VIDEO ON ATM NETWORKS

With the integration of digital video compression and network technology, networked multimedia applications, such as video on demand, videoconferencing, digital library, distance learning, and interactive visual communications are becoming more and more important. In these applications, compressed digital video is a major component of multimedia data (1,2).

The delivery of real-time video over networks has characteristics different from conventional data transport. For example, sufficient bandwidth for a guaranteed timely delivery of video information is required. The bandwidth required depends on the video compression algorithm and the desired video quality. For real-time, two-way communication, low delay (latency) is necessary. Long delay between users can make the communication ineffective. Low delay variation is another requirement for multimedia delivery even in one-way applications to avoid large buffers and possible problems in clock recovery. Another concern is the effect of transmission impairments. Because of the nature of compressed video, a single transmission error may propagate to subsequent frames and cause synchronization failures and severe degradation of visual quality at the receiver. The traditional error recovery method for data transport, such as retransmission, is not suitable because it causes too much delay. Therefore, special video coding and a forward error correction code is needed at the encoder, and error concealment techniques are needed at the decoder to minimize degradation of video quality.

The asynchronous transfer mode (ATM) is a cell-based, high-speed, networking technology supported by broadband integrated service digital network (B-ISDN) international standards. It is defined for operation over a number of physical media supporting bit rates ranging from megabits per second to gigabits per second. Because ATM is developed to support integrated services that include video, voice, and data, it provides the features necessary for supporting multimedia applications. ATM is emerging as an ideal networking technology for multimedia transport because of its high bandwidth, flexibility in bandwidth usage, low delay, low-delay variation, variable bit-rate capability, and guaranteed quality of service (QoS).

Although ATM was designed to support integrated services, it has some limitations which affect video transport. Being a cell-switched technology, it introduces packetization delay. When the network is congested, it may result in cell loss and cell-delay variation. In this article, we provide information related to video, ATM, and issues related to the transport of compressed digital video over ATM networks.

VIDEO FUNDAMENTALS

Analog Video

Video is a time sequence of two-dimensional frames (pictures). Each frame is represented by a sequence of scanning lines. As shown in Fig. 1, there are two ways of displaying or scanning a frame, progressive scan and interlaced scan. In the interlaced scan, a frame consists of two *interlaced* fields. In the progressive scan, a frame consists of only one field. A movie in a theater is in a progressive format. An analog TV signal is in an interlaced format consisting of 30 frames (60 fields) per second. The interlaced format conserves bandwidth because it sends only half a frame every 60th of a second, but because of the property of human eyes, gives the impression that we are viewing 60 complete pictures per second. However, interlaced scan results in motion artifacts when some part in the image moves between the two half-frames.

A color image can be represented by three component signals. The RGB color component system is one way to represent a color in the three primary colors. Red (R), Green (G), and Blue (B) components are combined to synthesize a color. Alternatively, a color can be represented using a luminance

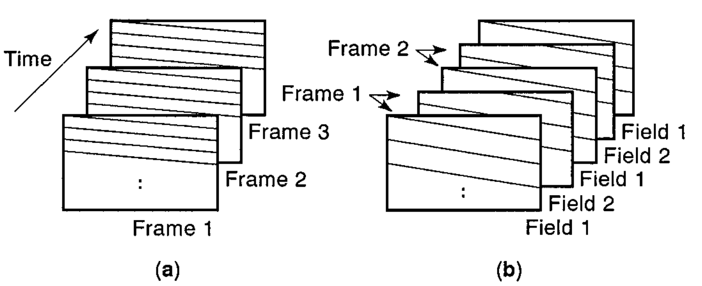

Figure 1. (a) Progressive and (b) interlaced scan: In the progressive scan, a frame consists of only one field. In the interlaced scan, a frame consists of two interlaced fields. The interlaced scan conserves bandwidth compared with the progressive format. However, the interlaced results in motion artifacts when some part in the image moves between the two fields.

Table 1. CCIR 601 Specification

	NTSC	PAL/SECAM
Luma sampling freq.	13.5 MHz	13.5 MHz
Chroma sampling freq.	6.75 MHz	6.75 MHz
Frames/second	30	25
Number of luma samples/line	858	864
Number of chroma samples/line	429	432
Number of active luma samples/line	720	720
Number of active chroma samples/line	360	360
Number of active lines/frame	486	576
Sample resolution	8 bits	8 bits
Data rate	167 Mbps	166 Mbps
Color subsampling	4:2:2	4:2:2

(brightness) signal and two chrominance (color) signals. Because the human visual system is less sensitive to color information than brightness information, chrominance signals can be represented with lower resolutions than luminance signals without significantly affecting the visual quality.

The current analog color TV standard used in North America and Japan was developed by the National Television Systems Committee (NTSC). The NTSC standard defined a YIQ color system. In this representation, Y is used for luminance, and I (in-phase) and Q (quadrature-phase) are two color-difference signals modulated by a 3.58 MHz color subcarrier. The luminance and the modulated chrominance signals are combined into a composite signal. Each channel of a TV signal has a video bandwidth of about 4.2 MHz, but requires a bandwidth of about 6 MHz to accommodate FM audio and channel separation. The phase alternating line (PAL) and the Sequential Color Avec Memoire (SECAM) standards used in Europe are based on a YUV color system. The YUV color system is similar to the YIQ color system except that the color-difference signals are defined slightly differently.

Digital Video

To use current state-of-the-art computers and digital networking technologies for processing, storing, and transmitting video signals, the analog video signal must be converted into a digital video signal. With compression techniques, digital video provides good quality video with a much lower bandwidth compared with that needed for analog video. With the much reduced bandwidth, many video applications become possible.

To convert an analog video into a digital format, each scanning line in the image is sampled, and the sampling points or pixels (picture elements) are represented by discrete values. CCIR 601 recommendation [International Radio Consultative Committee now changed to ITU-R (International Telecommunications Union-Radio)] (3) defines the format for digital television signals by digitizing NTSC, PAL, and SECAM signals. The number of active samples (pixels actually displayed) per line are specified to be the same in all systems even though the total numbers of samples per line differ. The important features of CCIR 601 are listed in Table 1.

The CCIR 601 defines a YCbCr color space which is a scaled and offset version of the YUV color space. Cb and Cr represent color difference signals of B and R from luminance signal Y. Because human eyes are not very sensitive to color signals, CCIR-601 specifies a sampling ratio of 4:2:2 between the luminance and the two chrominance signals to reduce the transmission rates of the Cb and Cr chrominance components. The 4:2:2 subsampling means that the color-difference signals Cb and Cr are sampled with half the sampling frequency of the luminance signal Y and for every four samples of Y, there are 2 samples of Cb and 2 samples of Cr. Figure 2 shows the sampling patterns of commonly used subsampling formats.

