

An Analysis of Universal Multimedia Switching Architectures

Radhika R. Roy
Ashok K. Kuthyar
Vanita Katkar

Distributed multimedia applications impose stringent and complex requirements on the switching and transport architectures of a network. An analysis shows that both synchronous transfer mode (STM) and asynchronous transfer mode (ATM) can meet the bandwidth and performance requirements of the emerging multimedia applications. ATM becomes the technology of choice, however, since it can:

- Transfer multimedia traffic efficiently, via its channel sharing capability,
- Employ statistical multiplexing using cell-switching to obtain high channel utilization, and
- Support connectivity efficiently in many-to-one, one-to-many, and many-to-many communications environments.

Introduction: A Confluence of Technologies

High-level cell-switching technology is becoming more cost-effective than traditional packet-switching, due to the:

- Integration of more functions into the hardware, and
- Falling price of hardware.¹

The present high-speed technologies permit bandwidth-intensive distributed multimedia applications to be switched cost-effectively and efficiently. For example, large-capacity storage disks and high-speed fiber-optic networks are making it possible to integrate diverse media, such as audio, video, and data.² This paper defines data to include text, still images, and graphics.

Most of us have seen the AT&T "You Will" television commercial in which a man on a tropical beach communicates via a portable workstation with his less fortunate peers back in their office. His workstation screen shows a two-way video capability, text and graphics areas, and a collaborative work area where changes made in a document by one participant are visible on all other participants' workstations.

Although apparently a very simple application from the user's point of view, the combining of audio, video, and data (called a multimedia application) places very stringent performance requirements on the network.

This is especially so if, as in the commercial, the applications involve real-time interactions between senders and receivers. These performance requirements include:

- Limits on voice and video signal delays, so that they are not perceptible to users;
- Synchronization of intramedia and intermedia signal streams;
- Low bit-error-rates (BER) to accommodate requirements, especially for data; and
- Very low probabilities of losing cells during transmission.

In this paper, emphasis is on how multimedia applications affect the network by presenting data streams with varying characteristics, and not on how the user interfaces with the applications.

Let's return to the commercial for an example. The video communications between the beach comber and his office mates would involve video encoders in each workstation. These encoders, which use bit compression techniques to reduce the bandwidth necessary to transmit a video signal, introduce a certain amount of intrinsic processing delay that would exceed any delay in a typical voice signal. Yet, the voice signals might also be encoded to compress the bit rate, introducing their own delay. Thus, the voice and video signals, each with different amounts of pro-

Panel 1. Acronyms Used in This Paper

ATM—Asynchronous transfer mode
BER—Bit error rate
FR—Frame relay
HDTV—High definition TV
MIS—Multimedia information service
MMC—Multimedia collaboration
MMM—Multimedia mail
SAC—Synchronous-to-ATM conversion
SMDS—Switched Multi-megabit Data Service*
STM—Synchronous transfer mode
VBR—Variable-bit rate
VCR—Video cassette recorder
VOD—Video-on-demand

cessing delay, have to be synchronized by the network before they are presented to the other participants.

Delay is also involved in any collaborative work in which electronic files have to be accessed by the workstations, formatted, and transmitted to all other participants. A certain amount of file search and processing delay is expected in this operation, but not so much that the natural rhythm of the conference would be adversely affected by delays.

Interactive data traffic, in which the participants react to each other in real time, also requires very fast response times. This is made all the more complicated for network management because the bandwidth for these applications vary between a few kilobits per second (Kbits/s), for terminal-to-terminal communication, to multi-gigabits per second (Gbits/s) for, say, video conferencing that involves collaborative work on text and high-resolution graphics.

The functional requirements for distributed multimedia applications also are very complex. For example, a simple point-to-point voice call can evolve, at the users' discretion, to become a multimedia call, or a multipoint multimedia call if additional participants are added on. Or a call may, in the extreme, exist without any connections. That is, since audio, video, and data connections can be established, then dropped and re-established without a new call setup being required, it is possible that a call can exist but be without any connections.

The requirements imposed by multimedia applications on the network are enormous, and pose tremen-

dous challenges for system engineers to design an optimal, yet cost-effective, global switching system. To list but a few of these challenges, the hardware architecture must provide high processing power to handle multimedia applications. A multimedia operating system must support multimedia data types, real-time scheduling of calls, and fast-interrupt processing, when a computer's central processing unit has to execute complex functions of multimedia applications. Storage and memory requirements include very high-capacity, fast access times, and high transfer rates. The high bandwidth, low latency, and low jitter required for multimedia applications also will necessitate new networking and protocol architectures.

Requirements for Multimedia Applications

To understand the basic switching requirements for multimedia networking, these applications can be categorized based on:

- Whether the delivery of media signals is real-time or non-real-time among the communicating entities, or participants;
- The types of connections between endpoints; and
- The symmetry of the traffic flow.

