Lecture 2: Basic Concepts and Terminologies

(09/02,09/2009)

Lecture Outline

1. `Data types and their characteristics`
2. `Data communication and networking terms`
3. `Network types`
4. `Network QoS`
5. `Application QoS`

1. Data types and their characteristics: Fig.1.2, p.6

(0) Data types: they describe the contents of data communications, i.e. what is been sent over communication media

(1) Text

a. **Unformatted text**: also called **plaintext**, enables pages to be created which comprise strings of fixed-sized characters from a given character set;

b. **Formatted text**: also called **richtext**, enables pages and complete documents to be created that are comprised of strings of characters of different styles, size and shape with tables, graphics, and images inserted at appropriate points;

c. **Hypertext**: this enables Web pages to be created in the form of an integrated set of documents with defined hyper links.

(2) Images

a. **Computer-generated images**: also called **computer graphics** or simply graphics;

b. **Digitized images**: they include digitized documents and digitized photos and pictures. Scanned pictures or documents are typical examples. Both types of images are represented by a two-dimensional matrix of individual elements known as **pixels**.

c. Monochromatic and color images:

(a) Each pixel in a monochromatic image is represented by a *gray level* value. A good quality monochromatic image can have 8-bit, 256 levels of greys.

(a) Each pixel in a color image is represented by a value that denotes the color of that pixel. 8-, 16- and 24-bit colors are all typically used.

· The more colors used, the better the quality, and the more data needed for a given image.

· Typical screen sizes are 640×480 pixels for a VGA and 1024×768 for SVGA.

· For a VGA screen a 8-bit per pixel image requires 307.2 kbytes while on a SVGA screen a 24-bit pixel image requires 2.359296 Mbytes.

(3) Video

a. Video images can be generated by digital video recording equipments such as video camecorders. Such digital images consist of a sequence of digitized images called **frames**.

b. Analog video signals have to be converted to digital form first. A **digitization format** has to be selected first for the conversion.

(4) Audio

a. **Analog audio**: almost all audio signals are generated in an analog form by a microphone.

(a) **Analog-to-digital converter** (ADC): converts analog audio signals to digital form.

· This conversion procedure involves sampling the amplitude of the speech/audio signal at regular intervals. The higher the sampling rate, the better quality is the reproduced analog signal.

· The quality of reproduced analog signal is determined by both the sampling rate and the bits used per sample. In speech singal a typical sampling rate is 8kHz with 8bits per sample. This yields a digital signal of 64kbps, which is the bit rate used in telephone network.

· For CD quality audio, a common sampling rate is 44.1kHz with 16 bits per sample, which produces a bit rate of 705.6 kbps, or for stereo audio, 1.411 Mbps.

(b) **Digital-to-analog converter** (DAC): converts digital audio signals back to analog form.

(c) To save bandwidth, converted digital signals normally are compressed using various *compression algorithms*.

b. **Continuous data**: they are generated in real-time contiuously. It is also called **real-time data**. The audio devices at both end have to operate in **streaming** mode.

(a) The bit rate at which the source data stream is generated can be either **constant bit rate** (CBR) or **variable bit rate** (VBR).

(b) CBR is used for continous audio data. VBR is used for video due to compressions which produce different amount of data from the source due to the interrelationships of frames in the video.

b. **Block-mode data**: they are block of data generated in a time-independent way, such as emails, static images and so on.