Video Compression

The digital video format in CCIR 601 results in a high data rate (about 166 Mb/s). Different applications may use different digital video formats which result in different uncompressed data rates. The uncompressed data rates of some common video formats are listed in Table 2. These data rates

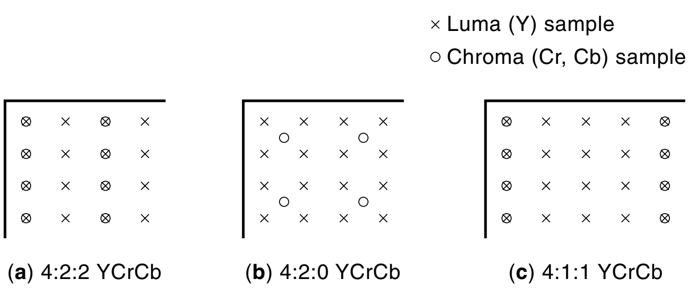

Figure 2. Examples of subsampling format: Subsampling is used to reduce the transmission rates of the Cb and Cr chrominance components. There are two samples of Cb and two samples of Cr for every four samples of Y in the 4:2:2 subsampling.

Table 2. Bandwidth Requirement of Broadband Services

Formats	Uncompressed Bit Rate, Mbps
CIF, 352 × 288 pixels (video conferencing) 4:2:0 30 frames/s	36
QCIF, 176 × 144 pixels (video conferencing) 4:2:0 30 frames/s	9
Digital TV (CCIR 601)	166
HDTV, 1280 × 720 pixels (high-definition TV) 4:2:2 30 frames/s	442
HDTV, 1920 × 1080 pixels (high-definition TV) 4:2:2 25 frames/s	829

are too high for transmission over low-cost networks. Video compression is a process for reducing the amount of data required to represent a video signal by removing spatial and temporal redundancies in the video signal. Spatial redundancies exist in a frame (i.e., between pixels) and temporal redundancies exist between frames.

Several video compression standards have been developed. JPEG (4) is a standard developed by the Joint Photographic Experts Group of ISO for still picture compression. It removes spatial redundancies using an "intraframe coding" technique. In JPEG, an image is partitioned into 8×8 blocks. Each block of 64 pixels is transformed into the frequency domain by the discrete cosine transform (DCT). The 64 DCT coefficients are quantized and coded by an entropy coding technique where more frequent coefficients are represented by shorter codewords. JPEG is also called Motion-JPEG when used for coding video, where each individual frame of the video sequence is coded by JPEG. With JPEG, a typical compression ratio of about 5 to 10 is achieved. Higher compression ratios are achieved with lower video quality.

For videoconferencing over a basic rate narrowband integrated service digital network (N-ISDN) which has only about 128 kb/s of bandwidth, the performance of JPEG is not adequate. H.261 (5) was developed for videoconferencing over ISDN at px64 kb/s where p ranges from 1 to 30. H.261 adds *motion-compensated* interframe prediction to the transform and entropy coding used in JPEG. To remove temporal redundancies, the prediction for each block includes a *motion vector* to indicate where in the preceding frame that block of the image is likely to have come from.

The Motion Picture Experts Group-Phase 1 (MPEG-1) (6) video coding standard was optimized for encoding entertainment video at about 1.5 Mb/s (mainly for CD multimedia applications). MPEG-1 uses a motion-compensation algorithm more sophisticated than that in the H.261 standard, allowing for both forward and backward prediction. Adding backward prediction increases the coding delay and the coder complexity, but improves the performance. The coding delay and the coder complexity are not very important for storage media applications where the encoding is done off-line and relatively few encoders are needed. MPEG-1 also allows motion vectors twice as precise as H.261. The MPEG-1 quality target is VHS video quality.

MPEG-2 (7) was developed for general higher quality applications which can afford higher bit rates. Typical MPEG-2 rates are from 3 to 12 Mb/s. It is also chosen by most countries for coding high-definition television (HDTV) at about 20 Mb/s. It handles interlaced video better than MPEG-1 by using adaptive field/frame coding modes. MEPG-2 defined several *profiles,* each targeted for a different type of application. For example, there is a simple profile (which is a subset of the other profiles) for a two-way communication application. Several scalable coding profiles were also defined. Scalable coding is also called layered coding. Using the MPEG-2 scalable coding algorithms, a video is coded in two or three layers. The base layer is decoded to get a video with a lower signal-to-noise ratio, lower resolution, or lower frame rate. The enhancement layers are decoded and added to the base layer to obtain high quality video. Scalable coding is considered important for transporting video over networks where the network is congested or the signal is heavily corrupted. In these cases, the base layer is better protected so that, if the network cannot support the enhancement layers, the user still obtains basic video quality by decoding the base layer. Besides the video coding part, MPEG standards also include several other parts covering audio coding, systems, conformance testing, etc.

The MPEG-2 video coding standard also adopted by ITU-T for broadband visual communication applications is designated as the H.262 standard. Based on the experience gained from MPEG-1 and MPEG-2, H.261 was later refined into a standard called H.263 (8) which is optimized for videophone transmitted on public switched telephone networks (PSTN) or local area networks (LAN). In H.263, four advanced coding modes were developed to improve the coding performance. H.263 is mainly for two-way communications. The coding algorithm limits the use of backward prediction to achieve short end-to-end delay which is important for two-way communication. With H.263, a reasonable quality videophone is achieved at about 20 to 40 kb/s.

MPEG-3 was originally intended for coding HDTV but later dropped because MPEG-2 is also suitable for coding HDTV. The main MPEG standard after MPEG-2 is MPEG-4 (9). MPEG-4 focuses on supporting video coding for a wider range of bit rates and applications emphasizing content-based interactivity, compression, and universal accessibility. In the MPEG-4 model, every frame of a video sequence consists of a number of arbitrarily shaped regions called video object planes (VOPs). Figure 3 shows the overall structure MPEG-4 encoding and decoding. The coding handles rectangular regions, as in MPEG-1 and MPEG-2, and also areas that change in shape and position from frame to frame. MPEG-4 encodes

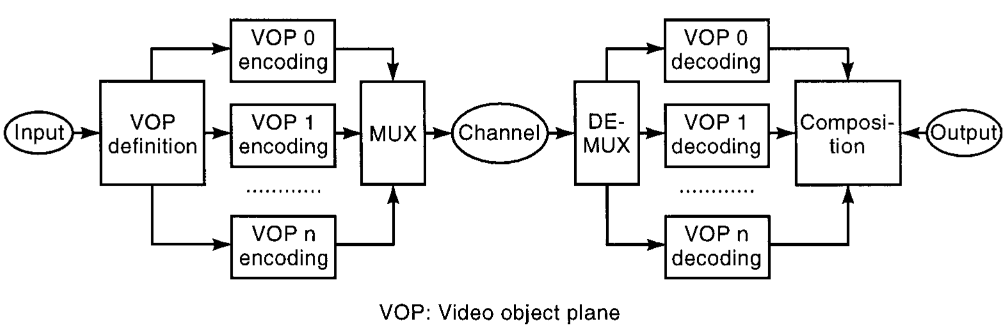

Figure 3. Encoder/decoder structure of MPEG-4: Input video sequence is segmented by VOPs, encoded separately, and multiplexed for transmission. At the decoder, each VOP is decoded after de-multiplexing and combined for displaying.

the shape, the motion and the textural information for each VOP. The information is stored or transmitted in separate VOP layers that are decoded separately. This enables many new applications, such as creating special effects and content-based retrieval.