The major categories of multimedia applications are:

- Multimedia collaboration (MMC), where multiple participants located at different locations can work collaboratively;
- Multimedia information services (MIS), where individuals can access network-based or third-party databases;
- Video-on-demand (VOD), where subscribers can access any video presentation;
- Multimedia mail (MMM), where subscribers can send and receive enhanced versions of electronic mail that include video and audio storage capabilities.

All four categories are expected to be used by both business and residential customers.

The key attributes that differentiate each type of application are shown in Table I. These attributes will have a major impact on determining the appropriate multimedia networking architecture. It is clear that most MMC, MIS, and VOD will be real-time applications, while MMM, a form of electronic mail, will usually be a non-real-time application.

Multimedia Collaboration (MMC). Multimedia collaboration enables a number of participants to use vari-

Table 1. Attributes for differentiating multimedia applications

Multimedia application category	Media [audio, video, data (text, still images)] delivery		Connectivity			Traffic flow characteristic	
	Real time	Non-real time	One-to-many	Many-to-one	Many-to-many	Symmetric	Asymmetric
Multimedia collaboration	Yes		Yes		Yes	Yes	
Multimedia information service	Yes		Yes				Yes
Video on demand	Yes		Yes				Yes
Multimedia messaging		Yes	Yes	Yes		Yes	Yes

ous media in either real-time or non-real-time, while located at multiple sites.³ Multipoint communications can be one-to-many, called multicast; many-to-one; or many-to-many. Each participant will be able to send and receive audio, video, and data signals, and perform certain collaborative activities with other participants by using a multimedia workstation and other devices equipped with a built-in or attached video camera and audio capabilities. The multimedia workstation uses the concept of a virtual work space that, given the current level of terminal sophistication, consists of a space, or window, on a screen for the participants' video images; a collaborative information window for reference materials, such as text, still images, graphics, and other video applications; other function-specific or general-use windows; plus, of course, the audio signals.

This basic, though not particularly simple concept of a virtual work space is expected to become even more complex as computer technology continues to advance. Thus, any network plans being made today must take into account that multimedia applications are in their infancy, and are guaranteed to become more complex.

MMC is primarily used for real-time communications, but collaborative activities also can be done on a non-real-time basis. A real-time application would be a user adding or deleting multiple parties, or media, as desired during a call, or actively editing a file. An example of a non-real-time application could be a chess game car-

ried on via e-mail, with participants viewing, whenever they desired, the stored video signals of their opponents making each move.

Real-time multimedia also will have to support audio and video conferences in much the same manner that people participate in person-to-person conferences today. In addition to expected features, such as a graceful way to inter-

rupt a speaker, the technology will have to support such ancillary interactions as 'break outs' or sub-conferencing, 'whispering' or side conversations during the main conference, and e-mail note passing. In addition, conference recording storage capabilities will be required for MMC, including those features commonly found today on a VCR, including playback, pause, rewind, random access, slow-motion, and fast-forward.

Real-time multimedia collaboration also has to provide acceptable real-time responses between participants. The biggest performance challenge in MMC occurs when conference participants continuously transmit audio and video data streams. Whereas a face-to-face conference has a chairperson to moderate the participants, some sort of network protocol will have to be developed to act as a 'traffic cop' for the multiple data streams in a multimedia conference.

Other issues are being studied to ensure optimal communications architectures for multimedia collaboration, including such functions as the mixing of audio and video streams to provide both intramedia and intermedia synchronization continuously, and sharing applications between multiple users at different sites while still providing consistent views of all meeting places.

Multimedia Information Service (MIS). Multimedia information service gives end-users access to multimedia information. Like MMC, multimedia workstations or other devices are used to interact with multimedia servers and

navigate through a wide variety of information. A wide range of electronic information services could include such obvious examples as electronically accessible print news from newspapers, magazines, and journals; stored images of television news, medical information, real-estate services, and sales catalogues; ecological information and library services. Once multimedia capabilities are commonplace in the network, MIS will electronically bring together information seekers and content providers in as yet undreamed of applications.

Communications between an MIS end-user and the service application server is usually point-to-point, although with multiple end users accessing the service, it will appear to be point-to-multipoint from the server's perspective. In some situations, multiple servers will be involved for a given transaction if all the information is not available on a single server. The outputs of these servers, therefore, have to be coordinated to provide an integrated response to the end-user. The downstream bandwidth (from the server to the user) usually is much higher than that of the upstream bandwidth (from the user to the server). This on-demand interactive application can be characterized as point-to-point and multipoint, bi-directional, asymmetric service.

The MIS system must provide a number of functions, such as call and connection setup and control, a client-server architecture, the dynamic allocation of network and multimedia database resources, media synchronization, fast retrieval times, error-free transmission of textual data, graceful degradation under fault conditions, and, of course, the multimedia information databases.

Video-on-Demand (VOD). Video-on-demand is an interactive service that lets the user access a large selection of movies or video programs via a point-to-point connection. This connection gives the user individual and instantaneous control of the storage medium, in terms of program start, fast-forward, pause, slow motion, random access, and rewind actions. Indeed, the service should have at least all the features offered by a VCR today.