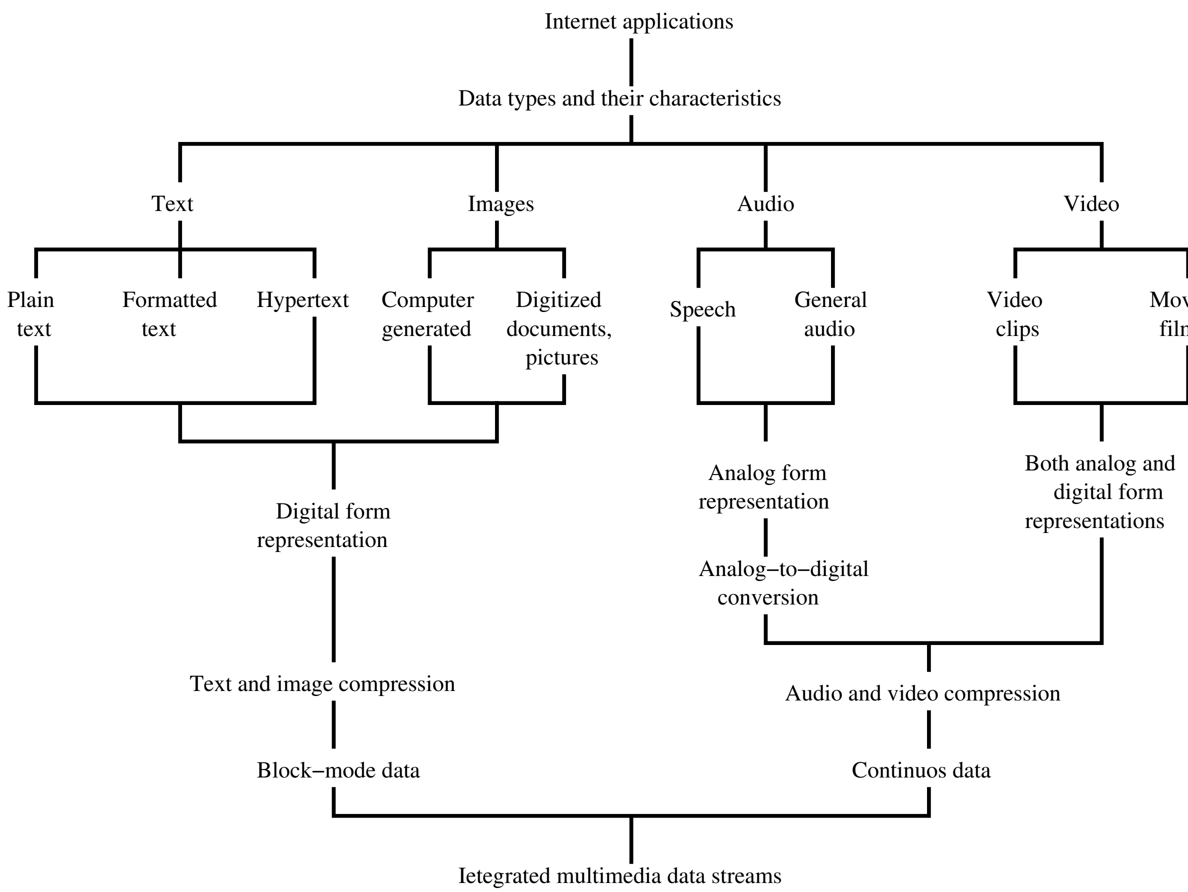

Figure 17: Data types used in Internet applications (Fig.1.2,p.6)

2. <u>Data communication and networking terms</u>: Fig.1.3, p.8

(1) Communication modes: the operation mode of the communication media and applications (Fig.1.4, p.9)

(2) Physical communication modes: the operation mode of the communication media

a. **Simplex**. Used when data is transmitted in one direction only.

