Networking Constraints in Multimedia Conferencing And the Role of ATM Networks

Radhika R. Roy

This paper deals with two closely related issues:

- Characteristics of multimedia conferencing, and the typical bandwidth requirements for audio, video, and data, including text, still images, and graphics.
- Results of an investigation examining how an end-to-end asynchronous transfer mode (ATM) network architecture can satisfy the stringent requirements of multimedia conferencing.

The delay within the network has been calculated to estimate the upper- and lower-delay budgets for the video codec, to ensure that the end-to-end transmission delay is not perceptible to users. For some service categories, bandwidths have been estimated that will satisfy these nonperceptible delay limits, mean opinion score (MOS) performances, and response-time requirements for real-time multimedia conferencing.

Introduction

The past several years have witnessed a rapid growth of multimedia conferencing, due to falling prices of equipment and transmission.[1,2] The business case[3] for multimedia conferencing has improved significantly in recent years, and will continue to improve in the future. Significant factors propelling the acceptance of multimedia conferencing in the marketplace include:

- A reduction in equipment and transmission costs,
- Savings in travel time and contract negotiation time, and
- An increase in productivity.

Emerging standards, such as the Motion Pictures Experts Group (MPEG), the Joint Pictures Expert Group (JPEG), and px64 coding standards for *N x* 64-kilobits/s compressed video, are displacing proprietary algorithms and facilitating multimedia communications. Organizations, such as the Interactive Multimedia Association (IMA) and the Multimedia Communications Forum (MMCF), are promoting interactive multimedia applications and communications and, thereby, reducing barriers to the widespread use of multimedia technology. The availability of video chip sets has provided the core technology needed for building multimedia desktop computers. Several factors will enable users to communicate naturally, and with little or no transmission impairments perceptible to the user:

- Availability of multimedia workstations (WSs) equipped with high-quality audio and video codecs, and display capabilities,
- Increasing use of the embedded base of unshielded twisted pair (UTP) cable for high-speed bit rates (50/100/155 Mbits/s), and
- Emergence of local area networks (LANs) supporting audio, video, and data.

Characteristics of Multimedia Conferencing

Multimedia conferencing provides real-time communications among multiple sites through the transfer of multiple data. Multipoint communications may be one-to-many (multicast), many-to-one, or many-to-many. The conferencing transmission environment consists of a video space on a screen, for the participants' images; a collaborative information space for reference material, such as text, still images, graphics, and video; and, of course, the audio signals.

The communication among multiple, geographically separated sites often includes diverse network facilities. Users at different locations share visual applications and interact by introducing images and pointing to, or annotating, text or images. Once an image is displayed, several sites may point or annotate at the same time. The quality of video is very important in providing realism for remote participants. Picture quality is primarily determined by the bandwidth available per view, as well as by the ability of the video codecs to exploit that bandwidth. This multipoint communication service has to be designed to support a high degree of concurrent interactivity, with acceptable real-time response, particularly over long-distance networks.

Multimedia conferencing requires that all users should communicate naturally, without perceptible audiovisual difficulties. Thus, the effects of end-to-end or round-trip delay on speech and video quality, as well as maintenance of both intra- and inter-media synchronization, are key considerations in designing systems for real-time conferencing. These performance criteria are influenced by human factors, such as conversational modes. Although several studies[4-8] show the delay effects of speech quality, very few studies[9,10] have been made to obtain a subjective assessment of audiovisual quality due to delay introduced in an actual communication network. Emerging video compression standards[11-13] indicate that one-way end-to-end delay should be less than 150 milliseconds, with round-trip delay less than 300 milliseconds for conversational modes.

Panel 1. Acronyms Used in This Paper

ADPCM — Adaptive delta pulse code modulation
ATM — Asynchronous transfer mode
BER — Bit-error rate
CBR — Constant bit rate
CD — Compact disk
CLR — Cell-loss rate
Codecs — Coders/decoders
CIF — Common intermediate format
FDDI — Fiber distributed data interface
FM — Frequency modulation
IMA — Interactive Multimedia Association
JPEG — Joint Pictures Expert Group
LAN — Local area network
LD-CELP — Low delay-code excited linear predictive coding
MMCF — Multimedia Communications Forum
M/D/1 — Poisson arrival times/constant service times/single server
MOS — Mean opinion score
MPEG — Motion Picture Experts Group
OC3 — Optical carrier (155 Mbits/s)
PCM — Pulse code modulation
QCIF — Quarter common intermediate format
T3 — Transmission facility (45 Mbits/s)
UTP — Unshielded twisted pair
VBR — Variable bit rate
VC — Virtual circuit
VP — Virtual path

Audio

Audio with a full-duplex interruptability feature, that is, the ability of participants to speak at any time, is the most important medium for multimedia conferencing. General human factor studies also indicate that, among all the media of multimedia conferencing, audio is the most critical. Depending on the application, the audio bandwidth can vary significantly. Table I shows current trends in audio coding, bit rates, and mean opinion score (MOS) performance characteristics.[14,15] MOS scores provide an objective comparison of subjective testing, such as users' perceived quality of network delay. Table I also includes achievable performances at reduced bit-rates for compact disc (CD) and frequency modulation (FM) stereo audio.[16]

The one-way, end-to-end delay for CD or FM stereo audio coders using proprietary algorithms may be about 85 milliseconds. The codec delay for pulse-coded modulation (PCM), or adaptive delta PCM (ADPCM), is negligible, while the delay for low delay-code excited linear predictive (LD-CELP) coding may be about two milliseconds.