The set-up movie device attached to the set, along with the television monitor and the remote control, lets viewers connect to the video server and browse through a selection of movies or other contents, such as news stories, documentaries, software, do-it-yourself programs, or games. Multimedia workstations, high definition televisions, or other devices could be used for this application.

The primary characteristics of this service are:

- Retrieving the program content from the video server at the user's request;
- A very fast completion rate of the user's request;
- High-quality transmission for all media;
- The degree of interactivity being determined by the user, such as interactive VOD with a choice of instantaneous or delayed access;
- Instantaneous control of the video program by the user, including start, fast-forward, pause, random access, and rewind actions; and
- A very high quality of service, including a very low bit-error-rate (BER), very low jitter rate, and the maintenance of intramedia and intermedia synchronization.

This on-demand, bi-directional, asymmetrical service will require a very high bandwidth for downstream transmission from the server to the user, compared to that of the upstream transmission from the user to the server. Although the bandwidth required for the individual user to access and control the service may not be significant, the total volume of bandwidth generated by all users could be substantial to the network.

Multimedia Messaging (MMM). Multimedia messaging, or electronic mail, is a non-real-time messaging service that transfers multimedia information in a store-and-forward mode. The end-to-end communication characteristics of MMM can be point-to-point or multipoint, and bi-directional, asymmetric or unidirectional.

MMM is considered to be an extension of existing electronic messaging systems, with the addition of audio or video features in the body of the message. Although the use of MMM is in its infancy, it is expected that this application will be popular with both business and residential customers alike.

The performance requirements for multimedia mail are not stringent, compared to those for real-time multimedia communication, because MMM is a non-real-time, store-and-forward, communication system. The service should, however, support a unified mailbox, in which messages incorporating a variety of media can be stored, controlled, and accessed.

Summary of Multimedia Requirements. The key requirements for multimedia communications that should be supported by the underlying switching and transport systems can be summarized as follows:

- Bandwidth allocation for each medium;
- Performance parameters appropriate to each media, such as delay, BER/cell loss rate, and synchronization; and

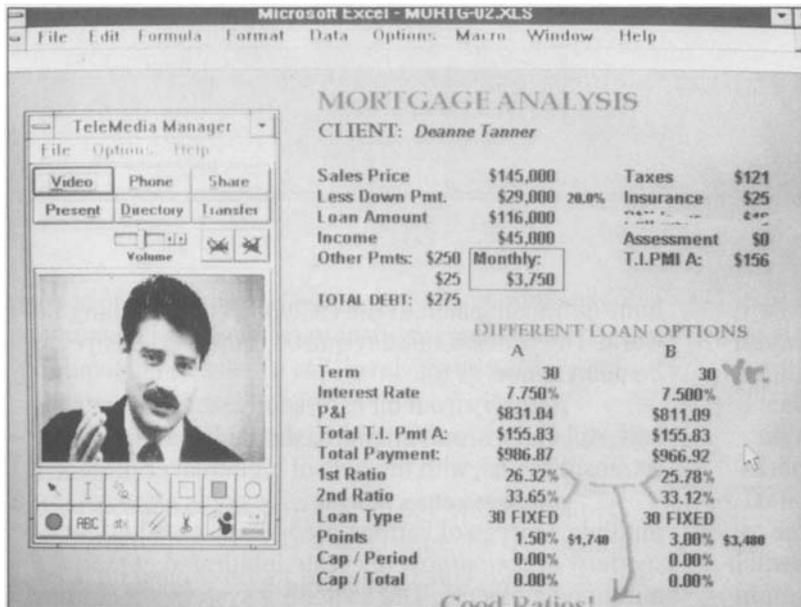


Figure 1. Multimedia applications involve two-way, real-time voice, data, and video data streams, including collaborative work areas where changes made by one conference participant can be viewed in real time by all others. Such applications place very stringent performance requirements on the network.

- Functional features, including call and connection control, routing, congestion control, quality of service (QoS) parameter control, and media bridging.

These multimedia applications have many more functional requirements than we have discussed, such as conference management, multimedia databases, multimedia storage, video server technology, user directories, security issue, and others. While important, these requirements have relatively minor influence on the selection of switching or transport architectures. As such, they are not addressed in this paper.

Multimedia and Switching Requirements

Multimedia collaboration imposes complex requirements for transferring real-time traffic within acceptable response times for point-to-point or multipoint, and multipoint-to-multipoint communications. There is a tremendous challenge to designing an optimal switching communication architecture that guarantees performance for all media. Although MIS, VOD, and MMM have certain unique characteristics, the requirements of these applications can be met by any switching systems that fulfill the requirements of multimedia collaboration.