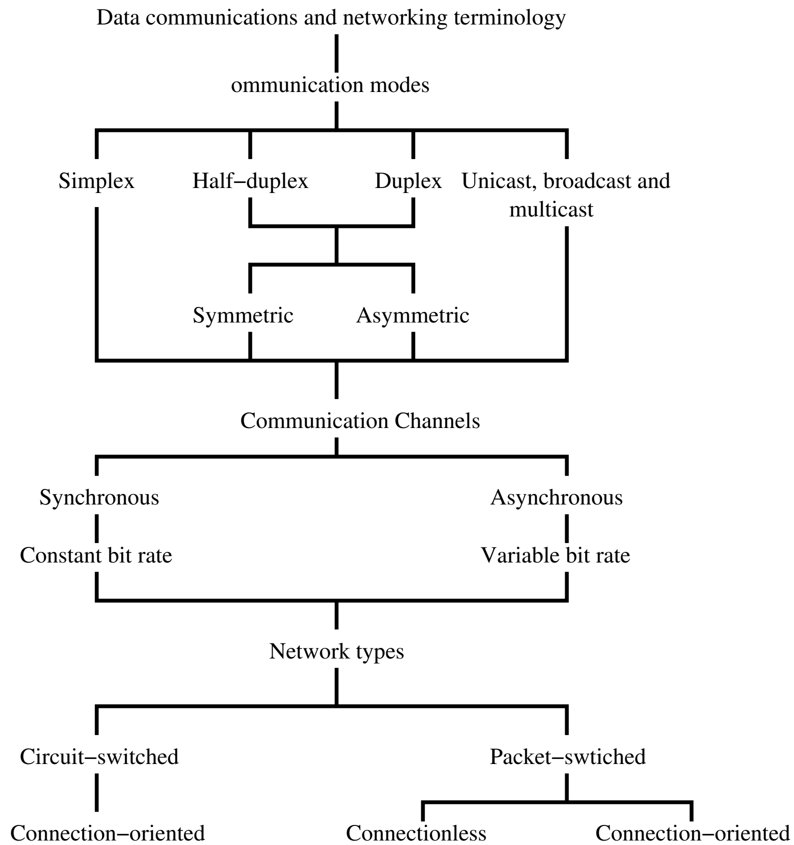

Figure 18: Data communications and networking terminology (Fig.1.3,p.8)

b. **Half-duplex**. Used when data is transmitted in both directions alternately, i.e. two DTEs take turns to perform transmissions and there will never be simultaneous transmissions by both ends.

c. **Duplex**(also referred as **full-duplex**). Used when data is transmitted in both directions simultanenously.

d. **Symmetric** vs **asymmetric** half-duplex and duplex:

(a) **Symmetric**: if the bit rate associated with the flow of data along each of the two directions is the same. Otherwise, it is called **asymmetric**.

(b) Interactive video conferences are typically symmetric duplex. Web browsing is typically asymmetric duplex.

(3) Logical communication modes: the operation mode of applications

a. **Unicast**. Used by a pair of applications to exchange information. Data sent by one device on behalf of one of the two applications is delivered to ther other device on which the other application is running.

b. **Broadcast**. Used when data sent by one application in a particular device is intended to be received by all other devices so that a corresponding application on each of these device can receive the data.

c. **Multicast**. Similar to broadcast, except that the data sent by a specific device on behalf of an application is intended to be received by members of a *multicast group*.

3. Network types

(0) Two types of communication channels:

a. Those that operate in a time-dependent way known as **circuit-mode**, and those that operate in a time-varying way known as **packet-mode**.

b. Circuit-mode channels are also called **synchronous communication channels** as they provide a constant bit rate service at a specified rate. Packet-mode channels are also called **asynchronous communication channels** as they provide a variable bit rate service. The actural bit rate is determined by the variable transfer rate of packets across the network.

(1) Circuit switched

a. Basic scheme: Fig.1.5, p.11. Fig.19(a) provides another illustration.

b. Basic elements:

(a) A set of *switching offices/exchanges* interconnected through communication links.

(b) Each DTE has a network-wide unique address.

c. Operations:

(a) Before a communication session can take place, a physical connected path has to be made in the communication media.

(b) Upto 1960s for PSTNs phone connectiosn were through phone operators

(c) **Call/connection setup delay**: could be 10 sec or longer for each call.

(d) Automatic switching equipment significantly reduces the delay.

(e) When a communication session terminates, either the source or the destination will initiate termination of the connection. The messages used to set up and terminatea connection are known as **signaling messages**.

(f) The bit rate associated with a connection is fixed and in general is determined by the bit rate of access circuits that connect the source and destination to the network.

c. Issues

(a) Advantages: reliable. Once a connection is established the path (resources) is guaranteed. The only delay is the propagation.

(b) Problems:

· Need of reserved physical connection paths. Non-flexible. The non-availability of a single section would prevent a connection establishment.

· Call/connection setup delay: for PSTN, the delay can be a fraction of a second for local calls can be upto several seconds for international calls. For ISDN, this delay is typically in the range of tens of milliseconds to several hundred milliseconds.