ATM NETWORK

ATM

ATM is a cell-based technology supported by B-ISDN international standards (9a). All types of traffic are transmitted after dividing the data into small fixed-size packets called ATM cells. ATM is defined for operation over a number of physical layers ranging from megabits per second to gigabits per second. It supports ATM connections offering services from a few bits per second to nearly the capacity of the underlying physical layers. It supports both constant bit-rate (CBR) traffic and variable bit-rate (VBR) traffic. Compared with conventional packet switching, ATM offers low-delay, high throughput, guaranteed quality of service (QoS), and bandwidth on demand.

An ATM cell is 53 bytes long and consists of a five-octet header and a 48-octet payload field. The cells are switched for transporting to the destination based on the information in the cell header. The switching speed is maximized with a fixed cell structure, predefined paths, and no link-to-link error recovery. Cell sequential integrity is preserved in the cell delivery. The payload field length is relatively small to limit packetization delay and queuing delay in switching, as required by real-time interactive multimedia applications.

Two kinds of interfaces were defined in ATM, the user-network interface (UNI) and the network-network interface (NNI). The UNI specifies the interface between an ATM switch and user systems. The NNI specifies the interface between two ATM switches. The UNI and NNI have slightly different cell formats and the same cell length, as shown in Fig. 4.

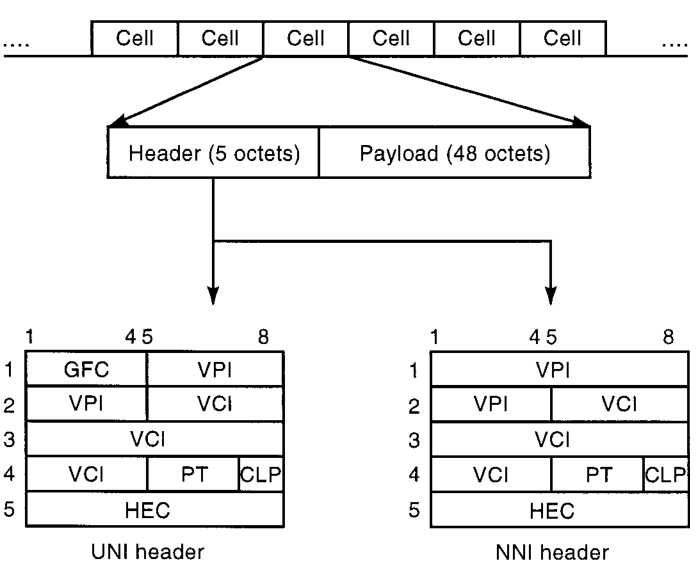

Figure 4. ATM cell structure: An ATM cell consists of a five-octet header and a 48-octet payload. There are two types of headers defined for the UNI and NNI.

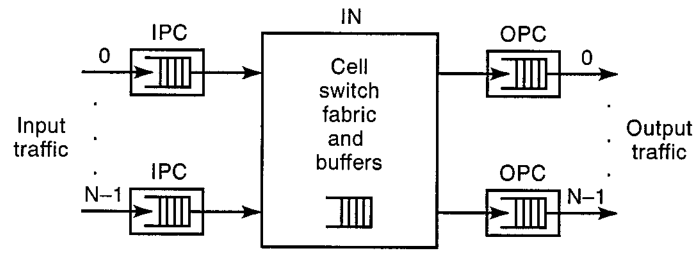

Figure 5. An $(N \times N)$ reference ATM switch: The IN is responsible for routing cell payload, traffic multiplexing, and congestion monitoring. When the network experiences congestion, the IN drop lower priority cells, and even high priority cells upon extremely adverse network congestion.

Present at the UNI only, general flow control (GFC) field controls the amount of traffic entering the network. GFC has only local significance, which means that the information in this field is not carried end-to-end. A possible usage of GFC is arbitrating cell transfer between several terminals sharing a UNI. An ATM virtual channel is identified by the combination of a virtual path identifier (VPI) and virtual channel identifier (VCI). Each virtual channel is associated with a VCI. A virtual path is a collection of virtual channels. Based on the VPI and VCI, the ATM switches route the virtual circuits across the ATM network. The values of VCI and VPI are valid only in the local link. At each ATM switch, these values are reassigned. A 3-bit payload-type (PT) field indicates the payload type of a cell, such as a user-reserved cell or a maintenance cell. The cell loss priority (CLP) bit allows assigning two different priority classes to ATM cells. The CLP determines whether or not a given cell should be dropped by the network during periods of congestion. This explicit loss priority is set by the source node or the network. The last byte in the header is header error control (HEC), a cyclic redundancy check (CRC) byte for detecting and correcting header errors.

In Fig. 5, an (N × N) reference ATM switch model is illustrated. The ATM switch consists of N input port controllers (IPC), N output port controllers (OPC) and an interconnection network (IN). The IN is responsible for routing cell payload, traffic multiplexing, and congestion monitoring. The port controllers and IN contain buffers for storing ATM cells temporarily to route and multiplex cells. If the network experiences congestion (i.e., buffers are full), the IN drops lower priority cells and even high priority cells upon extremely adverse network congestion.

The ATM layer structure (10,11) is shown in Fig. 6. The ATM layer contains the five-byte ATM header described previously. The ATM adaptation layer (AAL) carried in the payload is responsible for acting as the interface between a user application and the ATM layer. The AAL is subdivided in two sublayers, the segmentation and reassembly sublayer (SAR) and the convergence sublayer (CS). The SAR sublayer supports segmentation and reassembly operations. The CS depends on the particular service and supports different functions, such as clock recovery and data structure recovery. The service-specific convergence sublayer (SSCS) has particularly service-dependent functionalities, and the common part convergence sublayer (CPCS) provides services, such as padding and CRC checking. Different combinations of SAR sublayer and CS sublayers provide different service access points to

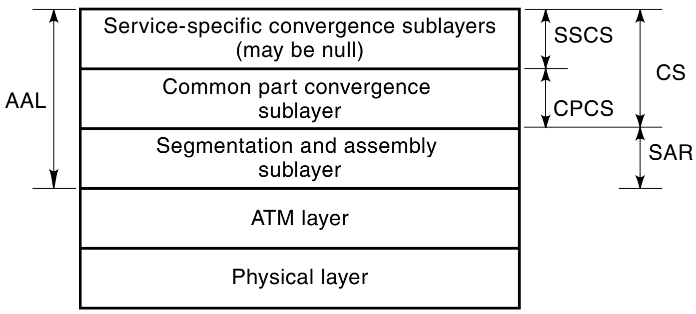

Figure 6. ATM layer structure: The AAL consist of the SAR and the CS, and is responsible for acting as the interface between a user application and the ATM layer. The SAR supports segmentation and reassembly operations. Different combinations of SAR sublayer and CS sublayers provide different service access points to the layer above the AAL.

the layer above the AAL. In some applications, the SAR and/ or CS are empty.

Connection between two end systems over ATM networks is established via different ways. Connection that is established by the end users on demand is referred to as Switched Virtual Circuits (SVC) and is set up by a signaling protocol. Through the signaling protocol, the ATM switch creates and configures an entry of routing table for data transfer. When a switch receives a cell, it examines the VCI field in the header and matches an appropriate entry in its table and transfers a cell to the appropriate output port. For fast switching operation, a Permanent Virtual Circuit (PVC) may be used to connect two end systems. PVC does not carry signaling procedure. Instead, an entry of the routing table for PVC is preconfigured and statically mapped. However, a failure of any link that PVC crosses results in the failure of the whole connection because it is statically preconfigured. In Soft PVC (SPVC), upon detecting a link failure, the ATM switch on the UNI will automatically establish a new connection via different links.

