A Standards-Based Multimedia Conferencing Bridge

David N. Horn
Terry G. Lyons
John C. Mitchell
Dale L. Skran

Standardizing video, audio, and data exchange formats under the auspices of the International Telegraph and Telephone Consultative Committee (CCITT) will make possible a large new market in standard videoconferencing devices. This paper describes a possible solution for multipoint, multimedia conferencing in such an environment.

Introduction

Current trends in information systems technology include integrating computing and communications, multimedia multiparty conferencing, distance learning, computer supported collaborative work (CSCW), hypermedia systems, and other distributed multimedia services.

The impact of these advances may revolutionize society. Multimedia conferencing is not only a way to reduce travel so people can work together in a global marketplace; it also is a method to help people work together more effectively and more productively, regardless of their locations. By bringing computer tools and on-line information access to the meeting room, the decision-making process can be improved, and intellectual property such as designs or documents can be created during the meeting.

The volume of information bombarding us is overwhelming. Hypermedia systems (i.e., multidimensional, multimedia documents), may need to interactively access information located at many places over a network. Computer-based information processing tools will soon become essential to assimilate this information.

Video compression and the Integrated Services Digital Network (ISDN) are "enabling" technologies that help make this revolution possible. Another required enabling technology is a set of standards, protocols, and a signaling system to access these revolutionary new services.

These new services present two major communications requirements.

Multimedia Requirements. The network must transport multimedia rather than today's primarily audio-oriented unimedia. Multimedia includes audio, video, and other data, i.e., all digital equipment *excluding* audio and video, even though the audio and video may be digitized. To transport multimedia information, multiple network connections may need to be associated with each other and also synchronized. An ISDN Basic Rate Interface (BRI) provides two 64 kilobytes per second (kb/s) B channels plus a 16 kb/s D (signaling) channel, to offer 128 kb/s of usable bandwidth. Because multichannel call routing is not yet generally available, a user may have to request two separate B channel calls. These may be routed over the network through totally different routes and hence incur different delays.

Multiparty Operation Requirements. Group work implies multiparty operation. Thus, two new requirements must be considered:

- Signaling for managing multimedia, multiparty calls. The signaling aspect is complex. There is a need for flexible conference dynamics, e.g., adding or dropping parties, and changing media bandwidth allocation. Also, each user may have more than one line to the network, and each site may have more than one user. Access privileges (e.g., password control for entering a conference) and billing also must be considered.
- There must be provisions for media-specific bridging, i.e., audio bridging, video bridging, and data bridging. These media-specific bridging requirements are covered in detail in the next section.

Implementation Problems. The problems with implementing these multimedia, multiparty services on today's ISDN network include:

- Signaling on the D channel is only between the endpoint and the nearest network termination (i.e., the private branch exchange [PBX] or the local central office). User-to-user signaling is limited at best, and is not universally supported through existing networks.
- All basic unrestricted digital network services are point-to-point.

Multimedia Conference Bridge System Architecture

This section introduces a solution based on a multimedia conference bridge, or, using CCITT terminology, a multipoint control unit (MCU). The MCU is connected to the network via one or more digital trunk interfaces. It can be connected to a central office switch, a PBX, or any other convenient point that is accessible via dial-up point-to-point services from the user terminals. Figure 1 illustrates the configuration.

To conduct a multiparty, multimedia conference, point-to-point clear-channel circuit connections based on the conventional (Q.931) ISDN signaling procedures are established between the user terminals and the MCU. The MCU performs several basic functions.

Conference Management. Conference management involves making and receiving calls, negotiating and dynamically managing the bandwidth allocated to each media, adding and dropping conferees, and managing the media-specific bridges.

Audio Bridging. In addition to the basic audio bridging function of audio summation and gain control, a multimedia audio bridge should perform the following extra functions:
- Audio energy level monitoring, for optional video control and talker indication.
- Transcoding between different types of speech codings. Transcoding allows conferees with different quality speech encoding methods to talk together, without dropping to the lowest common denominator in terms of quality.
- Muting idle channels to reduce background noise.
- Delaying the audio to match any video delay introduced by video processing at the video bridge.

