Chapter 3

ATM and Multimedia Traffic

In the middle of the 1980, the telecommunications world started the design of a network technology that could act as a great unifier to support all digital services, including low-speed telephony and very high–speed data communication. The concept of a network capable of integrating all ranges of digital service emerged. The name given to this network was broadband integrated services digital network (B-ISDN).

Several groups and telecommunication companies worked in parallel on alternative proposals for the technical implementation of the network. At the end of a long process, ATM technology was selected to support the B-ISDN network. ATM as a technology designed to support various classes of service, is the solution of choice for supporting long-haul digital multimedia applications.

The possibility of setting up the virtual connections at speed of several dozen megabits per second with a variety of guaranteed levels for the bit rate and the jitter, should satisfy most applications. The typical transit delay of a couple to a tenth of millisecond propagation delay excluded is compatible with most of the applications of multimedia. For applications requiring a constant bit rate, the circuit emulation service can be used.

The residual cell loss rate of 10^{-8} to 10^{-10} is suitable for all types of real time transmission of voice and video streams.

The issues regarding the risks of congestion should not in practice affect users in the long term, because the manufacturers are expected to take the necessary measures to limit the statistical nature of the multiplexing if the quality of service cannot be satisfactorily guaranteed. Some early services may, however, suffer from serious teething problem [29].

3-1 ATM and Traffic

As mentioned the ATM Network can support variety of services, such as video, voice and data on a single infrastructure, to do so ATM networks must provide traffic management [30]. We can describe the traffics as time-based and non-time-based information [31], time-based information is sensitive to time varying as video, and voice, non-time-based is insensitive to time varying as image, and data.

Its traffic characteristics and the corresponding communication requirements can characterize an application. Its traffic generation process can formally specify the traffic characteristics of an application. Since the traffic generation process (or traffic pattern) is basically a sequence of packets generated at arbitrary instants, two stochastic processes can characterize the traffic pattern:

a) The packet generation process (or packet arrival process).

b) Packet length distribution function.

The communications requirements of an application include bandwidth, delay, and error guarantees. The bandwidth requirements of an application (in each direction) are typically specified in terms of peak and average bandwidth. For CBR applications, the peak and average bandwidth are the same. For image browsing applications, a full screen photo image of 3 Mbytes (1000 x 1000 x 24 byte), after compression 300 Kbytes by Joint Photographic Experts Group (JPEG) compression [32]. This requires about 24 Mbps link (peak) bandwidth to satisfy the response time requirements.

An application can be classified according to its information delivery requirements as a real-time or non-real time application. A real time application is one that requires information delivery for immediate consumption, for example, a telephone conversation. Non-real time application information is stored (perhaps temporarily) at the receiving points for later consumption, for example, sending electronic mail. Figure 3-1 shows one new view [33,34].

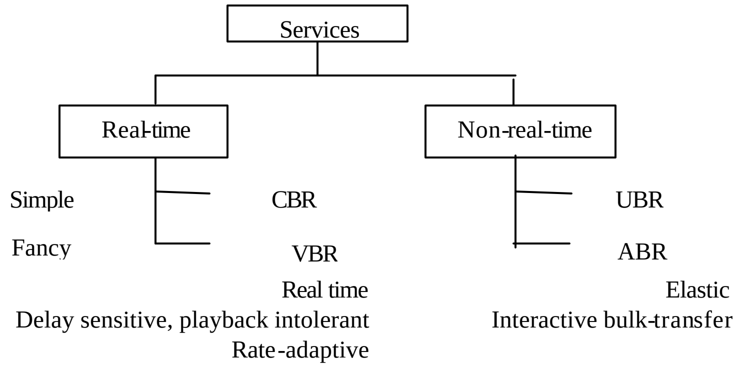

Figure 3-1 ATM Model "*Hierarchy*".

3-1-1 ATM Forum Traffic Categories

The ATM forum has defined the following traffic categories based on the different requirements Constant Bit Rate (CBR), real time and non-real time Variable Bit Rate (VBR), Unspecified Bit Rate (UBR), and more recently Available Bit Rate (ABR). These categories are discussed below.

