estable, sin fluctuación entre la PC y la puerta de enlace.

- **SIP—Protocolo de Inicio de Sesión**

Una forma más simple y modular, definida en RFC 3261. Describe llamadas telefónicas a Internet, videoconferencias y otras conexiones multimedia. Es un solo módulo pero interactúa bien con otras aplicaciones de Internet. Puede establecer sesiones de dos partes, múltiples (conferencia) y multidifusión (un hablante, muchos receptores). Puede contener audio, vídeo o datos. SIP solo maneja establecimiento, manejo, terminación de sesiones. Se utilizan otros protocolos como RTP/RTCP. SIP es un protocolo de capa de aplicación y puede ser UDP o TCP. Soporta una variedad de servicios como localizar, invocar, determinar sus capacidades, manejar mecanismos de establecimiento/terminación de llamada. Los números de teléfono son URL y utiliza el esquema sip (sip:ilse@cs.university.edu) para un usuario de nombre Ilse en un host del DNS de esa universidad. Pueden tener direcciones IPv4, IPv6 o número telefónicos reales. Se basa en texto, modelado en HTTP. Envía el mensaje por ASCII con el nombre del método (INVITE, ACK, BYE, OPTIONS,

CANCEL, REGISTER) con líneas extra para parámetros. Muchos se toman de MIME para que interactúe con otras aplicaciones de Internet.

Para establecer una sesión, crea una sesión, el invocador crea una conexión TCP con el invocado y envía un INVITE, sus encabezados describen la estructura del cuerpo del mensaje que contiene capacidades, tipos de medios y formatos. Si acepta la llamada responde con un código HTTP. El invocado puede proporcionar igual sus capacidades, tipos de medios y formatos. La conexión se realiza en tres vías, con un ACK para terminar el protocolo y confirma con un mensaje 200. Cualquier parte solicita terminación con BYE. OPTIONS consulta a una máquina sobre sus propias capacidades, se utiliza antes de un inicio de sesión. REGISTER es para que el SIP pueda rastrear y conectarse con el usuario que esté lejos de casa; se envía a un servidor de ubicación SIP que mantiene la pista de quién está en qué lugar. Ese servidor se usa para encontrar al usuario.

- **SIP—Protocolo de Inicio de Sesión**

Ambos permiten llamadas de dos partes y múltiples partes utilizando las computadoras y los teléfonos como los puntos finales. Ambos soportan negociación de parámetros, codificación y RTP/RTCP.

H.323 es un estándar típico de la industria telefónica que especifica toda la pila de protocolos y define lo que está permitido y prohibido. Es complejo y rígido y difícil de adaptar.

SIP intercambia líneas ASCII, es un módulo de carga ligera que interactúa con otros protocolos de internet pero no con sistemas telefónicos. Debido a que IETF es modular, es flexible y se puede adaptar. Peroo puede tener problemas de interoperabilidad.