To accommodate various multimedia services, four types of AALs have been defined to provide service specific functions (e.g., clock recovery and cell loss recovery). AAL1 was developed for constant bit-rate circuit emulation services. AAL2 was recently developed for applications which require a short delay (e.g., audio). AAL2 combines minicells of multiple users' information in a single ATM VCC to provide high multiplexing gain. Minicells, short length packets which carry users' information have variable lengths of 1 to 64 bytes to accommodate various applications with minimal overhead. AAL3/4 and AAL5 are for connectionless and connection-oriented data services. AAL3/4 is specially designed for services, such as Switched Multimegabit Data Service (SMDS). SMDS is a public, packet-switched service that needs to exchange large amounts of data over the wide-area network. Currently, AAL5 is the most widely supported AAL and is used for video transport.

Quality of Service

Today's network is usually optimized for a single medium, such as data, voice, or video. Data, voice, and video have different QoS requirements. Unlike traditional best effort data services, real-time multimedia applications, such as videoconferencing require delivering the signal on a certain schedule or it becomes useless. ATM is connection-oriented, which means there is a call-setup phase before a connection is established. During the setup phase, a node requests certain QoS parameters for the connection (12,13). If the network does not have the resource to guarantee the QoS requested by the user, the call is rejected. If accepted, the QoS is usually guaranteed for the life of the connection. Modifying the QoS during the connection may also be possible and is a research topic.

ATM guarantees QoS for integrated services across both local and wide areas. Six QoS parameters are defined as a measure of network performance of an ATM connection (13). These can be divided into two groups, negotiated QoS parameters and nonnegotiated QoS parameters. Negotiated QoS parameters include maximum cell transfer delay (maxCTD), peak-to-peak cell-delay variation (peak-to-peak CDV), and cell loss ratio (CLR) and may be negotiated between the end system and the network. The nonnegotiated QoS parameters which are not negotiated but provide the information about the network performance, include cell error ratio (CER), severely errored cell block ratio (SECBR), and cell misinsertion rate (CMR). SECBR is defined as the ratio of severely errored cell blocks and total transmitted cell blocks. A cell block is a sequence of N cells transmitted consecutively on a given connection where N normally corresponds to the number of user information cells transmitted between successive operation and management (OAM) cells. A severely errored cell block occurs when more than a defined number of lost cells or misinserted cells are observed in a received cell block.

ATM can support different traffic types, such as constant bit rate (CBR), variable bit rate (VBR), available bit rate (ABR), and unspecified bit rate (UBR). Each traffic type delivers a different QoS (13). CBR is intended to support real-time applications requiring tightly constrained delay and delay variation. QoS parameters specified for CBR are maxCTD, peak-to-peak CDV, and CLR. For VBR, two types of services are defined: real-time VBR (RT-VBR) and non-real-time VBR (NRT-VBR). The end-to-end, delay-sensitive applications, such as interactive video VBR conferencing are categorized as RT-VBR traffic, whereas delay-insensitive applications such as VBR video on demand are classified as NRT-VBR traffic. MaxCTD, peak-to-peak CDV, and CLR are defined as the QoS parameters for RT-VBR, and CLR as the QoS parameter for NRT-VBR. In ABR services, an end system obtains a fair share of the available bandwidth according to a network-specific allocation policy and adapts its traffic by controlling the source rate in accordance with the feedback carried by a flow control mechanism. The QoS parameter for ABR is the CLR. UBR is also called the *best effort service.* Because UBR services offer no traffic-related service guarantees, no QoS parameters are specified.

ATM Traffic Management

The goal of ATM traffic management (13,15,16) is to achieve high network utilization and guaranteed QoS for different types of ATM services. Because ATM is connection-oriented, a logical/virtual connection set-up and call admission control (CAC) is required to reserve the necessary network resources before information transfer. If resources with the required QoS are not available, the connection is refused. For a guarantee of negotiated QoS, new connections should not affect

Table 3. Category of Traffic Descriptors for Different Services

Service Class	PCR	SCR	MBS	MCR	CDVT
CBR	x	N/A	N/A	N/A	x
RT VBR	x	x	x	N/A	x
NRT VBR	x	x	x	N/A	x
ABR	x	N/A	N/A	x	x
UBR	x	N/A	N/A	N/A	x

the QoS for any of the existing connections. During the CAC, the end system provides a list of parameters which represent intrinsic characteristics of the source traffic called a traffic descriptor. Based on the traffic descriptor, the network allocates appropriate resources for the requested QoS of the end system by negotiation. Therefore, the negotiated QoS is guaranteed only when the end system complies with the negotiated traffic contract, which consists of traffic descriptors and a set of QoS parameters. The traffic descriptor consists of two elements, source traffic descriptors and cell-delay variation tolerance (CDVT). The source traffic descriptors which describe expected bandwidth utilization include peak cell rate (PCR), sustainable cell rate (SCR), maximum burst size (MBS) and minimum cell rate (MCR). SCR is an upper bound on the average rate of the cells of an ATM connection, and CDVT is defined as a measure of cell-delay jitter. Table 3 shows the traffic descriptors specified for different service types.

Once a connection is established, the network uses a policing mechanism to monitor each source traffic to detect the violation of the traffic descriptors negotiated between the user and the network. A source violates the traffic contract for several reasons, such as inaccurate estimation of traffice characteristics either maliciously or unintentionally. Violation of traffic contract causes network buffer overflow or cell loss of other connections. To minimize the effect of QoS degradation on other connections, the usage parameter control (UPC) and the network parameter control (NPC) are used as the policing mechanisms for the source traffic at the UNI and NNI, respectively. The cell from a traffic source that does not follow the contract is either rejected or tagged for deletion upon network congestion by setting the CLP bit of an ATM cell to a lower priority. The ATM Forum has proposed a standard for traffic control using the UPC parameters, PCR, SCR, MCR, and burst tolerance (BT). Using these parameters, the user interface determines whether or not the source conforms to the connection contract at the peak rate or the sustainable rate and the burst size, while supporting the minimum cell rate for the service.

As shown in the Fig. 7, the traffic descriptor can be simplified provided that the traffic characteristics of the source are altered. Traffic shaping is a method of smoothing the burstiness, or the PCR, of traffic at the source or UNI, so that the negotiated traffic descriptor conforms as much as possible. It results in less congestion and delay on networks. Traffic shaping is optional in ATM network standards.

For congestion control in ATM networks, preventive or reactive control is possible. Preventive congestion control ensures that congestion cannot arise, for example, by negotiating with each traffic source during connection setup to allocate appropriate resource. Reactive congestion control is used in conjunction with ABR traffic. When the network detects the possibility of congestion, it provides a feedback to the traffic sources. The sources can reduce the rate of generating data to avoid undesirable cell loss. One possible way to provide this feedback is via an Explicit Forward Congestion Indication (EFCI) bit. The EFCI bit is defined in the PT field of the ATM cell header. The switch indicates the congestion onto cells passing through the network by EFCI bit. The receiving terminal then sends a special message back to the sending side to request to reduce the traffic of source.