Constant Bit Rate (CBR):

The CBR category is intended for applications requiring tightly constrained delay and delay variation. Such as voice and video applications which are expected to transmit at a continuos rate. CBR services use ATM Adaptation Layer Type-1 [21], because it receives/delivers SDU (Service Data Unit) with a constant bit rate from/to the layer above. The CBR class of service is the preferred choice for many video dial tone service providers [35].

Variable Bit Rate (VBR):

The VBR category is intended for applications that share the requirements for tightly constrained delay and delay variation of CBR traffic, but which transmit at a variable rate. Compressed voice with silence suppression, and variable rate video codecs are examples of this category of traffic. ATM Adaptation Layer Type-2 is proposed for VBR services with a timing relation between source and destination [21], for example VBR Voice or video.

Undefined (or Unspecified) Bit Rate (UBR):

The UBR was originally intended for data application, which do not require tightly constrained delay or delay variation. The sources are not required to specify the bandwidth they will require.

Available Bit Rate (ABR):

The ABR mechanisms provide flow control back to the source to change the rate at which the source is submitting traffic to the network, ABR is intended for application that need a more reliable service than provided by UBR, such as critical data transfers and computer server applications.

3-1-2 Traffic Parameters

The performance of any application using an ATM network can be defined in terms of the following parameter [30]:

Throughput : Called **goodput,** bits per second delivered to the application.

Latency : The sum of the transmission delay (reduced by higher transmission speed), propagation delay (determined by physics), and queuing delay through each network element (switch).

Jitter : The variation in delay, or the variation in the inter cell arrival of consecutive cells. Certain applications, such as voice, are very sensitive to jitter.

Cell Loss : The amount of cell or packet loss the application can tolerate. Continuous services are relatively intolerant of cell loss.

3-2 Multimedia of Traffic Models

Multimedia application includes the voice, video, and data traffics, these traffics are different in nature and can applied at the terminal (TE) as shown in Figure 3-2.

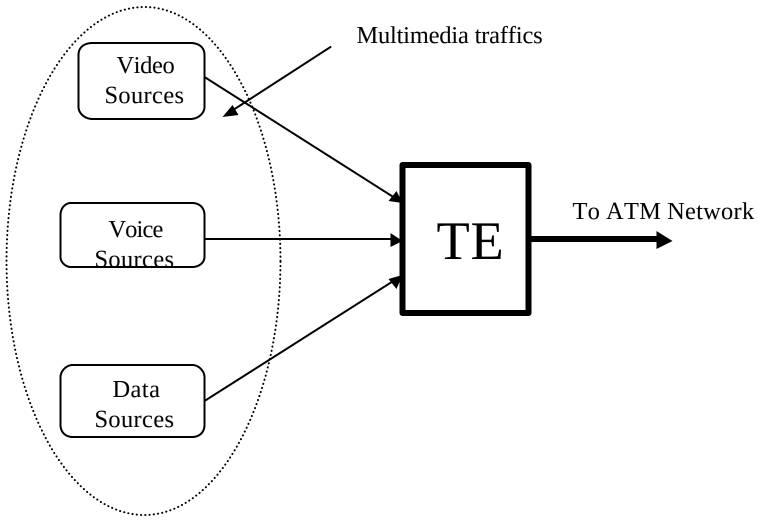

Figure 3-2 TE Trafics Configuration.

3-2-1 Voice Traffic

Figure 3-3 illustrates the block diagram of a station that encodes and sends a voice stream. Each voice source has a continuous time, analog signal is digitized by a coder as in [36]. The generated samples are accumulated in a packetizer, when the number of samples in the packetizer reaches the pre-determined cell length, header is attached then a voice cell is generated. The voice cell generation process may be synchronized to an external timing. The generated cells are stored in the transmit buffer in the order of their generation waiting for transmission.

Note that, in some LAN, the voice samples are transmitted directly using an assigned TDM channel on the network. In most other LAN protocols, the voice is transmitted in the form of cells where each cell consists of a number of voice samples within a packetization interval [37].