It is reasonable to assume that the common characteristics of these applications constitute the fundamental requirements that must be met by any switching system. The switching system must be capable of handling a wide range of different bit rates and traffic parameters, as well as any functional requirements, in flexible and cost-effective ways. The requirements that have the greatest influence on the design of the switching characteristics are addressed in this section.

Bandwidth Requirements. The range of bandwidth requirements for an individual medium can be summarized³ as follows:

- Audio bandwidth - Uncompressed at 64 Kbits/s to

1.536 Mbits/s; compressed at 16 to 384 Kbits/s; burstiness (peak/average) at 1-3.

- Video bandwidth - Uncompressed at 3 to 166 Mbits/s (high definition TV not included); compressed at 56 Kbits/s to 35 Mbits/s (HDTV not included); burstiness (peak/average) at 1-10.
- Data (text, still images, graphics) - Uncompressed at 155 bits/s to 12 Gbits/s; compressed at 800 Kbps to 1.2 Gbits/s; burstiness (peak/average) at 3 to 1000.

The bandwidth of any multimedia application depends on a combination of media, and the total traffic may vary between a few Kbits/s and multi-Gbits/s. Thus, any switching system must support this wide range of different bit rates and traffic parameters.

Traffic Characteristics. The switching system of a network must be informed of the expected traffic characteristics of a call, when a connection is requested, in order to optimize the switch's resources and accurately predict its ability to maintain a certain level of performance.

The burstiness of traffic depends on two parameters: *peak* and *average* traffic. In addition to the burstiness of traffic from multimedia applications, *burst length* and the *variance of traffic* are also useful parameters. Statistical multiplexing is less effective with longer lengths of bursts⁴ and, thus, fewer active users can be supported for a given amount of bandwidth. Therefore, a set of traffic descriptors should be used within the network switch to accurately characterize the traffic parameters for each medium. Algorithms may have to be used to calculate the composite bandwidth required for a given multimedia call, based on the peak bit rate, peak-to-average bit rate ratio, variance of the bit rate, and the burst length of each medium.

From the amount of traffic that a switch must handle in order to establish a multimedia call, it becomes

apparent that a switching system needs to employ sophisticated control mechanisms to enforce bandwidth for the bursty sources.^{5, 6}

Performance Requirements. The performance characteristics of real-time multimedia collaboration are usually more stringent than for any other applications, such as MIS, VOD, and MMM. As noted, MMM is primarily a non-real-time application, and its performance requirements are the least stringent. If a switching system meets the requirements for real-time multimedia traffic, therefore, it can be expected that it will meet the requirements of non-real-time traffic, as well. This section describes the general characteristics for performance *parameters*³ that must be satisfied by the switching system.

Delay. The delay requirements of multimedia applications are the most important parameters to consider in designing the fundamental constraints of the switching architecture. Many delay constraints are required to be maintained.

For example, a *one-way, end-to-end* delay for a real-time isochronous medium, such as audio or video, should be between 100 and 150 milliseconds, including propagation, network, and equipment delay.

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.

The *differential delay* between audio and video transmission should be no more than -20 milliseconds to +40 milliseconds.

The *response time* for retrieving an object, such as a file, should be between one and two seconds, with a 0.5-second delay or less for document browsing.

The *recovery* of data traffic from errors through retransmissions should be done before the time-outs of the higher-layer protocols.

The *combination* of multiple video streams into a single composite video stream (for display in a single video window) should be done with negligible delay—if all originating video sources are synchronized to the same video clock. Otherwise, combining the signals should not take more than one video frame time.

The *frequency* of call setup, tear down, and redirection for audio and video should be less than that of a mouse click or keyboard stroke.

Bit-Error-Rate and Cell Loss. *Bit-error-rates* (BER) should be very low, such as one cell in error for every

hour of transmission in the case of a cell-switching network. The precise value required, however, has yet to be determined.

Recovery from bit errors for data traffic, such as text, still images, and graphics, should be done through retransmissions, with the help of higher-layer protocols.

Synchronization. Multimedia applications include multiple sources of various media, separated either *spatially* or *temporally*, to create integrated composite multimedia objects. The switching system should have the capabilities to ensure both intramedia and intermedia synchronization:

Intramedia synchronization is required to determine the rate at which events must occur within a single medium stream. The limit of delay jitter has to be maintained by the given media stream.

Intermedia synchronization determines the relative schedule of separate synchronization streams, such as the intermedia synchronization between audio and video streams. The limits of differential delay have to be maintained between the different media streams.

Functional Requirements. Multimedia applications have many complex functional requirements, and the switching system should be flexible enough to support those functions in order to optimize the performance of multimedia communications.

The *call control* and *connection control* are two such important functions that must be supported. Call control involves establishing the parameters of the call, such as the characteristics of the call and the types of end points to be connected. Connection control, applied after the completion of the call control stage, includes such functions as routing, congestion control, QOS parameter control, and media bridging.