Video Bridging. Video compression hardware takes raw digitized video at approximately 45 megabits per second (Mb/s) and compresses it to a lower data rate (56 kb/s to 1.536 Mb/s) for transmission over the network. The hardware is still expensive, though very-large-scale integration (VLSI) technology is expected to bring the cost down rapidly. The compression and decompression causes a delay of several hundred milliseconds, and the picture quality is generally inferior to TV quality, especially for rapidly moving objects or images—such as normal printed text—with highly detailed content. Quality depends greatly on bandwidth: 128 kb/s video is much better than 56 kb/s video.

Given a conference where each party has only one source of compressed video and one display, one video bridging technique allows all users to view a particular user's video source. This means that the contents of the single compressed video source must be multicast (i.e., copied) to multiple destinations. The switching could be controlled either by a chairperson, who decides who sees what, or automatically by audio energy level, i.e. everyone sees the current loudest talker, except himself or herself, who sees the previous talker. Each viewer could also be allowed to make his or her choice of video source. Multi-window video, where the viewer can simultaneously can see multiple video sources in windows, is possible, but is currently expensive, adds delay, and causes loss of quality.

Data Bridging. The type of data bridging required depends on the application. Some applications may only

Panel 1. Terms and Acronyms in This Paper

AGC	AudioGraphic Conference (application)
AVPS	audio visual protocol stack
BAS	bit rate allocation signal
BRI	Basic Rate Interface
CCITT	International Telegraph and Telephone Consultative Committee
CELP	code-excited linear prediction
CIF	common intermediate format
CSCW	computer supported collaborative work
GCC	generic conference control
HSD	high speed data
ISDN	Integrated Services Digital Network
JPEG	Joint Photographic Experts Group
LSD	low speed data
MCS	multipoint communications service
MCU	multipoint control unit
MLP	multilayer protocol
PBX	private branch exchange
PCM	pulse code modulation
VLSI	very-large-scale integration

