

**DAT330 – Principles of Digital Audio**  
**Cogswell Polytechnical College**  
**Spring 2009**

**Week 12 – Class Notes**

**Internet Audio**

**Networks and File Transfers**

Computer networks are typically asynchronous, transmit data in discrete packets, and can interconnect many disparate devices. Successful communication across a network calls for arbitration to avoid usage conflicts. Networks are not concerned with the type of data being transmitted, and uses a common file structure for all data types such as e-mail, graphics, audio, and video. In addition, data delivery is not continuous, and often not in real-time. For example, delivery depends on network data rate as well as on current network traffic; transfer speeds can occur well below or above real-time. In some applications, a portion of the network bandwidth can be reserved to allow continuous real-time multimedia exchanges, such as video conferencing. Network communications are also limited by file format compatibility.

A network can link separated computers such that the separation is transparent to the user. Files can be downloaded from the central storage connected to a file server, processed in a local workstation, then conveyed to a distant end user. Using real-time methods, two recording studios can be bi-directionally linked so that overdub data can smoothly flow between them.

**Telephone Services**

Current digital telephony systems can route the user's voice (as a string of digits) through copper wire, fiber, and satellite to another user in a distant part of the world. Digital transmission systems can convey high bandwidth digital audio data without much signal attenuation and noise interference. Because phone calls are treated as data, the sound quality of a local call and a long-distance call is basically the same.

The audio frequency response for a typical phone system is 300-3400 kHz, with speech coded at 8-bit PCM (amplitude companding is used to improve the dynamic range) and a sampling frequency of 8 kHz to yield a bit rate of 64 kbps.

T-carrier networks are designed for long-distance transmission, over copper lines, optical fiber, microwave radio, and coaxial cables to convey digital audio, video, and data.

T-1 serial digital communication circuits provide a dedicated point-to-point bandwidth of 1.544 Mbps; the signal conveyed is called DS-1. At each end of the T-1 line there is a customer service unit (CSU). The CSU interfaces data from the customer premises equipment (CPE) or a data service unit (DSU) and encodes it for transmission along the T-1; a multiplexer or local area network (LAN) bridge could be used.

A T-1 circuit is capable of carrying CD data (1.41 Mbps) without data reduction, 24 voice channels, or a single video program with data reduction. T-1 comprises 24 subchannels (DSOs or slots) each carrying 64 kbps (8-bit, 8 kHz sampling rate); everything from data to voice traffic runs through these subchannels. Data bytes are applied to frames, with each frame holding 24 DSO's plus one framing bit. Thus, a frame holds 193 bits ( $24 \times 8 + 1$ ) and the frame

rate of 8 kHz yields a 1.544 Mbps overall rate.

Multiple applications, such as audio, video, and data, can share one T-1 line, with individual channels assigned to one or multiplexed DSO's as needed. T-1 is full duplex (bi-directional) and the assignments need not be identical.

T-1 lines are very reliable and have a BER of  $10^{-9}$ . Individual DSO's can be sent to different destinations with the digital access cross-connect system.

Other services include T-0, T-1C, T-2, T-3, and T-4 lines.

## **ISDN**

Integrated Services Digital Network (ISDN) is an advanced dial-up telephone service with full duplex operation. ISDN provides a digital connection between the consumer and the telephone exchange, and ultimately to long-haul digital transmission systems. An overall rate of 144 kbps is supported. Although ISDN uses existing telephone lines, specialized equipment is needed at the send and receive ends.

Basic rate ISDN, intended for home use, uses a copper wire to provide two 64 kbps circuits to send and receive audio or other information, and one 16 kbps circuit for dialing and other signaling functions.

ISDN is primarily offered to business customers, using coaxial or fiber cables, can provide more than 23 64 kbps channels totaling 1.544 Mbps of bandwidth.

Numerical data and compressed video signals sent over ISDN can achieve speeds exceeding that of dial-up modems. Timecode can be conveyed through ISDN systems.

Simple telephones can be used in voice-only ISDN hookups, but in order to exchange audio data the user supplies an AD/DA converter, data reduction codec, and ISDN hookup. The data sent between the user and the local telephone exchange is sent over copper wires; lines are terminated at a dedicated terminal adapter (TA) interface. The copper wire pair connecting to the user is called the user interface (UI) and a network termination (NT-1) converts this to a four-wire ST interface. The outgoing data is converted to a telecommunications format where it is directed over a carrier to the receiving party, which must have corresponding ISDN services and equipment.