VIDEO OVER ATM NETWORKS

One of the big advantages of ATM for transporting video is the high available bandwidth, which is up to several hundred megabits per second. Generally, higher bandwidth provides higher quality. Through virtual channels, ATM provides flexibility in bandwidth usage. ATM supports variable bit-rate traffic. Because compressed video has a variable bit rate, using ATM to carry compressed video reduces the requirement of the rate-smoothing buffer and the end-to-end delay. Through statistical multiplexing of multiple variable, bit-rate coded video, the network bandwidth is more efficiently utilized. ATM guarantees QoS. It is used to carry audio, video, and data for integrated multimedia services. It supports cell-

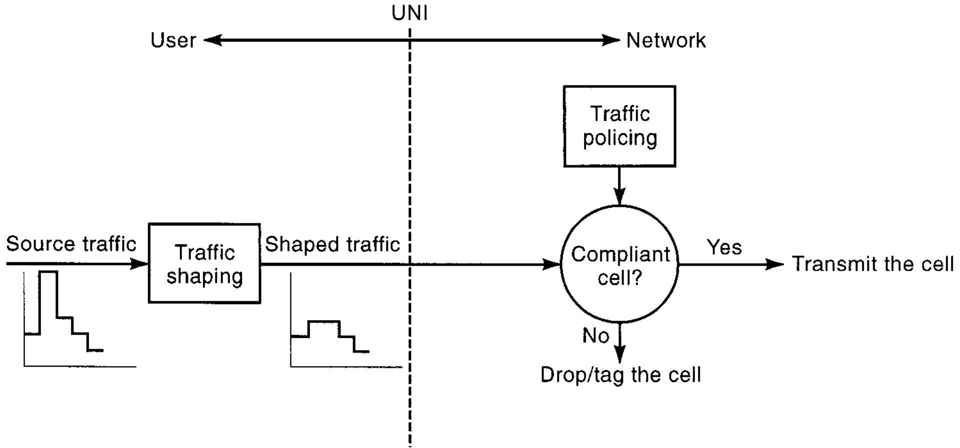

Figure 7. ATM traffic management: The traffic policing monitors the source traffic. If a cell is identified as noncompliant to a traffic contract, it is dropped or tagged as a lower priority. To comply with the traffic contract, the source traffic may be shaped.

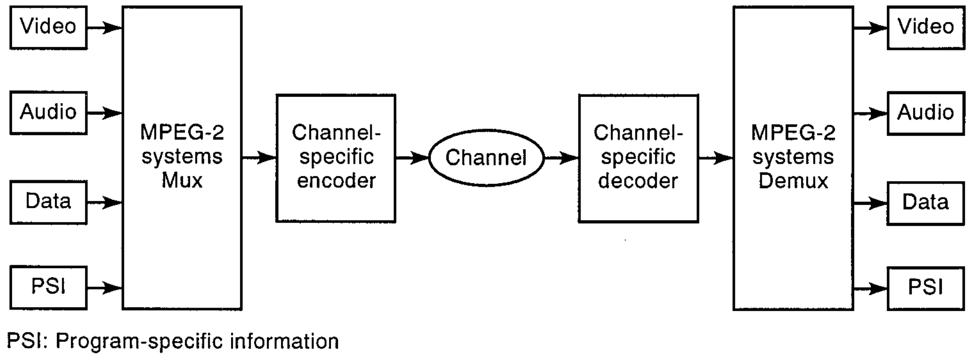

Figure 8. MPEG-2 end-to-end multimedia application: Each coded video, audio, data, and control streams are multiplexed into a single MPEG-2 system stream for storage and transmission. In the receiver side, the MPEG-2 system stream is demultiplexed before each coded stream is decoded.

loss priority and multipoint distribution in the network which are useful for video applications. However, the ATM network also has some limitations. The network causes cell loss, cell-delay variation, and packetization delay which are undesirable for real-time video applications. In the following, we discuss these issues in more detail.

MPEG-2 over ATM

Because MPEG-2 is for broadband video applications and the ATM network is for broadband communications, we first use MPEG-2 video as an example to discuss various issues related to transporting video over ATM networks.

A block diagram of an end-to-end multimedia application using the MPEG-2 System Multiplexer is shown in Fig. 8. To transport multimedia information over networks, MPEG-2 adopts a scheme involving two levels of packetization (17). The coded video, audio, data, and control streams are called elementary streams (ES). Each ES contains only one type of medium (e.g., audio, video, data, and control). ES's are first packetized into packetized elementary streams (PES) which consist of PES packets. Each PES packet can contain a variable number of coded bytes from one and only one ES. Multiple PES can be multiplexed into a single MPEG-2 systems stream for storage and transmission. Two types of MPEG-2 system streams are defined, program stream (PS) and transport stream (TS). Both stream definitions are packet-oriented multiplexes optimized for specific applications.

The PS is intended for a relatively error-free environment. The length of a packet is relatively long and variable. The TS is suitable for a relatively error-prone environment. The TS packets have a fixed length of 188 bytes (four types of packet header and 184 bytes of payload). This length was chosen on the basis of considerations for encryption (184 bytes is a multiple of eight bytes which is the block size of popular encryption algorithms), error correction needs (a length not greater than 255 bytes is desirable for Reed–Solomon codes), and ATM adaptation ($188 = 4 \times 47$ which fits into four ATM cells and still leaves four bytes for AAL functions if needed).

When transporting packetized video over ATM, a suitable AAL layer between the video stream and the ATM layer should be chosen. Because AAL1 is for constant bit-rate circuit emulation, it is suitable for real-time multimedia applications. However, because AAL5 is widely used, the ATM Forum has proposed AAL5 for transporting MPEG-2 TS packets. The default mode is to map two MPEG-2 TS packets into eight AAL5 cells. ITU-T has proposed AAL1 and AAL5 for transporting MPEG-2 TS packets (14). The mapping of MPEG-2 TS packets into AAL5 and AAL1 cells is shown in Figs. 9 and 10, respectively. A detailed comparison of AAL1 and AAL5 is in (21).

The network performance with respect to bit errors and cell loss is one of the factors considered during the specification of network adaptation. The requirement of the bit-error rate (BER) and cell-loss ratio (CLR) depends on the bit rates of the applications and on the user required QoS. Some examples of BER/CLR requirements for various audiovisual services are shown in Table 4.

Experimental results showed that performance is enhanced if errors are detected and corrupted data are not passed to the elementary stream decoders. The CRC32 error-detection capability of AAL5 supports this. In the ITU-T H.310 standard (18), when using AAL1 to carry video, one of the following three options is used: no FEC, RS(128, 124) without interleaving, and RS(128, 124) with interleaving. The actual choise of an option is carried out as a part of the capability exchange. It is expected that, as field experiences increase, the most appropriate option will be established. When using AAL5 to carry video, there is no cell loss or bit error-correction capability, because error detection and conceal-

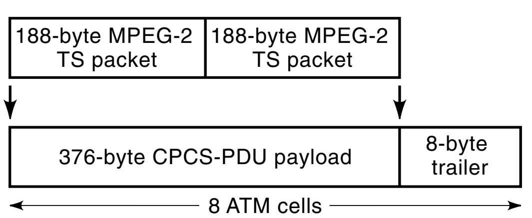

Figure 9. AAL5 mapping of TS packet: AAL5 can be used transporting MPEG-2 TS packets over ATM networks by mapping two TS packets into eight ATM cells.