As Table I suggests, the traffic characteristics of a variable bit rate (VBR) codec can be bursty. While an MOS of 4.0 or above may represent high-quality coding, an MOS of 3.5 represents a degradation in speech quality that is easily detectable by a user, but not bad enough to impede natural voice communication. It would, therefore, be appropriate to maintain an MOS of 4.0 or above for multimedia conferencing.

Table I. Trends in audio bandwidth and MOS performance

Coders	Bandwidth in KHz	Frequency band in Hz	Uncompressed bit rates in kbits/s	Trans-mission mode	Expected bit rates in kbits/s**			MOS
					Peak	Average	Peak/ average	
CD audio (propriatory algorithm)	20	10-20,000	1411.4-1536	CBR VBR	192 384	192 192	1 3	* *
FM stereo audio	15	20-15,000	1024-1536	CBR	128	128	1	*
Wideband audio (G.722)	7	50-7,000	128	CBR	64/ 56/ 48	64/ 56/ 48	1	*
PCM audio (Mu-law, G.711)	3.2	200-3,400	64	CBR VBR	64 64	64 32-21	1 2-3	4.3 *
ADPCM audio (G.721)	3.2	200-3,400	64	CBR	32	32	1	4.1
LD-CELP audio (G.728)	3.2	200-3,400	64	CBR	16	16	1	4.1

* Expected MOS may be between 4 and 4.5, but is yet to be supported by published results

** Some of the bit rates are compressed

ADPCM - Adaptive differential pulse code modulation
CBR - Constant bit rate
CD - Compact disc
CELP - Code excited predictive linear coding
FM - Frequency modulation
LD-CELP - Low delay CELP
MOS - Mean opinion score
PCM - Pulse code modulation
VBR - Variable bit rate

Video

Video signals are the most bandwidth-intensive traffic for multimedia conferencing. The actual video bit rates will depend on the following factors:

- The activity in the captured scene,
- The number of bits per pixel—for a given number of samples per line and a total number of lines per frame,
- The number of video frames transmitted per second,
- The chosen level of video quality, and
- The type of compression algorithms.

Many commercial codecs are available in the market to provide conferencing at different bit rates. Table II shows the trends in compressed-video bit rates and MOS performances[17] for videoconferencing using common intermediate format (CIF) and quarter CIF (QCIF). A CIF uses 360 x 288 pixels (358 x 288 pixels of active resolution), while a QCIF uses 180 x 144 pixels (178 x 144 pixels of active resolution).

An MOS of 3.5 will provide video quality between fair and good, that is, the perceptible impairment may or may not be slightly annoying. Thus, it would be appropriate to maintain an MOS level of 3.5 or above for multimedia conferencing.[18] Test results indicate that the available commercial codecs have the potential to maintain MOS levels of 3.5 to 4.5 at encoded bit rates of 0.768 to 1.5 Mbits/s for CIF video quality (360 x 288 pixels, non-interlaced, 4:1:1, 30 Hz) in constant bit rate (CBR) mode. But the back-to-back encoding-decoding delay will far

Table II. CIF/QCIF bandwidth and MOS performance

Video coding, resolution, and format	Frames per second	Encoding	Bits per sample	Transmission mode	Compressed video bit rates (Mbits/s)	MOS
360 *x* 288 pixels non-interlaced, *p x* 64 standards	30	4:1:1	8	CBR	0.768-1.5	3.5-4.5*
360 *x* 288/	10-15	4:1:1	8	CBR	0.384	3.3
180 *x* 144 pixels	10-15	4:1:1	8	CBR	0.256	3
non-interlaced	10-15	4:1:1	8	CBR	0.128	2.7
p x 64 standards	10-15	4:1:1	8	CBR	0.064	2

* Estimated

CBR - Constant bit rate   CIF - Common intermediate format   MOS - Mean opinion score   QCIF - Quarter CIF

exceed 150 milliseconds, the upper boundary before a user perceives the delay. Figures for the encoding-decoding delay of the commercial codecs have not been published.

A VBR codec can reduce the encoding-decoding delay, as well as maintain superior video quality, but at much higher peak bit rates. Table III shows some examples of video bandwidths and performances using VBR codecs[19] for conferencing. As the data in Table III indicates, the codec delay increases with the increase in buffer size, and vice versa. But the peak bandwidth decreases with the increase in buffer size. There would be no coding delay for the codec with negligible buffer at peak bit rates of 2.562 Mbits/s, but the encoding-decoding delay would be about 3 frames, or about 99 milliseconds, when some buffers are used at peak bit rates of 0.846 Mbits/s. Moreover, the traffic characteristics of a VBR codec can also be significantly bursty, that is, a peak-to-mean ratio of 3.43-10.69 for the examples shown in Table III. The MOS figures shown in Table III have been estimated, and their precise values need to be verified through experiments.