## **Asymmetric Digital Subscriber Lines (ADSL)**

Provides high capacity data channels, along with regular telephone service, on a single pair of copper wires. Like ISDN, ADSL needs two modems, but unlike ISDN, ADSL uses existing wires and existing phone numbers. Under optimal conditions ADSL can deliver 8 Mbps into the home, however ADSL does not have an evenly balanced data rate. The rate out of the homes might be considerably less, some 640 kbps.

ADSL competes with cable modems, and both have their pros and cons. Both systems are asymmetric.

CATV is designed to be unidirectional in data transmission, and suffers from the third-party syndrome which means that the more users log into a common line (shared by multiple homes), the slower the traffic will be.

ADSL is not as fast as cable modem hookups, but ADSL users enjoy a private line to the central switching computer. Speed is dependent on the distance away from the switching computer (top speeds are possible only for connecting distances less than 12,000 ft). Rates slow down over longer distances, and above 18,000 ft service may not be possible. ADSL (nor cable modems) does not provide sufficient bi-directional bandwidth for high-quality videoconference.

## Computer Networks

Networks are designed to interconnect multiple devices as networked nodes. For example, computers with network interface cards, each with a unique address can form nodes on a common network. Data is sent when a path is available, at the speed determined by the network interface. Both file exchange and random-access functions among computers are permitted. One node may transmit multiple audio channels, or a video channel; bandwidth of 30-40 Mbps might be needed. Typical office LAN's will not suffice, but multimedia networks can offer data reduction via perceptual coding.

A network must provide fast storing and loading of projects for editing and archiving, easily accessible directories, remote operation of systems elements, background back-up of files, communication among users, and multiple user access to one project as well as single user access to multiple projects.

Networks can be configured in a number of ways:

1. Bus configuration places nodes along a serial bus.
2. Ring configuration places nodes along a closed circuit.
3. Star configuration gives each node direct access to the central controller.

Star configurations are preferred because a disabled node does not affect other node performance. A central concentrator unit is needed to monitor and direct bus traffic in a star configuration. When the maximum number of nodes are placed on a network, or the system's longest length is reached, it can be extended with a repeater, a device that receives signals, then resynchronizes and retransmits them. A bridge isolates network segments so that data is only transmitted across the bridge when its destination is another segment. A router is a computer with two network interfaces; it can pass data between different types of networks and can optimize the routing for faster communication.

It is the job of a network to break a message into data packets of uniform size and code them with a destination address and a header that describes where each packet fits within the message; the packets are transmitted, and received where they are assembled into the message. If a packet is corrupted with errors it can be quickly retransmitted. Because the arrival speed of packets cannot be guaranteed, real-time transfer is not always possible.

All networks must define rules for access to physical storage and terminal devices; this limits the bandwidth of the network. Two common control methods are token passing and collision detection.

Token passing occurs when a node finishes a transmission and sends a token (bit pattern) to the next node; a node can transmit only when it has the token. Each token ring has both an input and output, so that connections are passed from one node to another, forming a ring.

With collision detection, if a collision occurs, priority is assigned, and the data is retransmitted.

In client-server networks, most applications, data, and the network operating system, are centrally placed on a common file server. In peer-to-peer networks; data is kept locally under the control of individual computers.

For audio and multimedia files, distributed storage is more efficient, in which nodes act as peers, each serving as both client and server, each with local storage. A distributed network with modest bandwidth (10-15 Mbps) might be adequate for many applications. However, when multiple users are on the network, data transfer speed might be limited.

Wide area networks (WAN) connect multiple stations beyond a single building, reaching up to global distances. WAN networks can use an ISDN interconnection. LAN distributes data through an office, building, or campus, and uses FDDI, CDDI, or ATM protocols.