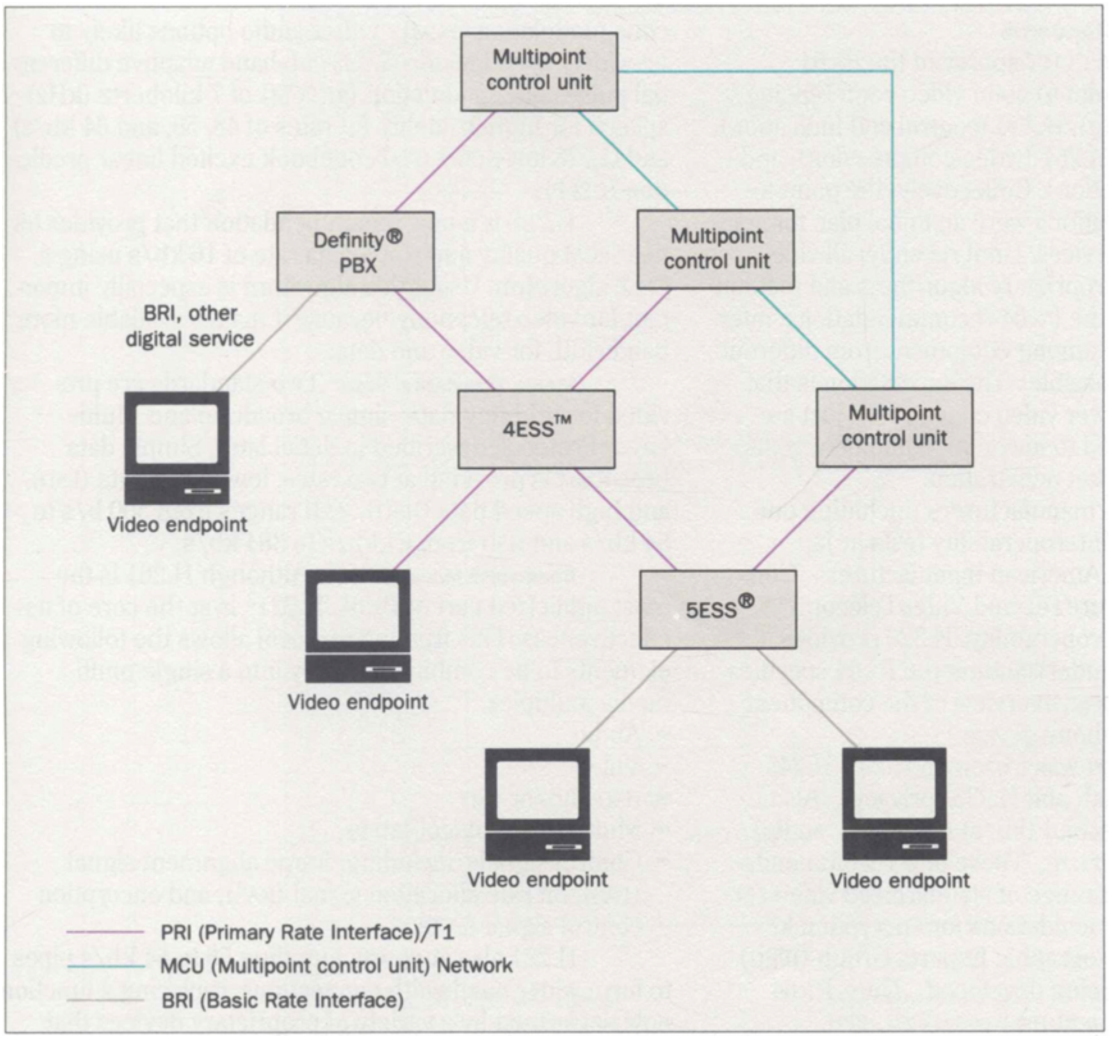

Figure 1. A diagram of a Multipoint Control Unit (MCU) layout, showing connections to the network through digital trunk interfaces.

require each party to receive data from a single source. Here, as with video bridging discussed earlier, the MCU must be able to multicast streams of data from one source to multiple destinations. Neither flow control nor contention resolution is needed. To receive data from a different source, the data bridge has to be reconfigured. More demanding applications, such as CSCW, may require any party to be able to send data to any set of destinations at any time. To satisfy this need, a true multipoint packet data protocol is presently being defined by the CCITT, and will be described more fully later in this paper.

Returning to the overall system architecture (see Figure 1), the MCU can be deployed anywhere in the network, including the customer's premises or the premises of a third party service provider. Multiple MCUs can be linked together in one conference, and one MCU can simultaneously support multiple conferences. All interactions with the network use only existing point-to-point services. Conference and application-related signaling between users and the MCU, and between users, is conducted "in-band" within the B channel circuits between the users and the MCU, and is transparent to the network. This implies a layered signaling scheme where protocols such as Q.931 are used for signaling between the network and the user/MCU for connections. In-band signaling is used for multimedia, multipoint call control and higher level service control.

Emerging Multi-Media Standards

Late in 1990, the CCITT approved the P×64 recommendations for point-to-point video conferencing, including H.221 (framing), H.230 (control and indication), H.242 (channel setup), H.261 (video compression), and H.320 (endpoint description). Collectively, the point-to-point P×64 recommendations were an initial plan for standardized videophone services. Until recently, all video codecs were based on proprietary algorithms and technology. With the arrival of the P×64 recommendations, interoperability between and among equipment from different manufacturers is now possible. The expectation is that this event will lead to lower video codec costs, just as Group 3 fax standards led to lower fax equipment costs and an explosion in market penetration.

In May 1991, 14 manufacturers (including one from the U. S.) passed interoperability tests in Japan. By March 1992, several American manufacturers (Compression Labs, Inc., PictureTel, and VideoTelecom) demonstrated P×64 interoperability. H.320 provides a good starting point for understanding the P×64 specifications because it contains an overview of the component elements of a videotelephony device.[1]

In May 1992, text was frozen on H.231, H.243 (multi-point control units), and H.233 (privacy). Also, G.728 [16 kilobits per second (kb/s) toll-quality audio] was accepted in its final form. These new recommendations provide for a wider range of standardized video services. Additional recommendations for encryption key exchange and Joint Photographic Experts Group (JPEG) still image transfer are being developed. Thus, P×64 continues to gather momentum.