In addition to CIF, there are other video qualities that are considered by MPEG standards. Table IV depicts the trends in bandwidths for different video qualities.[20] These bit rates are estimated, taking into account the performance of the CBR codecs. Although available commercial codecs will have back-to-back encoding-decoding delay much higher than 150 milliseconds for the video qualities shown in Table IV, trends suggest it may be possible, in the future, to build CBR codecs with delays that are in the range of 70-110 milliseconds. Thus, the VBR codecs coding delays could be reduced to almost negligible levels from the current 40 to 150 milliseconds and higher. But the peak bit rates will be much higher than those shown in Table IV.

Data

Data traffic is generated during conferences through the sharing of text files, still images, and graphics among users or among computers and users. In a conferencing environment, the real-time response for data traffic is critical. The response time should be between one and two seconds[21] for data retrieval and transfer, and about 0.5 seconds for document browsing in a computing environment. Detailed studies have yet to be made, however, to find more precise values in real-time conferencing environments. The response time should be faster for real-time multimedia collaboration, in which multiple parties are participating on a collaborative work.

Data traffic cannot tolerate any bit errors, and recovery from errors has to be made before the time-outs of the upper-layer protocols. Table V provides some typical examples of bit-rate requirements for data traffic to meet the real-time response. Sometimes, data traffic sharing may also generate unpredictable computer-to-computer traffic.

In such an environment, it is quite possible that the peak-to-average traffic ratio may exceed 1000:1, rather than standard assumptions of 5:1 or 10:1.[22] It is important that a transport network should be capable of exploiting these highly bursty characteristics of data traffic to maximize the use of the bandwidth.

Table III. Video bandwidth and performance for VBR codecs

Video quality, coding resolution, and format	Scene sequence	Transmission mode	Encoder-decoder delay in frames**	Compressed video bit rates (kbits/s) Peak	Mean	Burstiness (peak/mean)	Mean opinion score (MOS)*
Low rate videoconferencing quality	Sequence consisting of a person listening and	VBR with negligible buffer	0	2562	239.6	10.69	4.5-5.0
360 x 288 pixels,	interspersing	VBR with	1	1400	239.6	5.84	4.0-4.5
non-interlaced	occasional	buffer**	2	934	239.6	3.90	3.5-4.5
4:1:1, 8 bits/sample	comments		3	847	239.6	3.53	3.5-4.5
30 HZ, p x 64 or MPEG-1 standards			4	822	239.6	3.43	3.5-4.5

* Estimated ** Delay increases with the increase in buffer size (one frame delay = 33 milliseconds)
MPEG - Motion Pictures Expert Group VBR - Variable bit rate

Performance Criteria

Multimedia conferencing requires that communications among participants must have a "natural feel" about it. The time is very critical between when a speaker moves or speaks and when that motion or sound is perceived by the other participants. If the delay is too long, conversation, particularly excited discussions, will become untenable as participants begin to interrupt and talk over each other.

The total one-way system delay, as noted earlier, should be less than 150 milliseconds, and round-trip delay is less than 300 milliseconds, to maintain the conversational, "face-to-face" nature of the applications.[11-13] Thus, it may be desirable to offer the customer a delay of much less than 150 milliseconds to provide nonperceptible delay, that is, delay that either is not perceptible to the user or is negligible enough not to be an annoyance. More objective and subjective tests must be performed in order to establish a precise limit for this delay.

The delay between audio and video transmission has to be limited to provide lip-synchronization of the audio and video signals. Lip-synchronization errors appear to be unobjectionable for video-to-voice lags if they are in the range of -90 to +120 milliseconds,[23] although much tighter limits are being considered in the standards, such as -20 to +40 milliseconds.[24] A real-time audio or video stream also requires intra-media synchronization. There should be negligible delay jitter for continuous media, such as audio and video. For example, delay jitter might have to be as low as 250 microseconds for a VBR video codec operating with negligible or no buffer[25] for multimedia conferencing. In practice, more delay jitter can be tolerated.

Video can tolerate more errors than data traffic. In general, a higher-quality video requires a much lower bit-error rate (BER) than a lower-quality video. Thus, each level of video quality may have a different BER requirement. More subjective testing should be performed to find a precise value of BER for different video qualities.

In the case of an ATM network, the requirement can be expressed in terms of cell-loss rate (CLR). The CLR objective should include cell losses due to both BER and network congestion. For example, the average period between cell losses may be 40 minutes for video application.[25]

While some of the parameters require further studies to determine precise values, the performance requirements for multimedia conferencing can be summerized as follows:

- One-way, end-to-end delay (including propagation, network and equipment) for real-time audio or video should be between 100 and 150 milliseconds.
- The MOS level for audio should be between 4.0 and 5.0.
- The MOS level for video should be between 3.5 and 5.0.
- End-to-end delay jitter should be very low, less than 250 microseconds in some cases, but a precise limit has yet to be determined.
- BER should be very low, such as one cell in error after every one hour in an ATM network. A precise value has yet to be determined.