Call and Connection Setup/Control. Distributed networked multimedia applications, especially MMC, require call setup modifications as required, call execution, and call release for communications between different users. It is clear from the characteristics of these applications that a call can no longer be viewed as a monolithic object, that is, a static point-to-point connection. Rather, it has to be managed as a complex interaction of multipoint communications that may change even *after* the call is set up.

The components of a composite multimedia call must be handled individually and dynamically. It becomes a prerequisite, therefore, to separate call and connection functions for controlling the various combina-

tions of connectivity and interactivity of a multimedia communications. Such combinations can include:

- Setup and release of a call, involving virtual circuits, that has no connection;
- Setup and release of a call and its connection(s) simultaneously;
- Modification of existing calls to add or drop connections, add or drop parties, or change the QOS parameters and bandwidth for heterogeneous communication between parties; and
- Context or persistency maintenance before, during, and after the communication session of a call.

Separating call and connection functions:

- Frees the switching resources from being tied up during call negotiation when the call is being established,
- Allows for independent routing connections in order to globally optimize the use of networking resources, and
- Permits flexible choices for providing switching resources.

Call control is viewed as a preparation to initiate a communication between the parties involved, setting the stage prior to the connection being established. The requirements for call control to support multimedia, multi-user applications can be summarized as follows:

- A call will have phases to establish, modify, execute, and terminate communications between different entities. All requests and negotiations for resources supporting the individual connections by a call can be done dynamically. It also may be necessary that a call be established or released with no connections, or that a call and its connections be established or released simultaneously.
- A call involves multiple media and multiple parties or entities, along with their specific characteristics and requirements. Therefore, the coordination functions for connections between parties or entities will be required in a call.
- A call contains relationships between users, media, and system resources, and their capabilities and compatibilities must be expressed to the network for a successful communication. A call between parties or entities also can be established in a heterogeneous communications environment. Either way, the use of these capabilities can also be augmented, modified, or deleted dynamically during a call.
- Each user can control the information of a medium

individually for an existing call.

- Intermedia relationships can exist in a multimedia, multi-user application call.
- Parties or entities can be added and removed dynamically at any time during a call. Parties will be notified of any changes.
- Different permissions to modify the call can be given to individual parties through authentication procedures. These permissions can be changed dynamically during the call to maintain ownership and call control rights. Services can be screened between the parties.
- A call will allow negotiations between different entities for system resources.

Connection control comes after the completion of all preparations of the call control stage. In the connection control stage, a real connection links the parties or entities. In brief, the functions of the connection control are:

- Establishing, modifying, executing, and releasing connections between parties or entities;
- Supervising the quality-of-service parameters;
- Providing flow control, congestion control, routing, and reservation, allocation, negotiation, and renegotiation of resources;
- Notifying the network of party or entity changes; and
- Managing ownership and control rights.

From the point of view of the switching architecture, it also may be necessary to guarantee end-to-end QOS parameters within and across networks for a given call, and connection control will be required to translate these requirements in real-time to establish and manage connections.

Routing. The switching system should be fast enough to select the routing strategy that maximizes the long-term network throughput and provide the lowest possible end-to-end delay. Connection management will select paths between the source and destination points that will meet these objectives to optimize the use of network resources.

Many routing algorithms that provide optimal dynamic routing, such as load balancing, shortest path, and load balancing with modified shortest path, are discussed in the literature.^{7,8} These studies primarily consider, however, a single medium for a given call. Routing algorithms now must be developed to consider multimedia traffic containing audio, video, and data, and the criteria for selecting a path should include the traffic parame-

ters of each medium to calculate the load for each connection. Therefore, switching systems must employ more complex routing algorithms than they use today to optimize network capacity.

Congestion Control. Congestion control is a primary concern of cell-switching networks, where the multimedia traffic will be switched in the form of cells. In general, the cell-switching network will provide neither link-by-link error control through retransmissions, nor link-by-link flow control. Instead, it will use end-to-end error control, internal virtual circuits, and hardware switching to:

- Avoid large delays due to cell processing and queuing time within the switching nodes, and
- Minimize the variable delays between the cells.⁹

From a switching point of view, a multimedia connection can be considered as a connection that carries several media. Each medium has its own characteristics and QOS requirements, and should be treated separately. Audio and video traffic, for example, cannot be retransmitted to recover from errors, but data traffic must be recovered from errors through retransmissions.

In high-speed switching networks—where propagation delays across the network typically dominate switching, and buffering delays and the traditional feedback from the network to control congestion are outdated—any action the source that is generating the traffic takes is too late to resolve buffering or switching congestion. This situation becomes worse if the network has to handle real-time audio or video traffic. Therefore, high-speed networks will have to provide preventive congestion control schemes to handle bandwidth-intensive multimedia applications.

Quality-of-Service Parameter Control. In cell-switching networks, the primary QOS parameters include cell delays, cell-delay variations, and cell-loss probability. In addition, BER, differential delay between audio and video, and end-to-end delay also are important performance parameters.