Media Standards: Video. The core of the P×64 series of recommendations is H.261,[2] a specification for using a hybrid of discrete cosine transform (DCT) and differential pulse code modulation (PCM) to compress real-time video down to rates varying from 56 to 64 kb/s to 1.92 Mb/s. H.261 covers both P×56/64 kb/s (p = 1, 2, ... 6) and M×384 kb/s (m = 1, 2, .. 5). Two video quality modes are provided for: CIF (common intermediate format) and QCIF (quarter-CIF). CIF provides for a 288 lines by 352 pixel picture, and QCIF for a 144 line by 176 pixel picture.

Media Standards: Voice. Several audio standards are encompassed in the P×64 recommendations. However, all endpoints are required to support G.711 pulse code modulation (PCM). Other audio options likely to be widely used include G.722 sub-band adaptive differential pulse code modulation (ADPCM) of 7 kilohertz (kHz) speech for higher fidelity (at rates of 48, 56, and 64 kb/s) and G.728 low-delay (LD) codebook excited linear prediction (CELP).

G.728 is a new recommendation that provides for near-PCM quality audio at a data rate of 16 kb/s using a CELP algorithm. Using this algorithm is especially important for video telephony because it makes available more bandwidth for video and data.

Media Standards: Data. Two standards are provided for bridging data, simple broadcast and Multi-Layer Protocol, described in detail later. Simple data broadcast is provided at two rates, low speed data (LSD) and high speed data (HSD). LSD ranges from 300 b/s to 64 kb/s and HSD from 64 kb/s to 384 kb/s.

Bandwidth Management. Although H.261 is the most publicized part of P×64, H.221[3] is at the core of its effectiveness. This framing protocol allows the following elements to be combined flexibly into a single multimedia multiplex:

- Audio
- Video
- HSD and/or LSD
- Multilayer protocol (MLP)
- Control signals, including: frame alignment signal (FAS), bit rate allocation signal (BAS), and encryption control signal (ECS).

H.221 also supports bundling 56 to 64 kb/s pipes to form wider bandwidth connections, replacing a function now performed by a variety of proprietary devices that do not interoperate. Thus, H.221 is an internationally accepted open architecture for "inverse" multiplexing.

For example, in a 128 kb/s call we might observe 7 khz audio encoded at 56 kb/s, with the remaining bandwidth allocated to video and control. Dynamically, this might change to 16 kb/s CELP audio and 14.4 kb/s LSD with the remaining bandwidth allocated to video. HSD at 64 kb/s could be provided with the video turned off. Figure 2 shows a sample allocation of 384 kb/s of bandwidth on an H0 channel.

Point-to-Point Conference Control. Point-to-point conference control is covered by H.230[4] and H.242.[5] There are two kinds of BAS codes, commands and capabilities. Each capability, when sent, informs the far end