Table IV. Bandwidth for different video qualities

Video quality, coding resolution, and format	Frames per second	Encoding	Bits per sample	Transmission mode	Compressed video bit rates in Mbits/s
Standard TV quality, 720 *x* 480 pixels, interlaced, MPEG-1 standards	30	4:2:2	8	CBR	4-9
Cable TV quality, 360 *x* 480 pixels, non-interlaced, MPEG-1 standards	30	4:1:1	8	CBR	2-4
Low rate video-conference quality, 360 *x* 240 pixels, non-interlaced, MPEG-1 standards	30	4:1:1	8	CBR	0.384-1.5

Note: Mean opinion score is expected to be 3.5 to 4.5 and higher, but results are required to be verified through experiments
CBR - Constant bit rate   MPEG - Motion Picture Experts Group

- The differential delay between audio and video transmission should be between no more than - 20 milliseconds to +40 milliseconds.
- Response time for object retrieval, such as a file, should be between one and two seconds, with a 0.5-second delay for document browsing.
- For data traffic, such as text, still images, and graphics, recovery from bit errors through re-transmission should occur before the time-outs of the higher-layer protocols.

Role of the ATM Network

Multimedia conferencing has unique traffic, performance, and connectivity requirements. They include transferring real-time audio, video, and data with guaranteed performance, as well as point-to-point, point-to-multipoint, multipoint-to-point, and multipoint-to-multipoint communications. Moreover, all media, except CBR-based systems, have bursty traffic characteristics. Existing transport networks, such as packet-switching or circuit-switched networks, cannot guarantee meeting all the requirements of multimedia conferencing, as well as saving bandwidth, by taking the advantage of bursty traffic.

The ATM transport network can transfer video, voice, and data traffic simultaneously. Furthermore, in providing the temporal and spatial integration of all objects required in multimedia conferencing—including voice, data, video, and imaging—the transport network should guarantee the performance of each medium. In addition to delay and bit-error-rate requirements, the maintenance of synchronization for real-time traffic is an important issue. The ATM network has the potential to transport both synchronous traffic, such as video and voice, and asynchronous traffic, such as data. ATM technology has been selected by international standard bodies as the basis for B-ISDN facilities. Essentially, ATM is a high-speed, virtual-circuit-oriented, cell-switching technology. The data rates specified for ATM interface with higher-speed systems, such as 2.4 Gbits/s and up, and lower-speed systems, such as 45 Mbits/s and less.

The ATM network can transfer real-time multimedia conferencing traffic with guaranteed performance through cell-switching.[25] The network can provide logical connectivity simultaneously to all end-points. Through statistical multiplexing of both real-time and non-real-time bursty traffic, the ATM network can provide a channel-sharing capability that is expected to offer economies of scale.

The optimization of multipoint connections (one-to-many, many-to-many), without consuming a large

Table V. Bit rates required for data

Data (text, still images, graphics) object size	Uncompressed object size in Mbits/s	Typical compression ratio	Typical response time in seconds: Retrieval & transfer of object	Typical response time in seconds: Document browsing	Peak bandwidth requirements in Mbits/s: Retrieval and transfer: Uncompressed	Peak bandwidth requirements in Mbits/s: Retrieval and transfer: Compressed	Peak bandwidth requirements in Mbits/s: Document browsing: Uncompressed	Peak bandwidth requirements in Mbits/s: Document browsing: Compressed
ASCII text, 8.5" x 11" page, (88 char/line x 55 lines x 8 bits/char)	0.029	2-4	2	0.5	0.015	0.008-0.004	0.058	0.029-0.015
8.5" x 11" color page (200 pixels/inch, x 24 bits/pixel)	90	10-20	2	0.5	45	4.5-2.3	180	18-9
Medium resolution, 8.5" x 11" color page (400 pixels/inch, x 24 bits/pixel)	359	10-20	2	0.5	180	18-9	700	70-35
High resolution, 8.5" x 11" color page (800 pixels/inch, x 24 bits/pixel)	1436	10-20	2	0.5	718	72-36	2,872	287-144
Graphics quality (1600 pixels/inch, x 24 bits/pixel)	5744	10-20	2	0.5	2872	287-144	11,488	1152-575

Note: Highly bursty traffic assumed
Char - Character

number of virtual circuits in an ATM network, requires further study.[26-28] A limited implementation could be achieved with a multicast server in each ATM switch, but a more efficient approach is to support multicast operation directly within the switch. Many-to-many connections are more complex to implement, but simple many-to-many connections can be implemented without great difficulty, especially if the majority of such connections remain within the campus or metropolitan area. Since multicast alone may not be sufficient to support flexible audio and video connections in a multipoint conferencing, bridging audio and video may be essential.

Many value-added services, such as format conversion, forming a composite video stream, customized video services, conference control, etc., could be provided using network-based bridges. These issues, however, are beyond the scope of this paper.