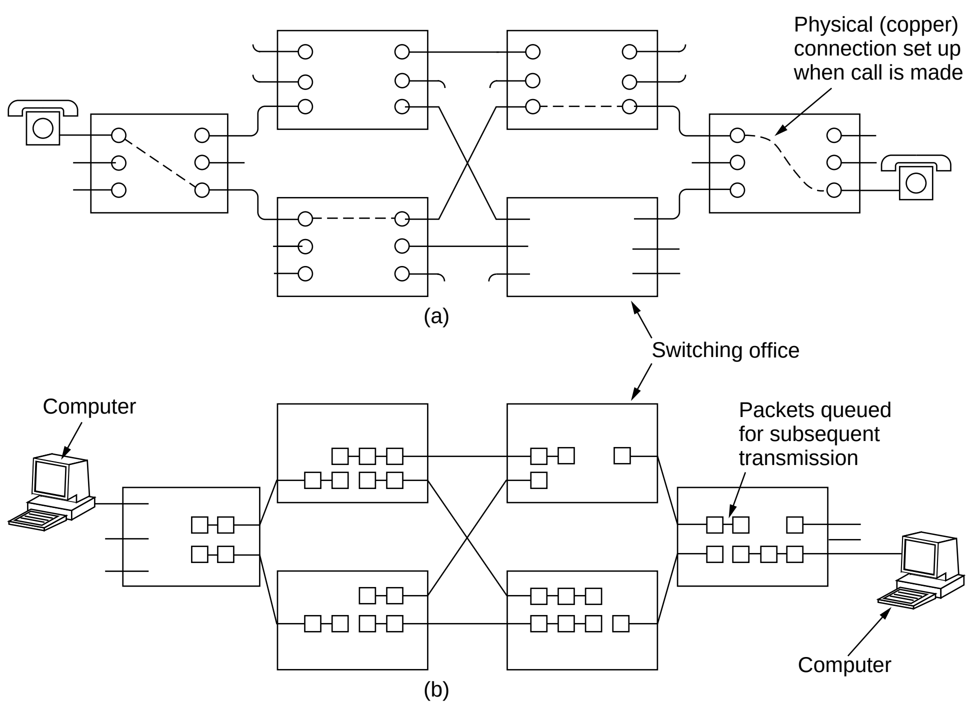

Figure 19: (a) Circuit-switching. (b) Packet-switching.

(2) Packet switched

a. Basic scheme: Fig.1.6, p.12. Fig.19(b) provides another illustration.

b. There are two types of packet-switched networks: **connection-oriented** (CO) and **connectionless** (CL).

c. Principles of CO packet-switched networks: Fig.1.6(a)

(a) A CO packet-switched network comprises an interconnected set of PSEs (packet-switched exchanges). Like in a circuit-switched network, every DTE in a CO packet-switched network has a unique address/number.

(b) **Virtual circuit**.

- Prior to a communication taking place, a connection has to be set up. This is like in circuit-switched network.
- However, the important difference is that in this case, only a variable portion of the bandwidth of each link along the connection path is utilized by the connection. Such a path is called a virtual circuit (VC).

(c) VC set up.

- To set up a VC, the source DTE sends a *call request* control packet to its PSE. The request contains the addresses of the source and destination DTEs, as well as a short identifier known as the **virtual circuit identifier** (VCI).
- Each PSE maintains a table which specifies which outgoing link should be used to reach a given destination address.
- Upon receiving a call request, a PSE selects an outgoing link based on the destination address in the call request. The PSE assigns the next free VCI for this outgoing link. The PSE then adds two entries in a **routing table**. The first is the incoming link/VCI and the corresponding outgoing link/VCI pair. The second is the inverse of these two VCIs (for route packets in reverse directions). The call request is then forwarded to the next PSE along the outgoing link.
- The above process will continue until the call request reaches the PSE that connects directly to the destination DTE. At that time, a VC, which is the collection of VCIs that were just entered to routing tables, has been set up. A *call accepted* control packet is sent back from the destination to the source over the VC.

(d) Operation of VC

- With a VC already set up, data transfer using the VC is simple. There is no need to include the full network-wide addresses of source and destination DTEs. Only the VCI is needed in the header of every packet.
- Upon receiving an incoming packet, a PSE will look at the VCI inside it. It will then consult its routing table for an entry that has the VCI in the packet as the first component. For example, if the VCI is 2, the PSE will look for an entry $(2, i)$, and then will replace the VCI value 2 in the

packet with new VCI value i, and forward the packet along the link that has a VCI i.

· The same procedure is followed to return data in reverse direction.