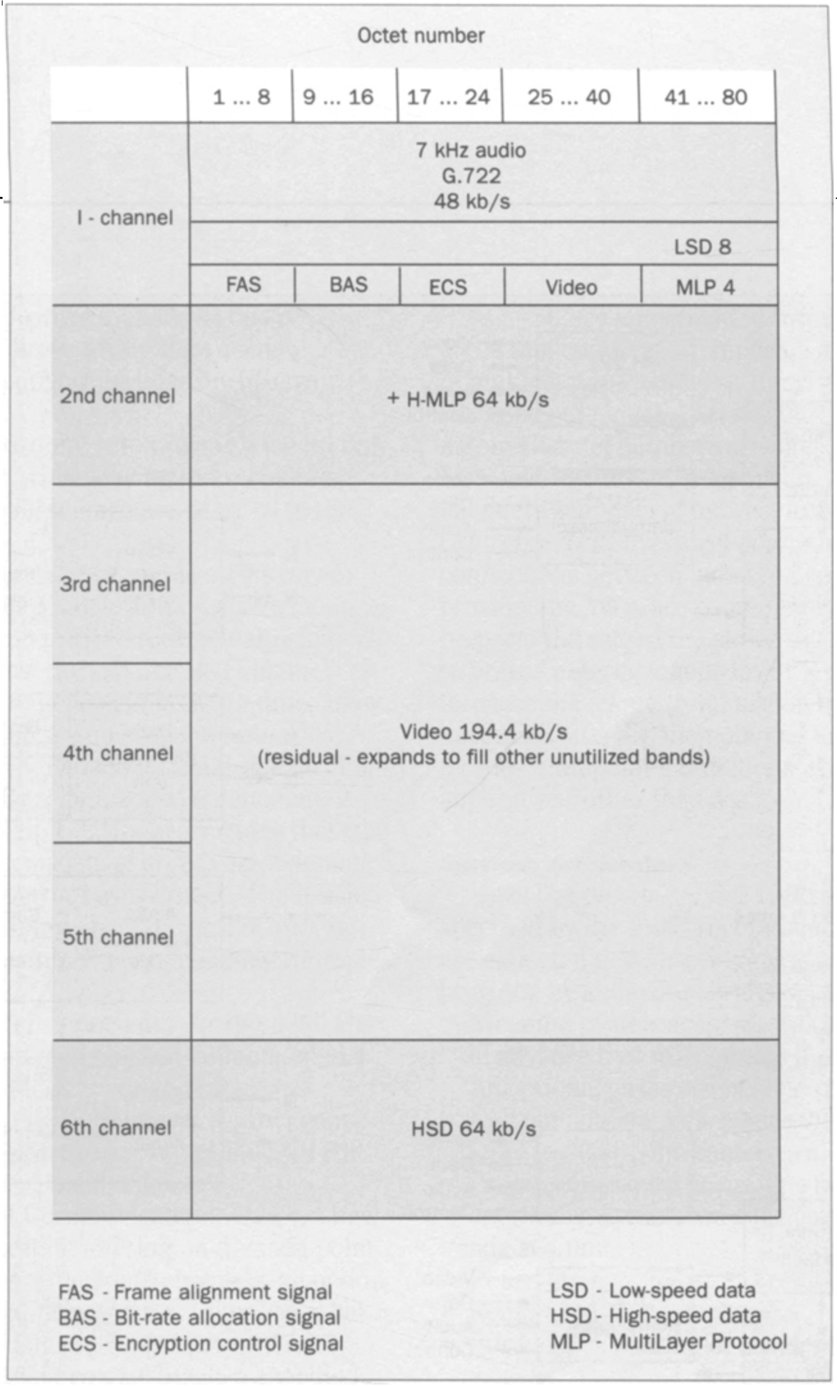

Figure 2. A specimen 384 kb/s bandwidth allocation.

that the sender can accept a given set of commands. For example, the audio capability "A-law" shows support for G.711 A-law PCM. Such a terminal could be expected to respond to the commands *A-law*, *OU*, and *A-law OF* (unframed and framed A-law PCM).

A session begins with a capability exchange. Each endpoint sends the other a set of capabilities that it wishes to disclose. The session is then presumably held at the highest common rates. The capability exchange may occur at any time during the call, and may be used to add or remove capabilities dynamically.

Multi-Point Conference Control. The MCU recommendation for P×64 has been frozen since May 1992. Final approval by the CCITT is expected early in 1993. Its major components are:

- H.231, multi-point control units for audiovisual systems using digital channels up to 2 Mb/s.
- H.243, basic MCU procedures for establishing communications between three or more audiovisual terminals using digital channels up to 2 Mb/s.

H.231 describes the overall architecture of an MCU, illustrated in Figure 3. H.243 addresses differences in