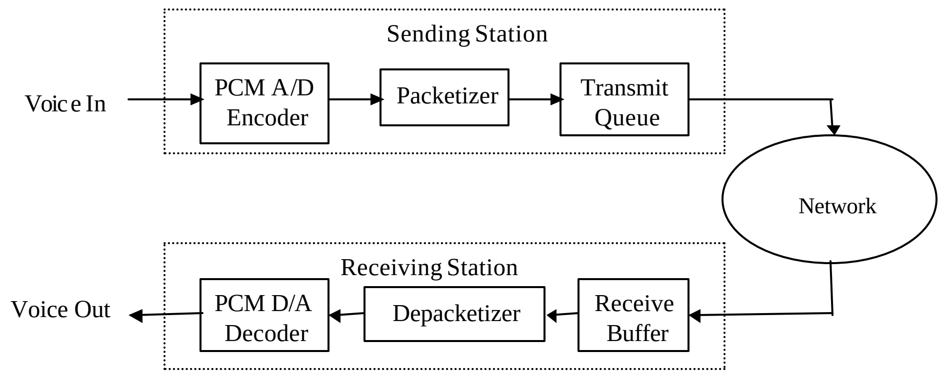

Figure 3-3 Block Diagram of a Station that encodes and sends a Voice Stream.

The voice cell delay time of end-to-end have to be in range of 250-600 ms. For voice communication (telephone), the information must be transmitted to the destination terminal in a transparent way also in ATM, as in the conventional line switching. Even if a little information is lost, the communication quality is not deteriorated. Consequently, the cell loss due to the buffer overflow (10^{-4}-10^{-3}) can be tolerated [1]. The information can be transmitted at a constant rate without fluctuation by the CBR service; a serve requirement is imposed that the end-to-end delay must be several milliseconds or less, excluding the transmission delay.

The voice is classified as voiced and unvoiced periods. In the voiced period, a cell is generated, in contrast, there is no cell generated in the unvoiced period. We can also call that the voiced period as a talkspurt and unvoiced period as a silent period. The voice source is represented by mean bit rate (MBR), and peak bit rate (PBR). Several models were introduced to model the burstiness and correlation characteristics of the cell arrival process from a voice source. The basic model is a periodic process alternating between a talkspurt and a silent period, as shown in Figure 3-4.

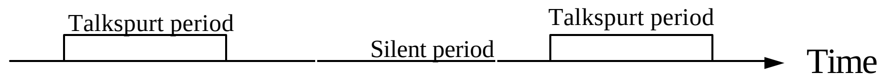

Figure 3-4 Single Voice Source Model.

Each period can be represented by an exponential distribution of means $1/a$ and $1/b$ respectively. The number of cells generated within the talkspurt period is then a geometric multiple of the cell length. Each voice source is sampled at 16 kHz and encoded using embedded PCM (Pulse Code Modulation). Hence, at a coding rate of 4 bits/sample, the source peak-rate is 64 Kbps. Let T represents the cell interarrival time, then the average arrival rate per source S (cells/sec) is given by equation (3-1).

$$S = \frac{(1/\alpha)}{T(1/\alpha + 1/\beta)} \qquad \text{--------------------------------------} \quad (3\text{-}1)$$

It is to be noted that the randomness introduced by the deterministic time is replaced by an exponential one. The first burstiness parameter set has $1/\alpha = 352$ ms and $1/\beta = 650$ ms which corresponds to a 35 % activity factor [38]. The source rate is 64 Kbps, then the cell's time is 6 ms. Table 3-1 shows the scheme used in North America (also used in Japan) plus the International (CCITT) standard [39].

(a) North American			(b) International (CCITT)		
Digital signal Number	Number of voice channels	Data Rate (Mbps)	Level Number	Number of voice channels	Data Rate (Mbps)
DS-1	24	1.544	1	30	2.048
DS-1C	48	3.152	2	120	8.448
DS-2	96	6.312	3	480	34.368
DS-3	672	44.736	4	1920	139.264
DS-4	4032	274.176	5	7680	565.148

Table 3-1 North American and International TDM Carrier Standards.

The talkspurt and silent periods can als o represented by the following:

$$y = 1 - e^{-\lambda t}$$

$(1- y) = e^{-\lambda t}$, take the logarithmic for both sides, we obtain

$$t = -(1/\lambda)\ \mathrm{Ln}\,(1 - y)$$

Where

(1-y): from 0.0 to 1.0

$(1/\lambda)$: mean value of period.