· When data exchange completes, the VC is cleared the appropriate VCIs are released by passing a *call clear* packet along the VC.

d. Principles of CL packet-switched networks: Fig.1.6(b)

(a) There is no need of connection establishment. A DTE can send a packet to another DTE at any time. Each packet has to carry in its header the full network-wide address of the source and destination DTEs.

(b) In a CL packet-switched network as each interconnected element has to make a decision as how to route (forward) an incoming packet based on the destination address, each element is called a **router** instead of PSE.

e. Issues

(a) **Best-effort service**.

· Both CO and CL packet-switched networks store each received packet in its entirety in a memory buffer.

· Then an error check is made against the packet. If the packet is found corrupted, the PSE/router simply drops it.

· If nor error of the packet is found, the packet will be sent along the outgoing link, as fast as the link can.

This behavior is called best-effort service. There is no guarantee of delivery.

(b) **Store-and-forward** networks and store-and-forward delays

· As every packet is first stored and checked for possible errors, packet-switched networks are also called store-and-forward networks.

· Store-and-forward introduces a delay.

· Additional delay is possible if there are several incoming packets that all have to be forwarded along the same outgoing link. These incoming packets may have to be queued at the PSE/router and extra delay occurs.

· The sum of the store-and-forward delays in each PSE/router contributes to the overall delay of the packet across the network. The mean of this delay is known as **mean packet transfer delay** and the variation about the mean is called **delay variation** or **jitter**.

f. Examples of packet-switched networks

(a) CL packet-switched networks: the Internet is the best example.

- IP based routing is a CL packet-switched delivery process.
- On the other hand, TCP uses the VC concept, but it is still a CL packet-switched based protocol. There is not VCI at each router.

(b) Two examples for CO packet-switched networks:

- X.25 international packet-switching network
- ATM networks

(4) Illustration of timing of events in the three switching techniques: Fig.20.

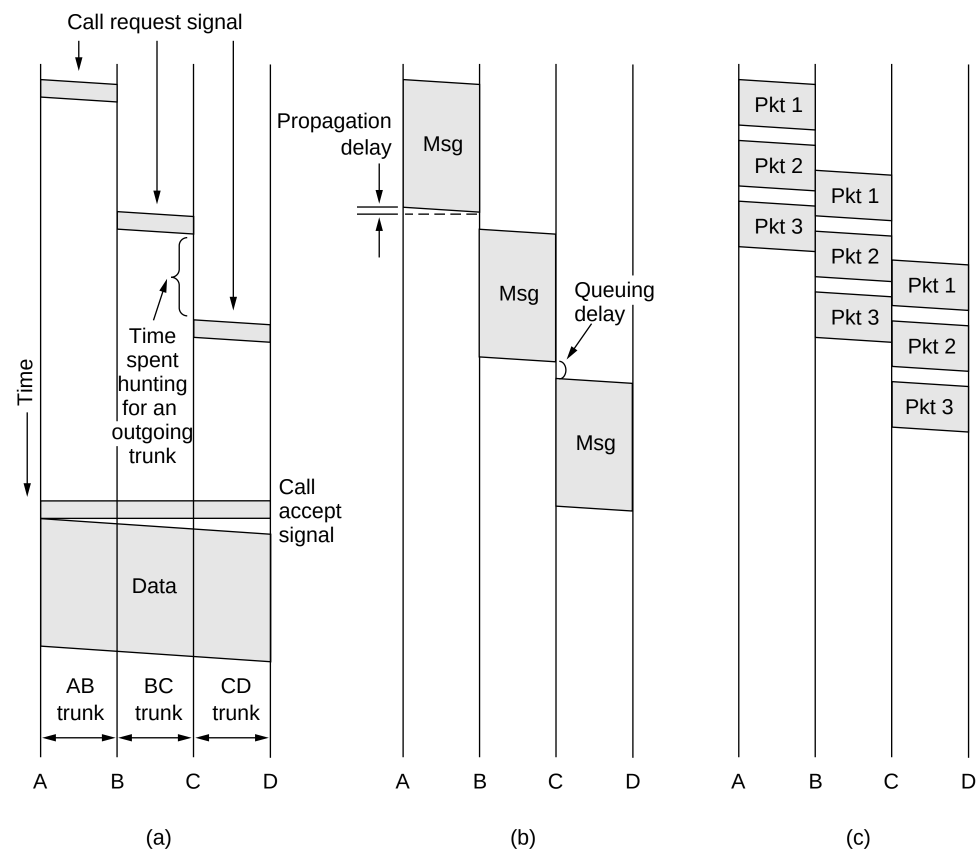

Figure 20: Timing of events in (a) Circuit-switching, (b) message-switching, (c) packet-switching.