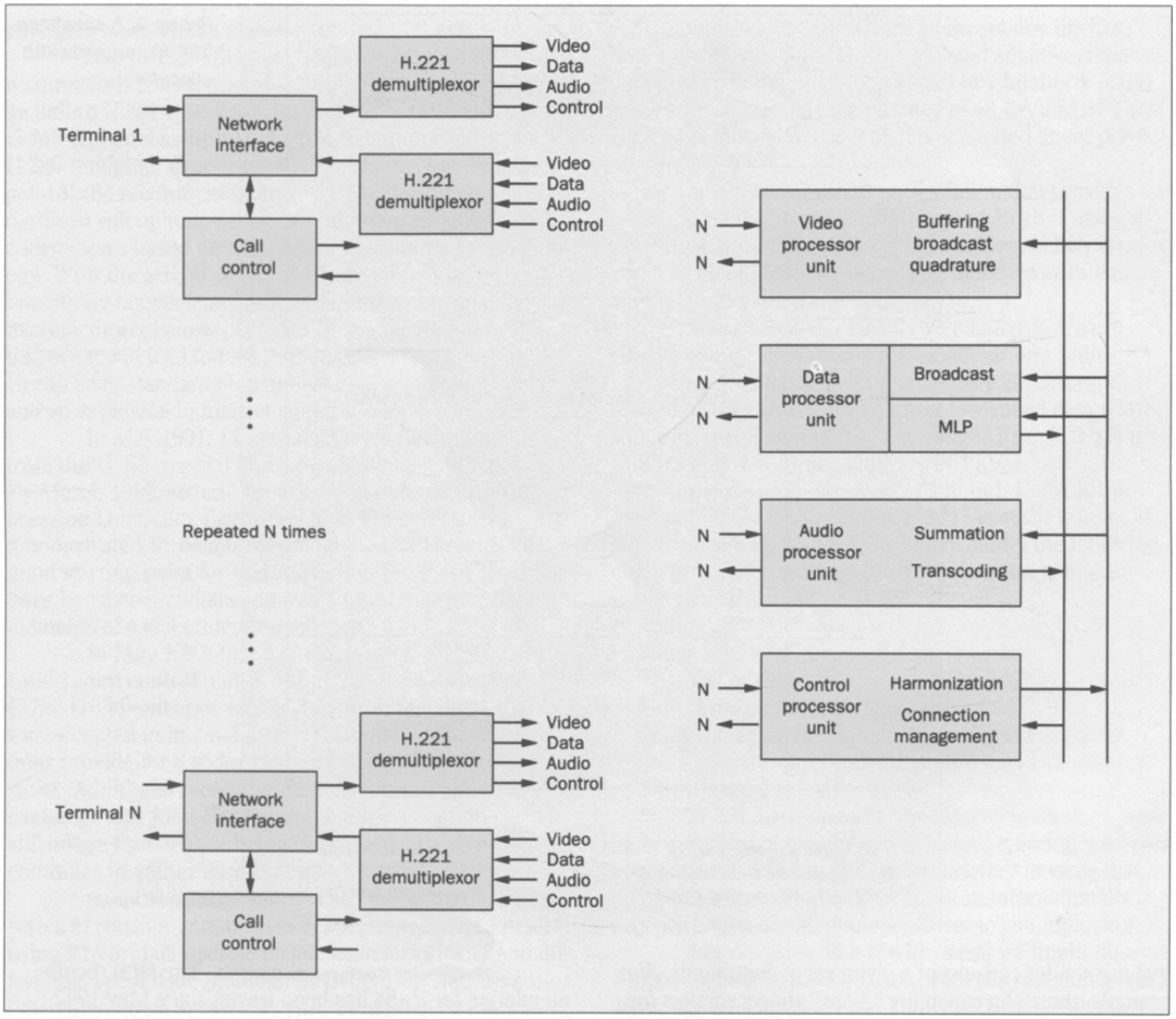

Figure 3. The generic architecture of an MCU following the H.231 standard, using H.221 multiplexers and demultiplexers.

operation between multi-point and point-to-point. It also describes management features that are significant only in a multi-point situation. Areas covered include:

- Terminal numbering
- Chair control of video broadcast
- In-band identification of endpoints for security purposes
- Multi-point control of LSD and HSD data broadcast, with the MCU controlling a token for each type of data

- Cascading of MCUs, a feature that allows customers to conference a larger group than a single MCU would normally support, or possibly to save on toll charges.

The overall philosophy is to provide a useful but simple set of commands, while leaving more complex features such as multi-level hierarchies of MCUs to the control of MLP.

Multilayer Protocol. The G- and H-series recommendations of Study Group XV described above focus on audio-video encodings and provide rudimentary support for data. They specify how the LSD and HSD channels of H.221 can be allocated to one transmitter at a time, allowing it to broadcast an arbitrary bit stream to other parties. Although H.221 permits opening another data channel labeled MLP (including a high-speed component, H-MLP), its use is not detailed. H.231 merely states that MLP data handling in an MCU consists of processing telematic information and conference control signals. The implied requirement, a distinguishing characteristic of MLP, is that it support interactive data exchange among multiple transmitters.