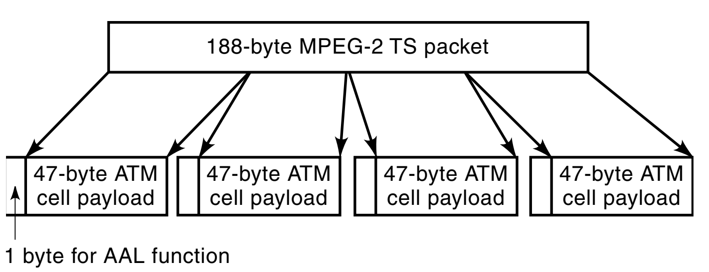

Figure 10. AAL1 mapping of TS packet: AAL1 can be used for transporting MPEG-2 TS packets over ATM networks by mapping one TS packet into four ATM cells.

Table 4. Example of BER/CLR Requirements for Various Audiovisual Services

Service	Nominal Bit Rate	User Required QoS	Required Max. BER/CLR Without Error Handling
Videophone (H.261)	64 kbps–2Mbps	30 min error-free	le-6/le-7 FEC in user level
MPEG-1 (audio included)	1.5 Mbps	20 min error-free	4e-10/le-7
MPEG-2 *VCR-quality*	4 Mbps	15 min error-free	BER < 3e-10
MPEG-2 postproduction	15 Mbps	1 h error-free	BER < 2e-11

ment were considered sufficient. The damaged part of the picture is replaced, for example, by the corresponding part of the previous frame.

Another concern for transporting video over ATM networks is the effect of cell-delay variation or jitter caused by ATM cell multiplexing. The most significant effect of cell-delay variation is on the reproduction of clocks in the decoder. Jittered clock references cause residual jitter in the reproduced clocks, which affects the reproduced audiovisual and other data there (e.g., a variation of reproduced color due to a jittered color subcarrier) or requires a long time for lock-in of the reproduced clock. Different methods of jitter reproduction may need to be implemented in the terminals. In MPEG-2 Systems, Program Clock References (PCRs) are samples of the time reference generated by the encoder system clock. The PCR time-stamps are transmitted regularly to the receiver in order to synchronize the decoder system clock. For this timing recovery scheme to work, the channel has to have constant delay so that the exact timing information carried by the PCR time-stamps can be preserved. In ATM networks, however, cells carrying the PCRs may experience cell-delay variation. This will cause the timing information carried by the PCRs to be inaccurate. In practical applications where the cell-delay jitter is small, the effect may not be noticeable. Depending on the applications, if the jitter is not acceptable, extra smoothing may need to be implemented to provide relatively jitterless PCRs for the timing recovery.

ITU-T ATM Multimedia Terminal Standards

ITU-T H.310 (18) and H.321 (19) provide the technical specifications of audiovisual communication systems for ATM networks. Specifically, H.321 adapts narrowband visual telephone terminals H.320 (20) to the ATM environment. H.321 maintains the overall structure and as many components of H.320 as possible, for example, H.261 as video coding specifications and H.221 for multiplexing audio, video, and data. On the other hand, H.310 supports the MPEG-2, H.261, and H.222 for multiplexing multimedia data. As shown in Fig. 11, a generic ATM audiovisual communication system within the scope of H.310 consists of terminal equipment, network, a multipoint control unit (MCU), and the constituent elements of the terminal equipment. H.310 defines two classes of unidirectional terminals: receive-only terminal (ROT) and send-only terminal (SOT). Bidirectional terminal types are defined on the basis of the ATM adaptation layer capabilities: RAST-1 (which supports AAL1), RAST-5 (which supports AAL5), and RAST-1 and -5 (which support AAL1 and 5). For detailed information, the readers are referred to the standards (18).

Variable Bit-Rate Coding and Statistical Multiplexing

From the viewpoint of network traffic management, the CBR video stream needs only simple bandwidth allocation because of its constant bit rate which is based on the peak cell rate (PCR). With the PCR, a simple multiplexing scheme of several CBR traffics is achieved guaranteeing the required delivery of cells. However, video quality fluctuates from time to time because some video frames are more difficult to code and require a higher PCR than that assigned to achieve the same quality as other simple video frames. Furthermore, the transmission of CBR traffic requires that the local buffer in the transmitter smooth the bit rate over a period of time to the assigned bit rate. The usage of this buffer causes extra delay. The VBR video coding has several advantages compared with the CBR. VBR traffic supports more uniform constant video quality because bit-rate generation is flexible (28,29). Ideally, it eliminates the necessity of the buffer in the encoder for bit-rate smoothing (in practice, a small buffer is still required to comply with the traffic contract with the network). Because ATM networks support VBR, it may be used for transporting VBR coded video. VBR video coding can produce better video quality, use less amounts of buffers, and result in shorter end-to-end delay. Rate control of VBR video coding conforming to the traffic descriptors and QoS parameters is still an active research area.

With VBR video, the network resource is utilized more efficiently through statistical multiplexing (30,31,32). In ATM networks, bit streams from different sources are multiplexed for transmission. Statistical multiplexing refers to the fact that the multiplexed bit stream is the sum of individual, statistically distributed, VBR source traffic. For example, the bit rate of a multiplexed bit stream from different CBR sources equals the sum of the peak bit rate of individual source traffic. However, different VBR traffic can be multiplexed with a lower bit rate than the sum of the average bit rate of individual source traffic because peak traffic usually takes a small amount of time compared with all of the traffic and it is unlikely that peak traffic occurs at the same time for all sources. Such statistical multiplexing of VBR source traffic over ATM networks provides statistical multiplex gain (SMG) which results in a total assigned bandwidth approaching the sum of the mean bandwidth of each source rather than the sum of each peak bandwidth, provided that a sufficiently large number of sources are multiplexed. If the network knows about the statistical characteristics of a source, an optimal path which achieves SMG while minimizing cell loss can be selected and used to transport the source traffic. Joint rate con-