a. Note: this diagram is from another source. Message switching is another technique similar to packet switching, except that data is sent in units of (logical messages), not in units of packets.

b. For circuit switching, there is a significantly visible interval of call latency time. However, once a connection is established, transmission is continuous.

c. Message switching has a visible *queuing delay* time, which is the time used for perform store-and-forward.

d. Packet switching is more flexible than message switching. There are time overlapping in packet sending and receiving among neighbor stations.

(5) Comparison of circuit-switching and packet-switching:

Item	**Circuit switched**	**Packet switched**
Call setup	Required	Not needed
Dedicated physical line	Yes	No
Each packet follows the same route	Yes	no
Packets arrive in order	Yes	No
Is a switch crash fatal	Yes	No
Bandwidth available	Fixed	Dynamic
Time of possible congestion	At setup time	On every packet
Potentially wasted bandwidth	Yes	No
Store-and-forward transmission	No	Yes
Transparency	Yes	No
Charging	Per minute	Per packet

4. Network QoS

(1) **Network Quality of Service (QoS) parameters**: they are the operational parameters associated with a digital communications channel through the network.

a. The QoS parameters collectively determine the suitability of the channel in relation to the user for a particular applications.

b. Different networks have different network QoS parameters. The QoS parameters for circuit-switched networks are very different from those of packet-switched networks.

(2) QoS of circuit-switched networks

a. Main QoS parameters:

(a) The bit rate;

(b) The mean bit error rate;

(c) The transmission delays.

b. The **bit error rate** (BER) of a digital channel is the probability of a binary bit being corrupted during its transmission across the channel over a defined time interval.

(a) For a constant bit rate channel, this equates to the probability that a bit is corrupted in a defined number of bits.

(b) For example, a mean BER of 10^{-3} means that on average one of every 1000 bits transmitted is corrupted.

(c) If BER is relatively small, some applications can tolerate it. For example, several corrupted bits in a segment of digitized speech signals are hardly noticeable.

(d) In some other applications such as financial record transmissions, any bit errors cannot be tolerated.

- For these applications the source data is normally segmented into blocks prior to transmission (to isolate errors and facilitate retransmissions in case of errors).
- The maximum size of the blocks is affected by the mean BER. A large block size that is close to the inverse of the mean BER implies that almost every block can get an error. For example, if the mean BER is 10^{-10}, then the block size must be considerably smaller than 1000.

c. Probability of a bit error in a block.

(a) Even if BER is very small and the block size is tiny, it is still possible that a bit in the block can be corrupted.

(b) If P is BER and N is the number of bits in a block, then the probability of a block containing a bit error, P_B, is given by:

$$P_B = 1 - (1 - P)^N$$

If $N \times P < 1$, then P_B is approximately equal to $N \times P$.

d. **Unreliable service**: also called **best-try** or **best-effort** service.

(a) Any blocks with bit errors will be discarded within the network. This can happen at any PSE/router or in the network interface at the destination (for both circuit-switched and packet-switched networks).

(b) For any applications that demand only error-free blocks are acceptable, mechanisms for error detection and retransmission are needed.

e. **Reliable service**: the source network interface or computer must divide the source information into blocks of a defined maximum size.

(a) The destination must be able to detect when a block is missing and notify the source.

(b) The source has to keep a copy of every transmitted block for retransmission.

(c) Error detection and retransmission introduce a delay. In addition, retransmission also introduce extra overhead in the form of control information in each block.

(d) Choice of block sizes:

- To reduce retransmission delays, smaller block sizes are more desirable.
- On the other hand, because each block has to contain the additional control information, smaller block sizes imply that more control information have to be sent.

Therefore, the choice of actual block sizes is a compromise.

Note: we just discussed that block size is also affected by the BER.

(e) **Example** (Example 1.1, p.15) If the desired probability of a block containing an error (hence being discarded) is 10^{-1}, derive the maximum block size that should be used over a channel which has a mean BER probability of 10^{-4}.