### **Ethernet**

Computer network used to connect personal computers, printers, disk drives, and other equipment over coaxial cable, twisted pair, and optical fiber. Ethernet uses asynchronous transmission and collision detection. The following bit rates are often used:

Ethernet – 10 Mbps

Fast Ethernet – 100 Mbps

Gigabit Ethernet – 1000 Mbps

10 Gigabit Ethernet – 10,000 Mbps

Because of the technical overhead, Ethernet is not used as a studio backbone to carry audio/video files. In many cases, a Storage Area Network (SAN) using Host Bus Adapter (HBA) cards and 1000BaseT or Fibre-Channel fiber optic cable, is a better choice for moving large, multiple files among many storage devices.

### **Asynchronous Transfer Mode (ATM)**

High-bandwidth network standard using low-delay switching, variable bandwidth, and multiplexing to provide flexible and efficient communications. Audio, video, and other data can be simultaneously delivered at a rate of 1 Gbps or more. ATM is used primarily in WAN and LAN with many users.

Traditional networks pool many channels of information, with each receiver picking its information from the stream; considerable routing data is needed, and this increases overhead. ATM architectures use switching and multiplexing to form a temporary dedicated, virtual channel within the transmission path bandwidth; the virtual channel is allocated a sufficient data rate, providing bandwidth on demand. Data packets or cells carry the name of the channel with much less routing data required. In addition, many ATM users can share a channel through multiplexing, yet maintain a fixed-time relationship between data cells. In this manner, slower moving data can be combined with fast data, thus assuring a continuous data flow. ATM is also referred to as Broadband ISDN.

The AES47 standard describes the transmission of audio over ATM channels. Linear PCM samples can be inserted into ATM cells in a number of ways including time order, channel order, or in multichannel groups. Data in the AES3 format can be conveyed, and there is a provision that carries up to 60 audio channels.

### **MediaNet**

Dedicated high-speed multimedia LAN network. Allows multiple users simultaneous access to materials on central disk drives. MediaNet is implemented on CDDI and FDDI protocols, and supports the Apple Filing Protocol, Networked File System, and ATM. With CDDI, the network can simultaneously handle multiple channels of compressed audio and video with throughput of 24 Mbps from node to node. With ATM, transfer rates of 120 Mbps can be achieved. Computers can be interfaced as either servers or clients. Disk drives are local in node computers, and each node can access data on any other node hard drive. Token rings are used for data traffic control. By using these types of networks, multimedia data can be directly manipulated remotely without copying files from place to place.

## **Wireless Networking, WiFi, and Bluetooth**

Wireless technologies are becoming more ubiquitous for many applications, such as cellular telephony. Both WiFi and Bluetooth can be used to convey non-audio and audio data over short distances, between computers, peripherals, and audio playback devices.

WPAN - is a wireless personal area network, operating over "personal" distances.

WLAN – is a wireless area network operating over longer distances. A hot spot can encompass a home, office, or school.

WMAN – wireless metropolitan area network operating over a large area.

WWAN – wireless wide area network can potentially have a global reach. For example, cell phones and General Packet Radio Services (GPRS).

Wireless networks are generally low power and rely on techniques such as frequency hopping spread spectrum (FHSS) to help prevent interference from other broadcast channels, as well as other microwave devices.

WiFi is used for WLAN applications and has two widely used standards: IEEE 802.11a and 802.11b; the former operates at a 5 GHz and the latter operates at 2.4 GHz. An 802.11b system might provide a theoretical maximum bit rate of 11 Mbps over a 300 m range, while an 802.11a system might provide a but rate of 54 Mbps over 50 m. Although actual user data throughput might be considerably less.

WiFi is also known as wireless Ethernet because it shares some parts of its specification with Ethernet. WiFi uses carrier detection and collision avoidance, as well as authentication and encryption.

Bluetooth is used for WPAN applications. It operates in a 2.4 GHz band, over distances ranging from 10-100 m, with bit rates up to 1 Mbps. Data can be conveyed point-to-point, or point-to-multiple point with a master and up to seven remote devices.

Data can be conveyed either asynchronously (ACL) or synchronously (SCO). SCO links between two devices may be used to convey real-time audio. Both transmission modes may operate simultaneously for audio channels.

Audio in conventional bluetooth is conveyed with a sampling frequency of 8 kHz yielding 64 kbps with companding for voice communication. Alternatively, audio can be streamed using ACL mode with real-time transport protocol (RTP) and can deliver packets up to 721 kbps. The Audio/Video Control Transport Protocol (AVCTP) can be used to control devices, and the Audio/Video Distribution Transport Protocol (AVDTP) can disseminate audio. The Advanced Audio Distribution Profile (A2DP) can also be used to convey audio via Bluetooth; an SBC subband codec is used, and other codecs such as MPEG can be optionally used.