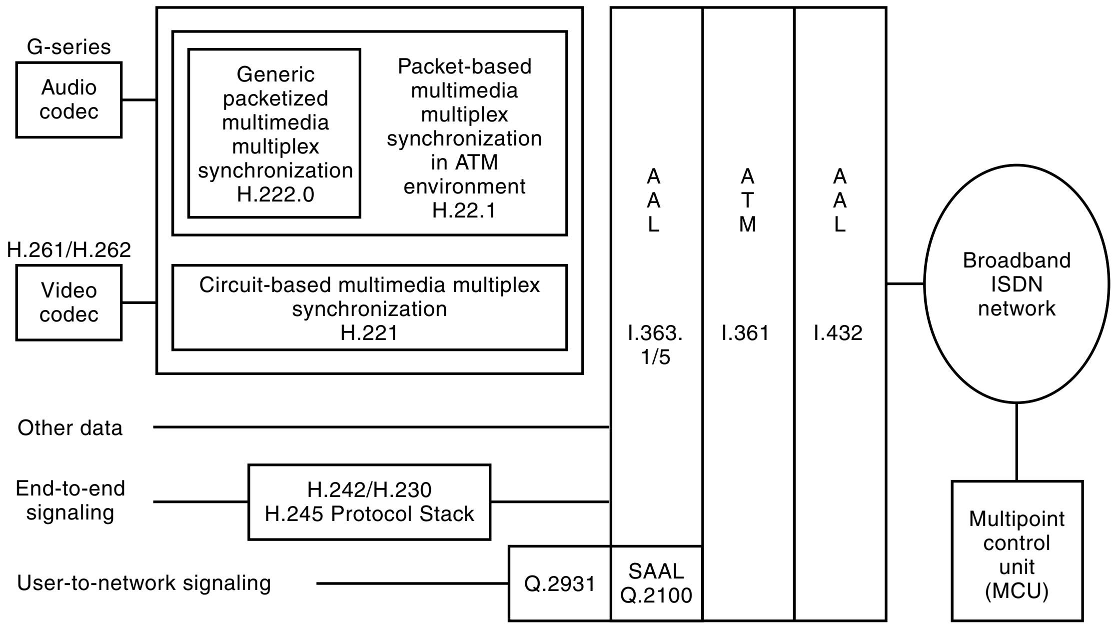

Figure 11. H.310 broadband audiovisual terminals: ATM audiovisual communication system within the scope of H.310. H.310 supports MPEG-2 and H.261 for video and H.222 for multiplexing multimedia data.

trol of VBR coding for multiple video sources to achieve high SMG and conforms to the ATM network constraints is also an active research area. This is particularly important for transporting multiple video channels over satellites where the bandwidth is relatively limited.

Scalable Coding

In scalable coding, encoder encodes the video in an ordered set of bit streams including one base layer and one or more enhancement layers (22). The decoder decodes subsets of the bit stream to reconstruct video with different quality. The minimum subset that can thus be decoded is the first bit stream in the set which is called the base layer. Each of the other bit streams in the set is called an enhancement layer. When addressing a specific enhancement layer, lower layer refers to the bit stream which precedes the enhancement layer.

Different applications require different video quality. Furthermore, the transmission networks consist of a heterogeneous mix of subnetworks with different qualities of service. To support the capability of handling the wide range of QoS, it is desirable to provide coded video for a range of different visual qualities and for a range of available network QoS. Simply encoding a separate stream for each combination of visual quality and QoS leads to multiple bit streams. Layered coding provides a more efficient way to encode a single representation of the source material that may be decoded and presented at a range of quality levels. A decoder can choose to decode a particular subset of these layers to scale the video to particular quality. The video is encoded at a low quality to form the base layer. The residual information is encoded as one or more enhancement layers. Decoding the base layer on its own results in a low-quality decoded sequence. The quality of the decoded sequence is improved by decoding the base layer together with the enhancement layers, as shown in Figure 12.

Scalable coding is also used to improve the performance in the presence of transmission errors (23). The base layer is transmitted with a high priority and protected with a more powerful error correction code resulting in a low error rate, whereas the enhancement layers are less protected and transmitted at a lower priority resulting in a higher error rate. This provides graceful degradation in the presence of errors. In ATM networks, the CLP bit sets different priorities for the base layer and the enhancement layers. Another advantage of scalable coding is the support of flexible transmission through heterogeneous networks consisting of interconnected subnetworks with different qualities of service. If a particular subnetwork cannot cope with the high bit rate of the complete

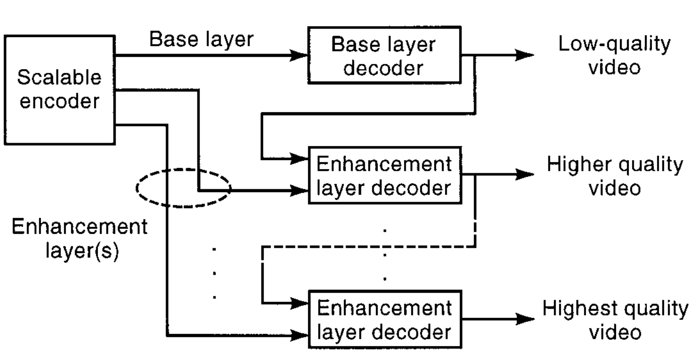

Figure 12. Layered coding: The video is encoded at a low quality and low bit rate to form the base layer. The residual information is encoded as one or more enhancement layers. Decoding the base layer results in basic quality. The quality of the decoded sequence is improved by decoding the base layer together with the enhancement layers.

video stream, then it should be possible to scale the video by extracting only the lower layers for distribution over that subnetwork. In this way, a video sequence is distributed over different networks with a range of qualities and bit rates. Scalable coding is also useful for video browsing applications, where only the base layer is decoded for fast browsing.

In MPEG-2, there are four different kinds of scalability defined: signal-to-noise ratio (SNR) scalability, spatial scalability, temporal scalability, and data partitioning. For SNR scalability, DCT coefficients of the encoder side are coarsely quantized to generate the base layer with a low bit-rate stream of low quality. The enhancement layers are produced by applying refined quantization to the residual error difference of the original video and the base layer. Spatial resolutions of the video sequence can be scalable using spatial scalability. In this scheme, scalability refers to the ability to modify the video resolution. Temporal scalability enables encoding a video sequence at a number of different temporal resolutions or frame rates. Both the base and enhancement layers have the same spatial resolution. The enhancement layer enhances the temporal resolution of the lower layer and, if temporally remultiplexed with the lower layer, provides the full frame rate. Finally, data partitioning is a technique that splits the DCT coefficients of each block into two layers, called partitions. Each partition contains a subset of the DCT coefficients. The lower spatial frequency coefficients are included in the base layer, and each enhancement layer consists of higher spatial frequency. A combination of more than two different types of scalability is called hybrid scalability.

In MPEG-4 where each video object is individually encoded and then compressed, another kind of scalability called content scalability is defined. In content scalability, the base layer contains basic or important video objects, and the enhancement layer contains additional objects. Content scalability allows users to manipulate the objects in a video.

ATM networks are well suited for transporting scalable video because of the VCI and VPI structures and the support of CLP. Scalable video coding is still an active research area.

Error Control

Because of the nature of the variable-length coding and interframe coding, compressed video is very sensitive to transmission errors. A single bit error in the video stream causes error propagation, synchronization failure, and results in severe degradation in video quality.

When a compressed video stream is transmitted over ATM networks, a cell loss means the loss of a large chunk of video data. Without special error recovery schemes, the video quality is very objectionable under the cell-loss situation. Layered coding discussed in the previous section helps the situation. However, the probability of error in the base layer still causes problems. The TS of MPEG-2 systems provides some error resilience. A TS packet which contains more important information (e.g., video headers) can be sent in duplicate so that, if one packet is lost, the information is still recovered from the other packet.