Study Group VIII has been developing a full MLP specification. The current study period culminated in a framework of five projected recommendations:

- T.121 AGC. AudioGraphic Conference, a core application that gives terminals a shared visual space of still images, annotations, and manipulations.
- T.122 MCS. Multipoint Communication Service, a new type of data communication offering on-demand point-to-point and multipoint virtual data channels, an option of uniformly sequenced delivery, and tokens to arbitrate resource contention.
- T.123 AVPS. Audio Visual Protocol Stack, a specified set of lower layer protocols to support MCS, for ISDN and for other important networks.
- T.124 GCC. Generic Conference Control, an application combining terminal-to-terminal, terminal-to-MCU, and MCU-to-MCU commands and indications.
- T.125 MCSP. Multipoint Communication Service Protocol, detailing how control information is exchanged over AVPS to implement the MCS service definition contained in T.122.

Figure 4 represents the relationship of these components. Agreement has been reached so far on the contents of T.122 and T.123. These are being sent to the CCITT plenary for approval in the first half of 1993.

MCS is implemented through a hierarchy of MCUs that cooperate to replicate and distribute the data of multiple transmitting terminals. Participants are assigned individual, addressable user identifiers. Destination channel numbers are indexes into distribution lists held by the MCUs. Receivers are allowed to join channels with data of interest to them (either known beforehand or advertised over other channels). Reliable connections between individual terminals and MCUs, as provided by AVPS, are combined into an overall flow that respects the rate of the slowest receiver. MCS is designed to bridge heterogeneous lower layers, allowing data conferences to include terminals on both ISDN and local area networks. MCS has the potential to become an open interface for multipoint interactions. It can support groupware applications other than AGC.

Services Architecture

The type of services offered by an MCU are not specified by the P×64 recommendations. MCU's services are expected to offer bridging features similar to those provided by audio-only bridges, including:

- Meet-me conferences, where conferees get a number in advance that will connect them to the conference. An optional password may be required to join it.
- Dial-out conferences, where the MCU calls each conferee to initiate the conference.
- Progressive conferences, in which a conferee calls successive parties and adds them to the conference one at a time.
- Reservation systems.
- Remote maintenance.
- Attendant screening of conferees for added security.

Limitations and Future Directions

The current standards have some limitations:

- Deployed ISDN networks do not fully support the multi-use information transfer capability for audiovisual services (previously designated 7 kHz audio) with interworking to ordinary telephone services.
- The bandwidth allocated to different media under H.221 is constant for long periods of time. It cannot expand to accommodate instantaneous demands for video or data during peaks in use, nor contract during intervals when a signal is idle.
- H.221 packages a single set of media streams into a P×64 digital connection. Transporting multiple video