## **The Internet and Internet Audio**

The internet is a global collection of interconnected networks that permits transmission of diverse data to one or many users. While a network is a collection of computers sharing resources between them, the internet is a network of networks.

The communication and message routing standard that forges the links that form the internet is a set of documents called the Transmission Control Protocol/Internet Protocol (TCP/IP). Using these protocols, networks can share information resources, thus forming the internet.

The internet is a packet-switched network. A user sends information to a local network, which is controlled by a central server computer. At the server, the TCP parses the message, placing it in packets according to the IP, with the proper address on each packet. The network sends the packets to a router computer that reads the address and sends the packets over data lines to other routers, each determining the best path to the address. Packets may travel along different routes, thus helping spread loads across the network and reduces the average travel time; however, real-time transmission is difficult because packets can be delivered out of order, multiple times, or dropped altogether. When the packets arrive at the destination address, the information is assembled and acted upon.

Packet networks, such as the Internet, operate on a first-come, first-serve basis thus throughput rate is unpredictable. There is no bandwidth reservation, but the Internet cannot guarantee a percentage of the network throughput to the sender. The number of packets delivered per second is continuously variable; buffers at the receiver can smooth discontinuities due to burst delivery, but add delay time to the throughput. Finally, packet networks operate point to point; this greatly increases the bandwidth requirements for multicasting. To overcome this, new transmission protocols have been developed.

The Internet sends information over its infrastructure according to standardized addresses. Each address is governed by the Domain System Structure (method that uniquely identifies host computers and individual users). In its origins, the Internet had six high-level domains created to distinguish the types of users: com (commercial), edu (education), gov (government), mil (military), org (other), and net (network).

Access to the Internet requires a computer, a communications link (LAN connection or modem), and a gateway.

## **MP3**

Music can be coded and transmitted in various file formats (AU, WAV, AIFF), but they have large bandwidths that can make transmission times quite unsatisfactory. The most expedient types of formats are those that employ data reduction algorithms to minimize the file size so that music can be transmitted more quickly. Bit rates of less than 96 kbps are considered ideal for stereo programs, but 64 kbps can be considered reasonably good sound quality.

Depending on the reduction ration, music can be either efficiently downloaded or streamed in real time. Although any file type can be uploaded and downloaded, only a few formats can satisfy the need for both speed and fidelity such as MP3 or WMA.

The MPEG Layer III (MP3) data reduction algorithm is used to decrease file size prior to electronic distribution. MP3 files can be up/downloaded to the Internet, stored on hard disk, CD/DVD, or in flash memory. In any case, the data must be passed through an MP3 decoder

for playback. The MP3 algorithm does not specify means for encryption or copy protection; thus contents can be copied indefinitely. There are many software programs available that convert music files into MP3 files, and to decode MP3 files on a computer. In any case, the audio source is stored as an uncompressed format (WAV, AIFF) and then compressed to MP3 using dedicated means. Most encoders allow for different levels of compression (28.8, 64, 112, 128, 192, 320 kbps) may be allowed. Higher bit rates provide stereo playback at 44.1 kHz sampling rate, but lower rates will not.

The MPEG coding scheme does not provide error correction. If needed, additional channel coding (CRCC checksums) can be performed, but this is not generally required for network applications. Error concealment (but not correction) is possible.

A number of alternative codecs have been developed for Internet applications. Some are based on MPEG standards, while others are not. For example, Ogg Vorbis is an open-source codec that provides good-quality lossy data compression. It has variable bit rates ranging from 64-400 kbps, and has many applications.

### **Music Downloading and Digital Rights Management**

Digital Rights Management (DRM) systems have been developed to control the use of intellectual property content, restrict its copying, and identifying illegal copies, in e-commerce systems. Although no single standard has been universally accepted, various systems have been devised so that copyright-protected music can be distributed via the Internet.

DRM systems ideally should be platform independent, have minimal object code file size, require minimal computation, and be revocable and renewable. In addition, DRM systems should be reliable, flexible, unobtrusive, and easy to use. DRM not only defines copy protection, but copy limitations and how the copyright holder can be paid for the copy.