3-2-2 Video Traffic

The video stream is encoded according to the standard coding such as H.261 [40] or MPEG [41,42]. In both MPEG and H.261, the frame is divided into number of 16x16 “macroblocks”, and macroblock can be coded differentially with respect to the previous frame. Moreover, in MPEG, the coded differentially with respect to both the preceding and the following frame. Figure 3-5 shows the block diagram of a station that encodes and sends a video stream over a communication network [36,41,43]. A frame is taken in the video camera, and sent as an analog signal into the frame grabber, where it is digitized. Then, the encoder compresses it. A rate buffer its purpose is to smooth out the variations in the encoder’s output rate, and follows the encoder which producing a constant bit rate stream.

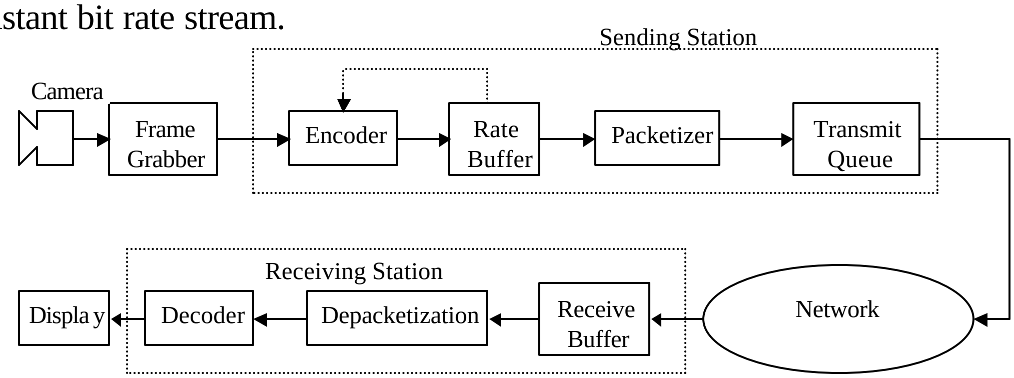

Figure 3-5 Block Diagram of a Station that encodes and Sends a Video Stream.

The buffer occupancy is used by the encoder as feedback to control the encoder output rate (and hence quality) so that the rate buffer doesn’t overflow or underflow. The constant bit rate stream is passed from the rate buffer onto the main memory through the system bus. In order to send streams over the network, the sending station packetizes streams (as cells). The cells are sent over the network to destination station. The destination station buffers the received cells to compensate for the delay variation due to the network. The contents of the buffer are passed to the decoder, which decompresses the stream and delivers it to the display. According to a coding standard such as H.261 or MPEG, a frame is composed of a number of Groups of Blocks (GOBs). Depending on the number of pixels in a frame, it is divided into either 3 or 12 GOBs. Each GOB is in turn divided into 33 macroblocks. A macroblock contains information for an area of 16 x 16 pixels and consists of three 'blocks', two for each color component and one for the luminance. A macroblock is the smallest unit that can be encoded/decoded without any future information. Also, in both H.261 and MPEG, a macroblock can be coded differentially with respect to the previous frame [41].

A delay constraint comes from the need to support interactive communications; it is well known fact that human beings can tolerate up to 200-250 ms of delay in two-way conversation. In the communications system using compressed video, there are two delays the first in the encoder and in the decoder that can be as high as 100 ms, as well as delays in the local networks to which the video stations are attached [44]. Therefore, a reasonable constraint for the wide-area component of the delay would be 40 ms. We applied two low quality compressed video stream such as 192 Kbps, and 384 Kbps (H.261), we have also applied high-quality compressed video stream such as 1.5 Mbps and 2 Mbps.

3-2-3 Data Traffic

The data traffic is a message arrived in specific distribution. The message comes in instant of time to be sent through the network. It is insensitive to the delay time, in other word, the data arrives to the destination at any time. The buffer size must be small as possible to reduce the cost.

Figure 3-6 Data Traffic Generation. At every Vertical Line, the Fixed Size Message Arrives

The data traffics arrival process is defined by two parameters. First parameter is message size, and the second parameter is the interarrival period time. Figure 3-6, depicts the data traffic generation. We suppose that the message size is fixed and the interarrival period time has exponential distribution with mean value such as 5 ms, or 10 ms.