Answer:

	$P_B = 1 - (1 - P)^N$
Hence	$0.1 = 1 - (1 - 10^{-4})^N$, and $N = 950$ bits
Alternatively	$P_B = N \times P$
Hence	$0.1 = N \times 10^{-4}$, and $N = 1000$ bits

(f) Delays. The total transfer delay (the book used the term *transmission delay* there. That term is used with a special meaning and hence should be reserved) of a block is determined by the following:

- The delays that occur in the source/destination DTE interfaces (the **codec delays**);
- The signal **propagation delay** while the block is been transmitted from the source to the destination.

Propagation delay is independent of the bit rate of channels. It will be formally defined and discussed in next lecture.

(3) QoS of packet-switched networks

a. Main QoS parameters:

(a) The maximum (allowed) packet size;

(b) The mean packet transfer rate;

(c) The mean packet error rate;

(d) The mean packet transfer delay;

(e) The worst-case jitter delay;

(f) The transmission delay.

b. The actual delay of each packet is a variable and is affected by both the bit rates of interconnecting links and store-and-forward delays at each PSE/router.

   (a) The parameter **mean packet transfer rate** is the average number of packets that are transferred across the network per unit of time (second, normally).

   (b) Mean packet transfer rate multiplying the maximum packet size provides the equivalent mean bit rate of the channel.

c. The **mean packet error rate** (PER) is the probability of a received packet containing one or more bit errors.

   (a) PER is the same as the mean block error rate (derived in previous example).

   (b) Hence PER is determined by both the maximum packet size and the worst-case BER of the transmission links that interconnect source and destination DTEs.

d. **Mean packet transfer delay** was already defined in the section about issues in packet-switched networks. It is the summation of the mean store-and-forward delay that a packet experiences in each PSE/router that it encounters along a router. The term **jitter** is the worst case variation of this delay.

5. <u>Application QoS</u>

(1) Network QoS vs application QoS

a. Network QoS defines what the network can offer in terms of quality of service. Different applications may have their own QoS parameters.

   (a) A video conferening application may define the minimum resolution of each video frame together with the size of each frame.

   (b) An audio application may require a minimum sampling rate in digitization.

b. Some of the application QoS parameters are related to and are meaningful only when the underlying network QoS parameters support them. Example such application QoS parameters:

   (a) The required bit rate or mean packet transfer rate;

   (b) The maximum startup delay;

   (c) The maximum end-to-end delay;

   (d) The maximum delay variation/jitter;

   (e) The maximum round-trip delay.

c. **Required bit rate or mean packet transfer rate** and **maximum end-to-end delay** are particularly important for applications that involve the transfer of a constant bit rate stream. In addition, for these applications the parameter **maximum delay variation/jitter** is equally important because if the rate of arrival of bitstream is variable with larter jitter, it will cause the problems ain the destination decoder.

d. The parameter **start-up delay** is important for interactive applications. It defines the amount of time that elapses between the time instance an application makes a request to start a session and the time instance the application receives a confirmation sent by the corresponding application at the destination.

   (a) This amount of time includes the time needed to establish a network connection as well as the delay introduced at both the source and destination DTEs to negotiate the session establishment.

   (b) RTD (round-trip delay) is an important part of the first time component. For interactive applications we would like start-up delay as short as possible, normally less than a few seconds. But a long RTD would make shortening this delay impossible.

(2) Circuit-switched networks are better for applications that involve the transfer of a constant bit rate stream.

   a. The call setup delay in such networks is often not important for those applications;

   b. The channels in circuit-switched networks provide a constant bit rate service of known rate.

(3) Connectionless packet-switched networks appear better for interactive applications:

   a. There is no network call setup delay;

   b. Any variations in the packet transfer delay are not important.

(4) Example (benefit of packet-switched network over circuit switched netwok): transfer of a large file of data from a server computer to a home client PC over the Internet.

   a. Access from a home PC to the Internet can be through PSTN (with a modem), an ISDN connection, or a cable modem (the case of DSL is similar to ISDN. Actually ISDN is one type of DSL).

      (a) Through PSTN, it is circuit-switched, providing a constant bit rate channel in the order of 28.8kbps (note: the latest V.90 and V.92 modem can support 56kbps). ISDN is also circuit-switched, the bit rate is in the order 64/128kbps.