The ATM network introduces cell loss and cell-delay jitter. Cell loss may cause parts of the image to be shown incorrectly, or it may cause complete loss of picture synchronization. A cell loss of 10^{-9} is good enough for video services.[25] Cell loss and cell-delay jitter may have an effect, however, on both encoder-decoder synchronization and the buffer size.

The ATM network is able to establish a single virtual path (VP) between a source site and a destination site for a given multimedia call. A virtual circuit (VC) for each medium—audio, video, or data— also can be established through the same VP. Using the same path to route the traffic of each medium facilitates maintaining both intra- and inter-media synchronization, especially for real-time traffic, such as audio and video.

Moreover, an ATM network uses very high-quality transmission facilities, such as fiber optics, and

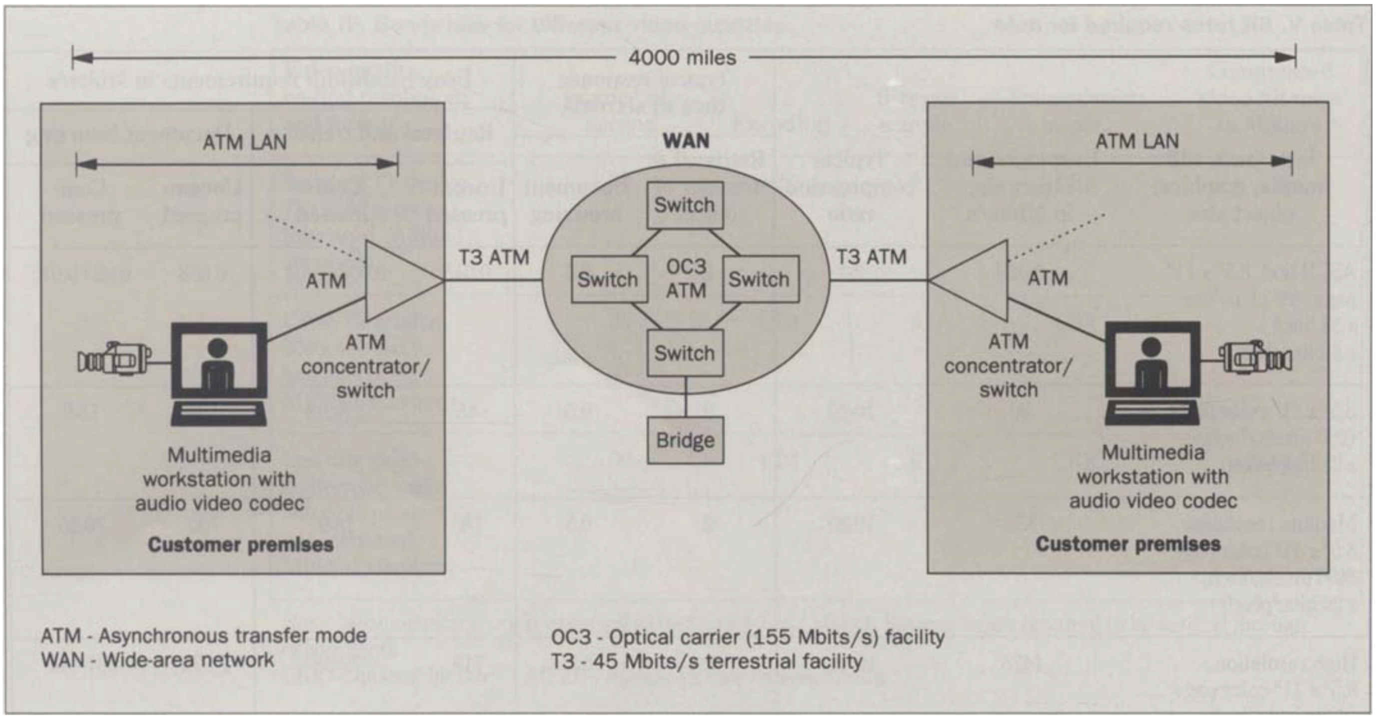

Figure 1. The end-to-end ATM configuration for multimedia conferencing includes both wired, cordless, and cellular voice and data terminals, with growing support for video; local, metropolitan and wide area networks; the local exchange, including loop plant, and interexchange lightwave networks; and customer services supplied by both the network and third-party vendors.

offers superior bit-error characteristics. An ATM switch handles multi-Gbits/s throughput, and each switch port handles 53/155/622-Mbits/s and 2.4-Gbits/s data rates. The ATM switch can be built to offer a very low cell-loss rate (much lower than 10^{-9}), and the cell delay jitter within the ATM network can be kept to a low value.[29]

Although ATM technology is intended for a wide area network (WAN), the ATM LAN[30] has recently emerged. It is composed of switches interconnected by 45/150 to 622-Mbits/s full-duplex links. These broadband LANs use ATM switching technology and provide:

- Higher aggregate bandwidth, much higher than a fiber distributed data interface (FDDI),
- Low latency,
- Low cell-loss probability,
- Higher availability, and
- Higher operational limit on the number of stations that can be attached on a single LAN.