DRM systems can encode music for various applications, such as Internet delivery, storage in portable players, or the recording of CD/ DVD discs. DRM systems establishes usage rules limiting device playback, number of playbacks, time limits, or impose no restrictions.

A rights expression language (REL) is used to establish and communicate usage rules, while in some cases XML (eXtensible Markup Language) is used as a format.

DRM systems rely on cryptography to allow secure delivery between authorized parties. Authentication is used so that properly encoded data can be read only by compliant devices and media. Content is transmitted and stored in an encrypted form to prevent unauthorized copying, playback, or transmission, and recordings cannot be downloaded or played by others. "Signatures" or "fingerprints" of audio content can be used to limit playback or to search for content. The signature can be used to identify unknown content by comparing its signature against those in a database. Watermarks in the file can be used to prevent unauthorized copying and to identify illegally made copies.

The consumer's convenience and fair use must be respected. DRM systems must balance the robustness of their effectiveness versus usability. They discourage casual copiers, but can fall short of complete protection against professional pirates and determined hackers.

## **Streaming Audio**

If the size of a file is greatly reduced, so that it can be received as fast as it can be played, the file can be streamed. In this application, the music begins to play as soon as a buffer memory receiving the signal is filled.

It is far more difficult to stream files than it is to download files since not only the file size has to be small, but it must also cope with the packet-switching transmission method of the Internet. Packets usually arrive sequentially, but not always, and some packets might be missing, and must be retransmitted, which incurs a delay.

A buffer is needed because otherwise interruptions in the flow of data would cause interruptions in the playback signal. In addition, the different computers used in playing streaming files have different processing power and different Internet connection speeds. Finally, data speeds across an Internet path differ according to the path itself and traffic conditions.

The Internet TCP protocol is efficient for packet transmission, providing robust error correction and verification. However, the required overhead slows processors and transmission throughput speeds in general. In addition, TCP allocates per-connection bandwidth proportional to available bandwidth, without considering how much bandwidth is needed.

And alternative to TCP is the User Datagram Protocol (UDP) which is a simpler packet protocol without error correction or verification. Clients must request a missing packet, rather than receiving it automatically, thus promoting better throughput.

Some popular types of streaming formats are RealNetwork's RealAudio, Apple's QuickTime, Macromedia's Shockwave, Microsoft's Windows Media (WMA).

## **Audio Webcasting**

For the delivery of data from one point to another, or unicasting, the Internet is rather efficient. However, unicasting is not efficient in delivering the same message to millions of users because separate complete transmission would be required for each one. Internet Protocol (IP) multicasting is designed to overcome this problem. Multicasting can send one copy to multiple simultaneous users by using terrestrial broadcast, satellite downlink, or cable.

Instead of relying on servers to send multiple transmissions, individual routers are assigned the responsibility of replicating and delivering data streams. In addition, multicasting provides timing information to synchronize audio/video packets.

The Real-time Transport Protocol (RTP) delivers real-time synchronized data. The Real-time Transport Control Protocol (RTCP) works with RTP to provide quality-of-service information about the transmission path. The Real-time Transport Streaming Protocol (RTSP) is specifically designed for streaming applications. The Resource Reservation Protocol (RSVP) works with RTP or RTSP to reserve network resources to ensure a specific end-to-end quality of service. In this way, music distribution systems can efficiently distribute music to consumers.

There are two types of multicasting: back-channel and scalable.

Back-channel multicasting sends data over UDP that does not guarantee delivery, but does include a back or resend channel. A packet resend option can be employed when multicast UDP packets are lost. Bandwidth is static and will not dynamically vary according to network conditions. This type of multicasting is preferred when content quality is important and the audience size is limited. It is not suitable for large audiences because of the system resources

required for each client connection.

Scalable multicasting uses a data channel only that is multicast with RTP/RTCP as the data packet transport and session reporting protocol. Two RTP ports are used: one to provide data transmission, while the other reports the quality of data delivery. Data is delivered via UDP. The bit rate change is not dynamically varied, but it requires minimal server resources for any size of audience. However, there is no recovery mechanism or packet loss.