Several error resilience and concealment techniques have been proposed to minimize the effect of transmission error on video quality (23–27). Traditionally, error correction in packet-switched data networks is performed with an automatic repeat request (ARQ) protocol, such as the transmission control protocol (TCP). With this form of error correction technique, it is possible to retransmit corrupted or lost data packets from the sender to the receiver. However, in many applications, including real-time videoconferencing, video transmission requires relatively low delay. Usually real-time video traffic requires a continuous stream of data to the decoder with a tight bound on delay and delay variation. Retransmission of lost data is not efficient or feasible because it causes too much delay. In these situations, adding redundant information into the original data maintains acceptable quality of service. Forward error correction (FEC) is one of the mechanisms for protecting data from transmission errors without using retransmission. With FEC, the sender appends redundant parity data to the original data before transmission, and the receiver uses this redundant information to detect and recover corrupted or lost data.

Current video coding standards provide two mechanisms, syntactic and semantic, for detecting an error in the received video stream. In syntactic detection, the error portion is easily detected by investigating the codeword to verify its legality. With semantic detection, decoding more than 64 DCT coefficients in a block is detected as an error in the video stream. For recovery of synchronization with the arriving bit-stream, unique codewords are defined in most current video coding standards. These codewords have unique bit patterns. Within an error-free bitstream, the synchronization codeword is not allowed to occur at places other than a synchronization point. When transmission error causes synchronization failures, the decoder jumps to the next synchronization codeword. This assures that the resynchronization for the decoder continues the decoding process with minimal degradation of video quality.

Another important issue for error control of coded video streams is error concealment. Error concealment is the technique for estimating the lost or corrupted block caused by transmission error to minimize the degradation of video quality at the decoder. In temporal concealment techniques, a lost or corrupted block is replaced with the block in the same location or is indicated by the motion vectors in the previous frame in the case of interframe coding. However, if there are high motion activities in the video sequence, it is difficult to estimate accurate pixels to replace the lost blocks. In spatial error concealment techniques, each lost pixel is estimated by interpolating spatially from the nearest undamaged pixels within the frame. When the impaired area is large, however, this method is not effective because of less correlation between the lost pixels and the pixels in a successfully decoded area. An effective cell-loss concealment technique is also the subject of continuing research.

BIBLIOGRAPHY

1. W. Verbiest, L. Pinnoo, and B. Voeten, The impact of the ATM concept on video coding, *IEEE J. Selected Areas Comm.,* **6** (9): 1623–1632, 1988.
2. G. Woodruff and R. Kositpaiboon, Multimedia traffic management principles for guranteed ATM network performance, *IEEE J. Selected Areas in Comm.,* **8** (3): 437–446, 1990.
3. CCIR Recommendation 601-2, *Encoding Parameters of Digital Television for Studios,* 1990.
4. ISO, *JPEG Digital Compression and Coding of Continuous Tone Still Images.* ISO 10918, 1991.

5. ITU-T Recommendation H.261, *Video Codec for Audiovisual Services at px64 kbits/s,* May 1992.
6. ISO/IEC 11172, *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mb/s,* May 1993.
7. ITU-T, ISO/IEC 13818-2, *Information Technology-Generic Coding of Moving Pictures and Associated Audio Information: Video,* Nov. 1994.
8. ITU-T Recommendation H.263, *Video Coding for Low Bit Rate Communication,* May 1996.
9. T. Sikora, The MPEG-4 video standard verification model, *IEEE Trans. Circuits Syst. Video Technol.,* **7**: 19–31, 1997.
9a. H. Dutton, *Asynchronous Transfer Mode (ATM) Technical Overview*, Englewood Cliffs, NJ: Prentice-Hall, 1995.
10. ITU-T Recommendation I.361, *B-ISDN ATM Layer Specification,* 1992.
11. ITU-T Recommendation I.362, *B-ISDN ATM Adaptation Layer (AAL) Functional Description,* 1992.
12. ITU-T Recommendation I.350, *General Aspects of Quality of Service and Network Performance in Digital Networks, Including ISDN,* Apr. 1992.
13. ATM-FORUM, *Traffic Management Specification Version 4.0,* April 1996.
14. ITU-T Study Group 13, *Integrated Video Services (IVS) Baseline Document,* March 1994.
15. ITU-T I.371 Draft Recommendation, *Traffic Control and Congestion in B-ISDN,* 1995.
16. H. G. Perros and K. M. Elsayed, Call admission control schemes: A review, *IEEE Comm. Mag.,* 82–91, Nov. 1996.
17. ISO/IEC 13818-1/ITU-T H.222.0, *Generic Coding of Moving Pictures and Associated Audio: Systems,* Nov. 1994.
18. ITU-T Draft Recommendation H.310, *Broadband Audiovisual Communication Systems and Terminals,* July 1995.
19. ITU-T Draft Recommendation H.321, *Adaptation of H.320 Visual Telephone Terminals to B-ISDN Environment,* July 1995.
20. ITU-T Recommendation H.320, *Narrowband Visual Telephone Systems and Terminal Equipment,* Mar. 1993.
21. S. Okubo et al., ITU-T standardization of audiovisual communications systems in ATM and LAN environments, *IEEE J. Selected Areas Comm.,* **15**: 965–982, 1997.
22. M. Ghanbari, Two-layer coding of video signals for VBR networks, *IEEE J. Selected Areas Comm.,* **7**: 771–781, 1989.
23. R. Aravind, M. R. Civanlar, and A. R. Reibman, Packet loss resilience of MPEG-2 scalable video coding algorithms, *IEEE Trans. Circuits Syst. Video Technol.,* **6**: 771–781, 1996.
24. H. Ohta and T. Kitami, A cell loss recovery method using FEC in ATM networks, *IEEE J. Selected Areas Comm.,* **9**: 1471–1483, 1991.
25. Q. F. Zhu, Y. Wang, and L. Shaw, Coding and cell-loss recovery in DCT-based packet video, *IEEE J. Selected Areas Comm.,* **3**: 248–48, 1993.
26. M. Ghanbari and V. Seferidis, Cell-loss concealment in ATM video codecs, *IEEE Trans. Selected Areas Comm.,* **3**: 238–247, 1993.
27. L. H. Kieu and K. N. Ngan, Cell-loss concealment techniques for layered video codecs in an ATM network, *IEEE Trans. Image Process.,* **3**: 666–677, 1994.
28. M. Nomura, T. Fujii, and N. Ohta, Basic characteristics of variable rate video coding in ATM environments, *IEEE J. Selected Areas Comm.,* **7**: 752–760, 1989.
29. M. Marafih, Y. Q. Zhang, and R. L. Pickholtz, Modeling and queuing analysis of variable bit rate coded video sources in ATM networks, *IEEE Trans. Circuits Syst. Video Technol.,* **4**: 121–127, 1994.
30. B. Maglaris et al., Performance models of statistical multiplexing in packet video communications, *IEEE Trans. Comm.* **36**: 834–843, 1988.
31. H. Sidhu and S. Jordan, Multiplexing gains in bit stream multiplexors, *IEEE/ACM Trans. Networking,* **3**: 785–797, 1995.
32. H. Saito, M. Kawarasaki, and H. Yamada, An analysis of statistical multiplexing in an ATM transport network, *IEEE J. Selected Areas Comm.* **9**: 359–367, 1991.

JEONGNAM YOUN
MING–TING SUN
University of Washington

VIDEOPHONE. See VIDEO TELEPHONY.