A call will be accepted based on the QOS parameters negotiated at the time of call setup. The network, before accepting a call, will examine the current load and decide whether the switching system can maintain the required performance levels. Congestion can occur when the system accepts too many calls, which will deteriorate the QOS. Thus, congestion control schemes that are used by the switching systems are very critical to

maintaining the promised QOS and preventing congestion.

Media Bridging. Media bridging provides the capability to communicate from any given point to any other end points, in the same or different networks, for audio, video, or data. In other words, bridging provides point-to-multipoint or multicast communications.

Multicasting is an important requirement for many multimedia applications, especially for MMC. In a circuit-switched environment, a point-to-multipoint dedicated connection is established for a call. Because these circuits are dedicated, not shared, for the duration of the call, they are costly propositions, especially for variable bit rate traffic, because of the low utilization of trunk bandwidth. In a cell or packet switching mode, however, multiple virtual connections can be set up for multicast connections. The bandwidth of each circuit path can be shared logically among different calls. This provides economies of scale by taking advantage of the bursty characteristics of multimedia communications.

Shared medium LANs, such as AT&T's StarLAN network or Ethernet,* have the unique property of providing multipoint connectivity by broadcasting in a limited local environment. But the same cannot be applied in a wide area network environment, because of the increased probability of packet collisions. Recently, however, a multicast protocol has been developed for multimedia traffic across wide area networks.¹⁰

Many algorithms also are available to implement broadcasting or multicasting to a group of connected entities. A limited implementation could be achieved by placing a multicast server in each switching system, but a much more efficient approach is to support multicast operation directly within the switch.¹¹

The idea of supporting multimedia bridging within the switch is an active research subject, made all the more complex by the fact that the requirements for bridging each medium may differ to some extent. For example, audio bridging requires the summation of all (or the loudest) audio incoming signals, and it then delivers only the sum to each endpoint. This summation has to be performed before the multicasting of the audio signals. In addition, if the endpoints use different coding standards, the conversion between the different types of speech coding techniques used may also be required.

In simple video bridging, a single video signal is sent to multiple destinations. In the case of multimedia collaboration, the video signal that is to be multicast may

have to be transmitted based on some agreed upon principles. For example, a voice control scheme may be based on audio intensity and, therefore, multicasts of the video will be associated with the highest audio intensity.

If a pre-designated chairperson controls the selection of the video source, the video bridge will multicast that video source, based on the chairperson's selection. In some instances, each viewer also might choose an individual video for multicasting. Therefore, video multicasting, which may change several times during the course of the conference, has to be coordinated with audio bridging. This inter-relationship between audio and video is the fundamental cornerstone for multimedia applications, especially for MMC.

Data bridging is required when more than one entity is allowed to send data. The ordering of data can be performed in accordance with some predefined procedures. For example, a simple multicast function can be performed when a single source has to transmit data to multiple destinations. In more complex interactions, however, a data bridge might receive data units from different sources. It then must send all data units to all destinations as they are received, without performing any ordering of the data units.

In another situation, the data bridge may be required to multicast data units received from multiple sources in a strict order. In all, there may be many more variations of these data bridging functions, depending upon user needs.

The performance objectives of media bridging is very stringent for multimedia applications. For example, lip-synchronization of audio and video signals and the fast processing of all required bridging functions, to name but two applications, present tremendous challenges to the switching systems. Needless to say, media bridging also is an active topic for research for the engineering community.

Switching Technique Alternatives

The switching systems that carry multimedia applications should have certain basic properties to satisfy the performance requirements of real-time audio and video, as well as data traffic.

Cell-switching networks should provide fixed, low end-to-end delay with little or no variation, and little or no variation in cell delay. In this respect, asynchronous transfer mode (ATM), synchronous transfer mode

(STM, also called circuit switching), frame relay (FR), and Switched Multimegabit Data Service* 802.6 (SMDS) switching techniques are four candidates for comparison. This section compares the properties of these different switching techniques.

The traditional packet-switching technologies using software switching and variable packet size have not been considered here, because these systems cannot guarantee the performance of real-time audio and video traffic. Most of these traditional switching technologies use link-by-link error and flow control, and some of them use connectionless routing. Consequently, these systems have large end-to-end delays, with significant variability and considerable packet delay jitter. Moreover, if these large traditional systems happen to transfer multimedia traffic, they can develop serious bottlenecks, even though they use sophisticated adaptive routing and congestion control algorithms.

Table II shows a high-level comparison between ATM, STM, SMDS, and FR. SMDS uses fixed cell size and connectionless routing, while FR uses variable packet size and connection-oriented routing. ATM uses fixed cell size and connection-oriented routing. STM circuit-switching uses connection-oriented routing and transfers information bit streams to maintain that order. Connectionless routing does not incur any call setup delay, since cells are transferred without prior establishment of the routing path.

Connectionless routing schemes, as noted, do not guarantee to maintain the order of the cells, and resequencing of the cells is usually required at the destination point. In addition, more processing is needed for connectionless traffic, since routing has to be performed on a per-cell or per-packet basis.