## **MPEG-4**

MPEG-4 is a group of tools (coding modules) for a variety of applications including interactive audiovisual coding, with an emphasis on very low bit rates and scalability of the coded bitstream that allows operation over the Internet and other networks. It supports high-quality audio and video, wired, wireless, streaming, and digital broadcasting applications. MPEG-4 specifies how to represent both natural and synthetic audio and video material as objects, and defines how these objects are transmitted or stored and then composed to form complete scenes. A scene might be simultaneously comprised of images, video, audio, and text. MPEG-4 will transmit all of these objects as multiplexed data streams along with a scene description.

The streams are demultiplexed at the receiver, the objects are decompressed, composed according to the scene description, and presented to the user. Moreover, because the objects are individually represented, the user can manipulate each object separately. The scene descriptions are used to describe the spatial-temporal synchronization and behavior during presentation. MPEG-4 also supports management of, and controlled access, to intellectual property.

## **MPEG-7**

MPEG-7 is a standard entitled "Multimedia Content Description Interface." It provides a way to characterize multimedia content with standardized descriptions and thus allow efficient searches for content. MPEG-7, as opposed to metadata, describes more intrinsic characteristics of a file. For example, for audio it might describe key signature, instrumentation, and other parameters derived from the content itself, as well as information derived from the signal such as waveform power, spectral spread, harmonicity, etc.

## **Encryption**

Used to prevent copying even as content moves across many platforms and media in various incarnations. Ideally, content protection systems should offer rights management, usage control, authentication, piracy deterrence, and tracking capabilities.

Cryptography provides one method to protect the contents of a files. Data is encrypted prior to storage and transmission, then decrypted prior to use. The encryption algorithm is a set of mathematical rules for rendering information unintelligible. To decrypt a file, a "key" is used to decipher the data.

Modern keys consist of binary numbers, perhaps from 40-128 bits in length. In many applications, there is a public key that many users can employ, as well as private keys reserved for privileged access. Many file formats, such as audio, have data placed into frames; the frame contains a frame header with auxiliary information that assists decoding. A key can be placed in the header so that the file can be decoded if the user is authorized to do

so. Illegitimate files would lack the key, and likewise unauthorized decoders could not playback the file. However, if the data is separated from the key, or if the key is hacked, the file can be accessed without any restriction.

Encryption may require a severe computation overhead, or a small increase in file size. Of course, encryption cannot protect the file after it has been decrypted.

## **Watermarking**

Digital watermarks offer another security mechanism that is intrinsic with the data and hidden in it. Watermarking ties ownership to content and can verify the content's authenticity. The aim of watermarking is to embed an inaudible digital code (tag) into an audio signal so that the identifying code can be extracted to provide an electronic audit trail to determine ownership, trace piracy, etc.

Source watermarks are attached to a specific media to identify and protect the content. Transactional watermarks are independent of the media and track usage. In some cases, the watermark itself is encrypted.

Watermarks should have the following characteristics:

1. Transparency - inaudible to the end user, and not affect the quality of the audio signal.
2. Robustness – can be recovered from the audio signal and can survive processing, data compression, or perceptual coding.
3. Secure – it cannot be removed, altered, or defeated by an unauthorized user.
4. Guard against false positives and negatives – watermarks should not be mistakenly identified where they do not exist, or misinterpreted.
5. Data overhead should be kept to a minimum.

Some watermarking methods employ signal phase manipulation, placing data in an undetectable frequency range, or employ spectrum-spread encoding. Many watermarking systems combine two or more methods to achieve a more robust solution.

There are many watermarking techniques, each with a different intent.

1. Copy control watermark – designed to prevent casual unauthorized copying. Can allow for unlimited copies, limited number, or no copies.
2. Forensic watermark – designed to deter piracy. Uses a significant amount of data for file authentication and detection of tampering, using a verification key.
3. ID3 tags – embed song and artist information during MP3 ripping. A Track Unique Identifier (TUID) can be added to a song to link it to the source album.
4. Audio fingerprinting – content-based identification technology. Contents are scanned and signal characteristics are extracted and stored in a database. The extracted information inherently identifies the content itself. When unknown music is played, characteristics are again extracted, and compared to those already stored. A match is used to identify the music work. Used to track music playback by broadcasters and webcasters for royalty payments, and to identify pirated music. Alternatively, they can be used to recommend similar music to a listener.