These LANs are designed to carry both real-time and non-real-time traffic. Multimedia WSs equipped with cameras and codecs can easily be plugged into the ATM LAN, opening up opportunities to build end-to-end configurations based on ATM technology. These LANs can offer services for multimedia conferencing with superior performance.

ATM Network Delay

The end-to-end ATM configuration for multimedia conferencing is depicted in Figure 1. ATM LANs and an ATM WAN are used to connect the end-points. The multimedia WSs used for desktop conferencing are connected to the ATM LANs. The following assumptions are made for this configuration to provide a high-level estimate of end-to-end delay:

- 4000 circuit miles (the continental United States) between the conferees,
- 8.3 microseconds-per-mile propagation delay through the terrestrial link,
- T3 (45 mbits/s) access trunk speed and OC3 (155 mbits/s) backbone trunk speed,
- Six to eight hops between the source and destination ATM LANs,

Table VI. VBR video bandwidth and end-to-end delay

Video quality	Transmission mode	Mean opinion score (MOS)*	One-way end-to-end delay in milliseconds: Codec delay	Network delay	Total delay	Bit peak rates for single stream in Mbits/s	Upstream peak bit rates in Mbits/s: No. of streams	Total bit rates	Downstream peak bit rates in Mbits/s: No. of streams	Total bit rates
Low rate video-	VBR	4.5-5.0	0	75	75	2.56	1	2.56	2	5.12
conferencing	VBR	4.0-5.0	33	75	118	1.40	1	1.40	2	2.80
quality,	VBR	3.5-4.5	66	75	141	0.93	1	0.90	2	1.96
360 x 288 pixels	VBR	3.5-4.5	99	40	139	0.85	1	0.85	2	1.70
non-interlaced	VBR	3.5-4.5	132	40	172**	0.82	1	0.82	2	1.64
4:1:1, 30 Hz										

* Estimated
** Exceeds perceptible delay limit of 150 milliseconds
VBR - Variable bit rate

Other notes:
1. See Table III for additional explanations
2. Similar examples can be created with different qualities of video
3. Network delay is estimated at between 40-75 milliseconds

- An 80% utilization for concentrators/switches of the ATM LANs,
- An 80% utilization for access or backbone trunk, and
- A 53-byte (payload + header). ATM cell

The ATM switching delay includes the processing delay, which is assumed to be less than 200 microseconds, the queueing delay, and the transmission delay of the cell. The queueing delay in the ATM network is approximated using what is called the M/D/1 queueing discipline.[31] The end-to-end virtual connection is already considered to be established, and the call setup time is not included in this delay estimate.

The network bridge delay is expected to be minimal, especially if its function is to replicate and transfer the cells through logical connections in a one-to-many or many-to-many communication environment. In this situation, the delay within the bridge will be comparable to that of the ATM switch. However, some value-added services, such as a combination of different video streams or other value-added functions, may cause additional delay. More importantly, if the video streams originating from different sources are not synchronized to the same network clock, as much as one frame time may be required to recombine the video streams to provide customized services within the bridge.

Propagation delay appears to be the dominant component of network delay. The network bridge delay may also be significant in providing customized services, assuming that the delay may vary between a few milliseconds and one video frame—that is, 33 milliseconds. The estimated one-way end-to-end network delay—including equipment and circuit propagation delay, but excluding video or audio encoder-decoder delay—may vary between 40 and 75 milliseconds. So, the target delay budget for designing the video or audio codec should be in the range of 110 to 75 milliseconds, including both encoding and decoding functions, in order not to exceed the total one-way, nonperceptible delay limit of 150 milliseconds.

Bandwidth Requirements for Multimedia Conferencing

The bandwidth requirements for conferencing will be influenced by the following factors:

- Nonperceptible delay limit for real-time continuous media, such as audio and video. The allocation of video codec delay budget (assuming audio codec delay is negligible, or less, compared to video codec delay) and network delay budget should not exceed the nonperceptible limit for end-to-end round-trip transmission.
- Video application and its MOS level.
- Audio application and its MOS level.
- Number of simultaneous video streams.
- Data object size and its response-time requirement.

Desktop multimedia conferencing may have many video windows[32] showing the participants. The user may select the video mode to display a single participant, or to display many parties simultaneously, by dividing the available video space.

In the present analysis, only two simultaneous video windows have been considered in order to make a lower bandwidth estimate: one window displaying the participant who is speaking, or anyone selected by the