Both connectionless routing and variable packet size will cause relatively large variable cell and packet delays, and may not guarantee the delay performance parameters required for real-time audio and video traffic.

Performance: STM is the best technology to guarantee performance for both real-time audio, video, and data, although ATM also can guarantee that performance parameters are met for all multimedia traffic. Both SMDS and FR will be able to guarantee the performance for data, but it is not clear whether these switching technologies will guarantee delay performance for real-time audio and video traffic.

Bandwidth: STM and ATM will provide multi-

Table II. Comparison of switching technologies multimedia traffic

Parameters		Technology alternatives			
		ATM	STM (Circuit switching)	SMDS (802.6)	Frame Relay
Support of multimedia traffic transfer	Data	Yes	Yes	Yes	Yes
	Audio	CBR	Yes	Yes	
		VBR	Yes		
	Video	CBR	Yes	Yes	
VBR		Yes			
Connectivity	One-to-many	Yes	Limited	Yes	Yes
	Many-to-one	Yes	Limited	Yes	Yes
	Many-to-many	Yes		Yes	Yes
Performance guarantee	Delay for audio/video	Yes	Yes		
	Packet loss for data	Yes	Not applicable	Yes	Yes
Bandwidth		1.5 Mbits/s to multi Gbits/s	Few bits/s to multi Gbits/s	1.5 Mbits/s to 45 Mbits/s	56 kbits/s to 1.5 (claim 45) Mbits/s
Guarantee of media synchronization		Yes	Yes		
Round trip delay		Low	Lower than ATM	Higher than ATM	Higher than ATM
Channel/trunk sharing capabilities		Yes	No	Yes	Yes
Channel/trunk utilization		High	Low	High	High
Congestion control		Required	Not required	Required	Required

ATM - Asynchronous transfer mode
 CBR - Constant bit rate
 SMDS - Switched Multi-megabit Data Service
 STM - Synchronous transfer mode
 VBR - Variable bit rate

Gigabits per seconds bit rates; SMDS and FR are currently defined at 45 and 1.5 Mbits/s, respectively, although it is claimed that FR can also transfer efficiently up to 45 Mbits/s.

Synchronization: STM and ATM technology will guarantee both intra- and inter-media synchronization of audio and video traffic, because of negligible or no delay jitter, low fixed end-to-end delay, and little or no variation in cell delay.

Connectivity: ATM, SMDS, and FR can provide flexible connectivity for one-to-many, many-to-one, and many-to-many communications environments. STM can provide one-to-many and many-to-one connectivity in a limited way, but it does not offer efficient many-to-many communications, as compared to cell- and packet-switching technology.

Support of Multimedia Traffic: Both ATM and STM can carry audio, video, and data traffic, but STM is not intended to transfer variable-bit rate (VBR) audio and video traffic efficiently. SMDS and FR have been standardized to carry data only.

Channel Sharing and Trunk Utilization: ATM, FR, and SMDS can both exploit the bursty characteristics of multimedia traffic and share channels through statistical multiplexing. As noted, however, STM cannot share a channel using statistical multiplexing. Consequently, trunk bandwidth utilization remains very low for STM, while ATM, SMDS, or FR technology supports very high utilization of the trunk bandwidth.

Congestion Control: One advantage of STM is that, with its dedicated channel for a call, it does not need to have a congestion control scheme after call setup, whereas ATM, SMDS, and FR require elaborate congestion control to guarantee the required QOS parameters negotiated at the time of the call setup. The success of these cell and packet-switching technologies to carry multimedia traffic will depend on the effectiveness of their congestion control schemes.

Summary of Comparisons. By this comparison, it is clear that ATM is the only

switching technology of choice that has the promise to carry multimedia signals efficiently by sharing channels through statistical multiplexing. In addition, ATM can guarantee the performance for multimedia traffic, including the synchronization for each medium. STM technology also will guarantee superior performance for multimedia traffic, but it lacks the capability of efficient trunk utilization. Moreover, there is no cost-effective scheme that can compete with ATM, on a multiparty variable-bit rate call, to transfer multimedia signals. Furthermore, ATM also can support circuit connections through synchronous-to-ATM conversion (SAC).

Therefore, ATM will support both circuit- and cell/packet-switching traffic and has the ability to be the core switching technology.

Conclusions

The key parameters of multimedia applications that most influence the criteria for switching architecture are:

- Media, such as audio, video, and data;
- Bandwidth, from a few Kbits/s to multi-Gbits/s;
- Traffic characteristics, including burstiness, burst length, and variance of bandwidth;
- Performance, including end-to-end delay, delay jitter, differential delay (primarily between audio and video), response time, cell and packet switching time and retransmission delay (for data traffic), bit-error-rate and cell loss rate, and intra- and intermedia synchronization; and
- Functional requirements, including call control and connection control, connectivity, routing, congestion control, QOS parameters control, and media bridging.