(b) Cable modems operate in packet-switched mode. The modems in a cable region time-share the use of a single high bit rate channel, typically at 27Mbps (More details are in Ch.5). The actual bit rate an individual application can obtain depends on the number of shared users. If the number of concurrent users is 270, then each user/application would get a mean data rate 100kbps.

b. For applications like file transfers, the main parameter of interests is not the mean data/bit rate but the total time to transmit the complete file.

(a) For PSTN and ISDN, this total time is directly related to the channel bit rate and size of the file.

(b) With cable modem, the actual channel bit rate available to a specific application is a variable (of course up to the limit of 27 Mbps). However although applications time-share the use of the 27 Mbps channel, when an application gain access to it, the file transfer can take place at a very high rate. If the file size is 100 Mbits, the minimum transfer time using the various Internet access modes are:

PSTN and 28.8kbps modem	57.8 minutes
ISDN at 64kbps	26 minutes
ISDN at 128kbps	13 minutes
Cable modem at 27Mbps	3.7 seconds

With cable modem if other applications also transfer files during the time the file is being transfered, the completion time of the file transfer will increase. However, the probability of multiple users requesting a transfer in this short time window is relatively low.

(5) **Buffering** technique in packet-switched networks

a. The use of which type of networks for what type of applications is not fixed. Often alternative network types can be used in many instances:

(a) Interactive applications can also be done in PSTN and ISDN as call setup delays now are quite small and hence acceptable.

(b) For constant bit rate applications can be performed using packet-switched network if the *equivalent mean bit rate* provided by the network is greater than the input bit rate and the *maximum jitter* is less than defined value.

b. **Buffering** is used to overcome (or at least reduce) the effect of jitter: Fig.1.7, p.19.

(a) The effect of jitter is overcome by retaining a defined number of packets in a memory buffer at the destination before playout of the information bitstream is started.

(b) The memory buffer operates in FIFO mode and the number of packets buffered there before output starts depends on the worst-case jitter as well as the bit rate of the information stream.

(c) **Packetization delay**:

· This is the (extra) delay incured at the source when information bit stream is converted into packets. This delay adds to the overall input-to-output delay.

· To minimize the overall delay, the packet size used for an application is made as small as possible, but sufficient large in size to overcome the effect of the worst-case jitter.

c. Example (Example 1.3, p.20). A packet-switched network with a worst-case jitter of 10 ms is to be used for a number of application each of which involves constant bit rate information stream. Determine the minimum amount of memory that is required at the destination and a suitable packet size for each the following input bit rates. It can be assumed that the mean packet transfer rate of the network exceeds the equivalent input bit rate in each case.

(i) 64 kbps

(ii) 256 kbps

(iii) 1.5 Mbps

Answer:

(i) At 64 kbps, 10 ms will transmit 640 bits. Hence choose a packet size of, say 800 bits with FIFO buffer of 1600 bits - two packets - and start playout of the bitstream after the first packet has been received.

(ii) At 256 kbps, 10 ms = 2560 bits. Hence choose a packet size of, say 2800 bits with FIFO buffer of 4800 bits (or 5600 bits to hold two packets).

(iii) At 1.5 Mbps, 10 ms = 15000 bits. Hence choose a packet size of, say 16000 bits with FIFO buffer of 32000 bits.

Note:If the computed packet size exceeds the network maximum packet size, then the equivalent number of packets must be sent before playout starts. For example, if the maximum network packet size was 8000 bits, then for case (iii) above playout would not start until two packets have been received and the FIFO buffer should hold four packets.

(6) **Service classes**

a. A number of standard service classes have been defined to simplify the the process of determining whether a particular network can meet the QoS requirements of an application.

   (a) Each service class has an associated set of QoS parameters and a network can either meet this set of parameters or not.

   (b) For networks that support a number of different service classes, in order to ensure the QoS parameters associated with each class are met, the packets belonging to different classes are given different **priorities**. The Internet makes such practice.

      · Packets containing real-time streams such audio and video are more sensitive to delay, jitter and loss than packets containing textual information.

      · Hence packets relating to multimedia applications involving real-time streams are given higher priority than the packets relating to applications such file transfer or email.