Table VII. Bandwidth requirements for multimedia conferencing

Service category	Media characteristics: Media	Media characteristics: Object	Transmission mode	Mean opinion score (MOS)	Delay	Upstream peak bit rates: No. of streams	Upstream peak bit rates: Peak bits in Mbits/s	Downstream peak bit rates: No. of streams	Downstream peak bit rates: Peak bits in Mbit/s	Total peak bit rates in Mbits/s (including all media): Upstream	Total peak bit rates in Mbits/s (including all media): Downstream
1	Audio	PCM audio (Mu-law, G. 711)	CBR/ VBR	4.3	Negligible codec delay	1	0.064	1	0.064	4.926	7.488
1	Video	360 x 288 pixels, non-interlaced 4:1:1, 30 Hz	VBR	4.5-5.0*	Negligible codec delay	1	2.562	2	5.134	4.926	7.488
1	Data	8.5" x 11" color page, 200 pixels/inch 24 bits/pixel	VBR	N/A	Less than 2 seconds response time	1	2.3	1	2.3	4.926	7.488
2	Audio	PCM audio (Mu-law, G. 711)	CBR/ VBR	4.3	Negligible codec delay	1	0.064	1	0.064	4.34	6.64
2	Video	360 x 288 pixels, non-interlaced 4:1:1, 30 Hz	VBR	4.0-4.5*	33 milliseconds codec delay	1	1.4	2	2.8	4.34	6.64
2	Data	8.5" x 11" color page, 200 pixels/inch 24 bits/pixel	VBR	N/A	Less than 2 seconds response time	1	2.3	1	2.3	4.34	6.64

* Estimated
CBR - Constant bit rate
PCM - Pulse code modulation
VBR - Variable bit rate

Other notes: 1. More service categories can be created by changing the objects of different media
2. See Tables I-VI for additional explanations

user, and one window comprised of composite video streams of all other participants. In addition, there can be a shared window where data files, information, or multimedia (data, audio, and/or video) documents and applications can be shared by all conferencees. Provision also can be made for a personal window, but the data from a personal window is local to the desktop and is not intended for sharing.

Table VI shows some typical examples for bandwidth requirements and delay components for low-rate videoconferencing video (360 x 288 pixels, non-interlaced, 4:1:1, 30 Hz). It is interesting to observe that if the network delay is 75 milliseconds, the codec delay should be kept below 75 milliseconds in order to keep the total delay from exceeding the 150-millisecond upper limit for nonperceptibility. Similarly, if the codec delay is increased to 100-110 milliseconds, the network delay has to be limited to 40 milliseconds (propagation delay is about 33 milliseconds). In this situation, any value-added service for video display should be planned very carefully, and processing delay should be negligible, or limited to a few milliseconds.

Table VII shows some estimated bandwidths for multimedia conferencing considering all media—audio, video and data—for both upstream and downstream. These bandwidths have been estimated assuming that these services will satisfy nonperceptible delay limits, MOS performances, and response-time requirements for real-time multimedia conferencing. Other examples can be similarly created, by taking different objects of each medium, for different service categories. In fact, bandwidths for multimedia applications can range from just a few hundred kilobits per second to multi-Gigabits per second.

Conclusions

End-to-end delay, MOS performances, and response times are the key performance parameters that determine:

- The design of video codecs,
- Multiplexing and switching equipment,
- The kind of transport network, and
- The amount of encoded bandwidth required for multimedia conferencing.

Because of its minimal delay for processing, queueing, transmission, and jitter, the ATM network will be able to meet the performance requirements for all media—as well as for their synchronization.

The ATM network also will have a very low BER and CLR to satisfy the stringent requirements imposed by all media in a real-time conferencing environment. The ATM network may have a one-way end-to-end delay within the United States from 40 to 75 milliseconds, including propagation and network equipment delay, but excluding the video encoder-decoder delay of the end-users' multimedia work stations. In order not to exceed nonperceptible delay limit, that is, the total end-to-end, one-way delay of less than 150 milliseconds, the video codec may have to be designed to offer delay between 75 and 110 milliseconds. A CBR codec offers higher delay, due to its relatively larger buffer size. The encoder-decoder buffer delay can further be reduced by use of a VBR codec with little or negligible buffer. On the other hand, a VBR codec will cause an increase in the peak bandwidth, and requires sophisticated controls, such as leaky-bucket mechanisms,[19] a congestion control technique.

The ATM network will have larger bandwidth-delay products, compared to traditional packet-switching networks. As a result, efficient transport protocols[33,34] are needed to achieve fast, high-performance throughput for bandwidth-intensive multimedia applications that consist of both a real-time, time-sensitive continuous media and non-real-time media. The issues related to transport protocols that meet stringent requirements of multimedia applications have not been addressed here.

The emergence of digital video compression techniques and transmission systems has generated a need for new objective performance measures. A very precise limit of nonperceptible delay for multimedia conferencing must be established through rigorous testing, as well as through greater participation in the standard-setting process. Moreover, the potential tradeoffs of video quality versus delay, bandwidth requirements, and willingness-to-pay must be determined to make multimedia conferencing services a success in the market place.

Acknowledgements

The author wishes to thank the following colleagues for their helpful comments on paper: B. H. Bharucha, R. V. Cox, W. H. Criss, J. Hartung, M. J. Karol, A. K. Kuthyar, S. K. Murthy, A. R. Reibman, and P. K. Verma.