The bandwidth requirements, traffic characteristics, and performance criteria are the fundamental parameters that should determine the basic switching technology. Multimedia applications have diverse, but stringent, requirements for audio, video, and data. The complex functional requirements of multimedia applications impose additional burdens on the switching system.

A given switching architecture is required to meet the stringent requirements in a cost-effective way to provide an optimized solution for multimedia networking. The comparison between switching techniques reveals that both STM and ATM can meet both performance and bandwidth requirements for the emerging multimedia applications. ATM needs to employ sophisti-

cated congestion control to guarantee the required QOS parameters. But STM appears to be incapable of providing an efficient solution in many-to-many communications. This technology also does not have channel sharing capability, which provides for the high utilization of the channel bandwidth by taking advantage of the burstiness of multimedia traffic.

ATM can provide statistical multiplexing for bursty multimedia traffic, thus yielding a high utilization of the trunk bandwidth. ATM also can support circuit connections. Therefore, ATM can become the core switching technique to support both circuit, cell, and packet switching applications.

The efficient support of additional functionalities, such as media bridging within the switching architecture, is a significant challenge because of the complexities involved. Routing algorithms used for switching also must be optimized to meet the diverse requirements of multiple media. The separation of call control and connection control is a fundamental requirement, in a distributed networked multimedia environment, to:

- Optimize the use of the network resources, and
- Maintain flexibility for the complex interaction between different networking entities.

The ATM switching system¹² appears to be flexible enough to support these functions, and to optimize the execution of multimedia communications.

* Ethernet is a registered trademark of Xerox Corporation; Switched Multi-megabit Data Service is a registered trademark of Bellcore

References

1. J. McQuillan, "From Packet to Cell: Better Performance," *Business Communications Review* (BCR), May 1994.
2. S. Ramanathan and P. Venkat Rangan, "Architecture for Personalized Media," *IEEE Multimedia*, Vol. 1, No. 1, Spring 1994.
3. R. Roy, "Networking Constraints in Multimedia Conferencing and the Role of ATM Networks," *AT&T Technical Journal*, Vol. 73, No. 4, July-August, 1994.
4. L. Dittmann and S. B. Jacobsen, "Statistical Multiplexing of Independent Bursty Sources in ATM Network," *Proceedings, IEEE GLOBECOM 1988*.
5. A. E. Eckberg, Jr., D. T. Luan, and D. M. Lucantoni, "Meeting the Challenge: Congestion and Flow Control Strategies for Broadband Information Transport," *Proceedings, IEEE GLOBECOM, 1989*.
6. M. Sidi, W. Z. Liu, I. Cidon, and I. Gopal, "Congestion Control Through Input Rate Regulation," *Proceedings, IEEE GLOBECOM, 1989*.
7. M. Schwartz, "Telecommunication Networks: Protocols, Modeling and Analysis," Addison-Wesley Publishing Company, 1987.

-
8. I. Cidon, I. Gopal and R. Guerin, "Bandwidth Management and Congestion Control in plaNET," *IEEE Communication Magazine*, Vol. 29, No. 10, October, 1991.
 9. W. Stallings, "ISDN and Broadband ISDN," McMillan Publishing Company, 1992.
 10. S. Casner, "Are You on the MBone?," *IEEE Multimedia*, Vol. 1, No. 2, Summer 1994.
 11. P. Newman, "ATM Technology for Corporate Networks," *IEEE Communications Magazine*, Vol. 30, No. 4, April 1992.
 12. T. M. Chen and S. S. Liu, "Management and Control Functions in ATM Switching Systems," *IEEE Network, the Magazine of Computer Communications*, Vol. 8, No. 4, July/August 1994.

(Manuscript approved August 1994)

Radhika R. Roy is a member of technical staff in the Interactive Information Services Architecture Department of AT&T Bell Laboratories in Holmdel, New Jersey. He works on the development of multimedia services, multimedia network design, and technology assessment. He has a B.S.E.E.

degree from the University of Engineering and Technology in Dhaka, Bangladesh, an M.S.E.E. degree from Northeastern University, Boston, Massachusetts, and a Ph.D. in electrical engineering from the City University of New York. He joined AT&T in 1990.



Ashok K. Kuthyar is head of the Interactive Information Services Architecture Department in AT&T Bell Laboratories in Holmdel, New Jersey. His group is responsible for developing new multimedia and image product and service concepts, providing a technology transfer from research to the business units, defining end-to-end architectures and requirements, and developing prototypes. He has a B.S.E.E degree from Bangalore University, in Bangalore, India, and an M.S.E.E. degree from Kansas State University in Manhattan, Kansas. He joined the company in 1986.



Vanita Katkar is a technical manager in the Next Generation Network Architecture Department at AT&T Bell Laboratories in Holmdel, New Jersey. Her group is responsible for developing the multimedia services creation architecture. She joined AT&T in 1980. She has a B.S. degree and an M.S. degree in mathematics from the University of Poona in Poona, India.

