7.1.4 Audio and Video Compression

Before audio and video can be transmitted over a computer network, it must be digitized and compressed. The need for digitization is obvious: computer networks transmit bits, so all transmitted information must be represented as a sequence of bits. Compression is important because uncompressed audio and video consume a tremendous amount of storage and bandwidth—removing the inherent redundancies with compression in digitized audio and video signals can reduce the amount of data that needs to be stored and transmitted by orders of magnitude. As an example, a

Approach	Unit of allocation	Guarantee	Deployment to date	Complexity	Mechanisms
Making the best of best-effort service	none	none, or soft	everywhere	minimal	application-layer support, CDN, over-provisioning
Differential QoS	classes of flows	none, or soft	some	medium	policing, scheduling
Guaranteed QoS	individual flows	soft or hard, once a flow is admitted	little	high	policing, scheduling, call admission and signaling

Table 7.1 ♦ Three approaches to supporting multimedia applications

single image consisting of 1024 pixels, with each pixel encoded into 24 bits (8 bits each for the colors red, green, and blue), requires 3 Mbytes of storage without compression. It would take seven minutes to send this image over a 64 kbps link. If the image is compressed at a modest 10:1 compression ratio, the storage requirement is reduced to 300 Kbytes and the transmission time also drops by a factor of 10.

The topics of audio and video compression are vast. They have been active areas of research for more than 50 years, and there are now literally hundreds of popular techniques and standards for both audio and video compression. Many universities offer entire courses on audio compression and on video compression. We therefore provide here only a brief and high-level introduction to the subject.

Audio Compression in the Internet

A continuously varying analog audio signal (which could emanate from speech or music) is normally converted to a digital signal as follows:

- The analog audio signal is first sampled at some fixed rate, for example, at 8,000 samples per second. The value of each sample is an arbitrary real number.
- Each of the samples is then rounded to one of a finite number of values. This operation is referred to as **quantization**. The number of finite values—called quantization values—is typically a power of two, for example, 256 quantization values.
- Each of the quantization values is represented by a fixed number of bits. For example, if there are 256 quantization values, then each value—and hence each sample—is represented by 1 byte. Each of the samples is converted to its bit representation. The bit representations of all the samples are concatenated together to form the digital representation of the signal.

As an example, if an analog audio signal is sampled at 8,000 samples per second and each sample is quantized and represented by 8 bits, then the resulting digital signal will have a rate of 64,000 bits per second. This digital signal can then be converted back—that is, decoded—to an analog signal for playback. However, the decoded analog signal is typically different from the original audio signal. By increasing the sampling rate and the number of quantization values, the decoded signal can approximate the original analog signal. Thus, there is a clear trade-off between the quality of the decoded signal and the storage and bandwidth requirements of the digital signal.

The basic encoding technique that we just described is called **pulse code modulation (PCM)**. Speech encoding often uses PCM, with a sampling rate of 8,000 samples per second and 8 bits per sample, giving a rate of 64 kbps. The audio compact disk (CD) also uses PCM, with a sampling rate of 44,100 samples per second with 16 bits per sample; this gives a rate of 705.6 kbps for mono and 1.411 Mbps for stereo.

A bit rate of 1.411 Mbps for stereo music exceeds most access rates, and even 64 kbps for speech exceeds the access rate for a dial-up modem user. For these reasons, PCM-encoded speech and music are rarely used in the Internet. Instead, compression techniques are used to reduce the bit rates of the stream. Popular compression techniques for speech include **GSM** (13 kbps), **G.729** (8 kbps), **G.723.3** (both 6.4 and 5.3 kbps), and a large number of proprietary techniques. A popular compression technique for near CD-quality stereo music is **MPEG 1 layer 3**, more commonly known as **MP3**. MP3 encoders typically compress to rates of 96 kbps, 128 kbps, and 160 kbps, and produce very little sound degradation. When an MP3 file is broken up into pieces, each piece is still playable. This headerless file format allows MP3 music files to be streamed across the Internet (assuming the playback bit rate and speed of the Internet connection are compatible). The MP3 compression standard is complex, using psychoacoustic masking, redundancy reduction, and bit reservoir buffering.

Video Compression in the Internet

A video is a sequence of images, typically being displayed at a constant rate—for example, at 24 or 30 images per second. An uncompressed, digitally encoded image consists of an array of pixels, with each pixel encoded into a number of bits to represent luminance and color. There are two types of redundancy in video, both of which can be exploited for compression. Spatial redundancy is the redundancy within a given image. For example, an image that consists of mostly white space can be efficiently compressed. Temporal redundancy reflects repetition from image to subsequent image. If, for example, an image and the subsequent image are exactly the same, there is no reason to re-encode the subsequent image; it is more efficient simply to indicate during encoding that the subsequent image is exactly the same.

The MPEG compression standards are among the most popular compression techniques. These include **MPEG 1** for CD-ROM-quality video (1.5 Mbps), **MPEG 2** for high-quality **DVD** video (3–6 Mbps), and **MPEG 4** for object-oriented video compression. The MPEG standard draws heavily on the JPEG standard for image compression by exploiting temporal redundancy across images in addition to the spatial redundancy exploited by JPEG. The **H.261** video compression standards are also very popular in the Internet. In addition there are numerous proprietary schemes, including Apple's QuickTime and Real Networks' encoders.

Readers interested in learning more about audio and video encoding are encouraged to see [Rao 1996] and [Solari 1997]. A good book on multimedia networking in general is [Crowcroft 1999].

CASE HISTORY

STREAMING STORED AUDIO AND VIDEO: FROM REALNETWORKS TO YOUTUBE

RealNetworks, a pioneer in audio and video streaming, was the first company to bring Internet audio to the mainstream. Its initial product—the RealAudio system released in 1995—included an audio encoder, an audio server, and an audio player. Allowing users to browse, select, and stream audio content from the Internet on demand, it quickly became a popular distribution system for providers of entertainment, educational, and news content.

Today audio and video streaming are among the most popular services in the Internet. Not only is there is a plethora of companies offering streamed content, but there is also a myriad of different server, player, and protocol technologies being employed. A few interesting examples (as of 2007) include:

- **Rhapsody from RealNetworks:** Provides streaming and downloading subscription services to users. Rhapsody uses its own proprietary client, which retrieves songs from its proprietary server over HTTP. As a song arrives over HTTP, it is played out through the Rhapsody client. Access to downloaded content is restricted through a Digital Rights Management (DRM) system.
- **MSN Video:** Users stream a variety of content, including international news and music video clips. Video is played through the popular Windows Media Player (WMP), which is available in almost all Windows hosts. Communication between WMP and the Microsoft servers is done with the proprietary MMS (Microsoft Media Server) protocol, which typically attempts to stream content over RTSP/RTP; if that fails because of firewalls, it attempts to retrieve content over HTTP.
- **Muze:** Provides an audio sample service to retailers, such as BestBuy and Yahoo. Music samples selected at these retailer sites actually come from Muze, and are streamed through WMP. Muze, Rhapsody, YouTube, and many other streaming content providers use content distribution networks (CDNs) to distribute their content, as discussed in Section 7.3.
- **YouTube:** The immensely popular video-sharing service uses a Flash-based client (embedded in the Web page). Communication between the client and the YouTube servers is done over HTTP.

What is in store for the future? Today most of the streaming video content is low-quality, encoded at rates of 500 kbps or less. Video quality will certainly improve as broadband and fiber-to-the-home Internet access become more pervasive. And very possibly our handheld music players will no longer store music—instead we'll get it all, on-demand, from wireless channels!