References

1. B. R. Halhed and D. L. Scott, "There Really Are Practical Uses for Videoconferencing," *Business Communications Review (BCR)*, June 1992.
2. B. R. Halhed and D. L. Scott, "Market Trend," *Business Communications Review*, October 1992.
3. S. Richardson, "Videoconferencing: The Bigger (and Better) Picture," *Communications*, June 1992.
4. R. R. Riez and E. T. Klemer," Subjective Evaluation of Delay and Echo Suppressors in Telephone Communications," *Bell System Technical Journal*, Vol. 42, 1963.
5. J. W. Emling and D. Mitchell, "The Effects of Time Delay and Echos on Telephone Conversations," *Bell System Technical Journal*, Vol. 42, 1963.
6. G. K. Helder, "Customer Evaluation of Telephone Circuits with Transmission Delay," *Bell System Technical Journal*, Vol. 45, 1966.
7. A. G. Vartabedian, "The Effects of Transmission Delay in Four-Wire Teleconferencing," *Bell System Technical Journal*, Vol. 45, 1966.
8. N. Kitawaki and K. Itoh, "Pure Delay Effects on Speech Quality in Telecommunications," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 4, May 1991.
9. CCITT Study Group XII, Geneva, Switzerland, February 24 - March 3, 1992.
10. H. W. Frowein, G. F. Smoorenburg, L. Pyters, and D. Schinkel, "Improved Speech Recognition Through Videotelephony: Experiments with the Hard of Hearing," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 4, May 1991.
11. CCITT SGXV Working Party XV/1, Expert Group for ATM Video Coding," November 7, 1991.
12. E. A. Fox, "Advances in Interactive Digital Multimedia Systems," *Computer Magazine*, October 1991.
13. D. L. Gall, "MPEG: A Video Compression Standard for Multimedia Applications," *Communications of the Association for Computing Machinery (ACM)*, April 1991.
14. N. S. Jayant, "High-Quality Coding of Telephone Speech and Wideband Audio," *IEEE Communication Magazine*, January 1990.
15. Thomas C. Bartee, "Digital Communications," Howard W. Sams & Co., Indianapolis, IN, 1986.

16. C. V. Cox and J. D. Johnston, private communication.
17. R. Schaphorst and D. Bodson, "Subjective and Objective Testing of Videoconferencing/Videophone Systems," GLOBECOM'91, December 1991.
18. CCITT SGXV, Working Party XV/1, Expert Group for ATM Coding, Document AVC-109, September 24, 1991.
19. A. R. Reibman and A. W. Berger, "On VBR Video Teleconferencing over ATM Networks," GLOBECOM'92, December 1992.
20. A. Puri, "Video Coding using the MPEG-1 Compression Standards," SID'92, *Digest*, 1992.
21. B. Shneiderman, "Response Time and Display Rate in Human Performance with Computers," *Computing Surveys*, Vol. 16, No. 3, September 1984.
22. R. D. Gitlin, T. B. London, L. J. Greenstein, and S. R. Ahuja, "Lucky Net: An Overview," GLOBECOM'91, December 1991.
23. G. Karlson and M. Vetterli, "Packet Video and its Integration into the Network Architecture," *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 5, June 1989.
24. Proposal Package Description for MPEG Phase 2, ISO-IEC JTC1/SC2/WG111 MPEG91/100, August 23, 1991.
25. W. Verbiest and L. Pinnoo, "A Variable Rate Video Codec for Asynchronous Transfer Mode Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 7, June 1991.
26. S. Turner, "Design of a Broadcast Packet Switching Network," *IEEE Transaction on Communications*, Vol. 36, No. 6, June 1988.
27. P. Newman, "ATM Technology for Corporate Networks," *IEEE Communications Magazine*, April 1992.
28. S. C. Liew, "Multicast Routing Algorithms for 3-Stage Clos ATM Switching Networks," Conference Proceedings, GLOBECOM'91, December 1991.
29. T. Aoyama, I. Tokizawa, and K. Sato,"ATM VP-based Broadband Networks for Multimedia Services," *IEEE Communications Magazine*, April 1993.
30. E. Biagioni, E. Cooper, and R. Sansom, "Designing a Practical ATM LAN," *IEEE Network*, March 1993.
31. M. D. Prycker, "Asynchronous Transfer Mode: Solution for Broadband ISDN," Ellis Horwood, New York, 1991.
32. K. Wantabe, S. Sakata, K. Maeno, H. Fukuoka, and T. Ohmori, "Distributed Desktop Conferencing System with Multimedia Interface," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 4, May 1991.
33. A. N. Netravali, W. D. Roome, and K. Sabnani, "Design and Implementation of a High-Speed Transport Protocol," *IEEE Transactions on Communications*, Vol. 38, No. 11, November 1990.
34. T. D. C. Little and A. Ghafoor, "Synchronization and Storage Models for Multimedia Objects," *IEEE Journal of Selected Areas in Communications*, April 1990.

(Manuscript approved November 1993)

Radhika R. Roy *is a member of technical staff in the Interactive Information Services Architecture Department of AT&T Architecture Area at Bell Laboratories in Holmdel, New Jersey. He works on development of multimedia services and technology assessment. He joined AT&T in 1990. Mr. Roy received a B.S.E.E. from the University of Engineering & Technology, Dhaka, Bangladesh, an M.S.E.E. from Northeastern University, Boston, Massachusetts, and Ph.D. in electrical engineering from City University of New York, New York.*