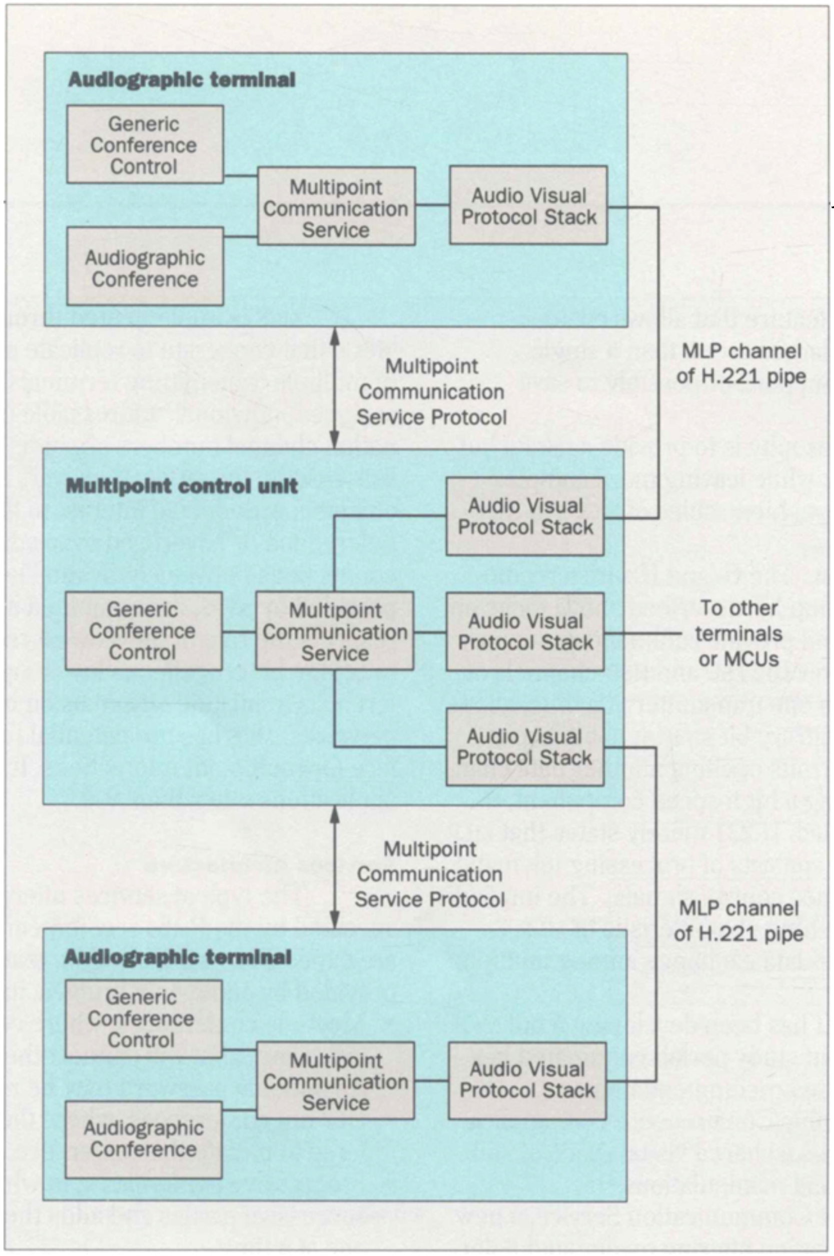

Figure 4. A diagram of how the multilayer protocol (MLP) data channel might function in a multimedia conference.

streams within a conference requires that multiple P×64 connections be set up. This affects the ability of cascaded MCUs to offer viewers a choice of video sources originating outside their respective parts of a conference.
- Encryption key management and authentication are not provided for in H.233 (privacy). A separate recommendation for these areas is currently being prepared.
- Standards for transfer of JPEG images in the context of video teleconferences have not yet been completed, but are currently being developed.

Future developments may redress these limitations and expand services in other ways:
- Audiovisual calls can be endowed with features that make ordinary telephony more convenient to use, including call forwarding, additional call offering, hold, transfer, and conference.
- Packet switched framing, which is natural for data, can also be applied to the audio encoding. Audiographic streams may be transported in this way through a multi-service frame relay network or a local area network. Many logical streams could be accessed through a single network interface, with efficient use of bandwidth on demand.
- Broadband ISDN applies similar concepts of cell relay to higher transmission speeds. It supports the multiplexing of several virtual circuits over a virtual path between endpoints. With the completion of H.261, work on video coding in Study Group XV (where the P×64 recommendation originated) has shifted to this domain.
- The multipoint communication of MCS, currently built on a set of point-to-point connections between

terminals and MCUs, could evolve to using future switched multipoint network services. This would require work in advance on multipoint transport layer protocols.

- New service offerings can be built around repositories for storing multimedia information. These may include message recording for later retrieval, broadcast distribution to large numbers of subscribers, or interactive sequencing through audiovisual programs.

Summary

We have presented an open architecture for implementing a new generation of multimedia, multiparty applications and services on the current network, using existing signaling and services. The signaling model is layered, using the existing messages to signal for point-to-point connections to a MCU, then using in-band signaling for higher-layer services.

The key to success and interoperability is to define and adhere to the right set of standards. Each layer should be powerful enough to support the sophisticated applications of the future, yet general enough to support arbitrary applications including the multimedia equivalent to "plain old telephone service."

We expect the development of new and powerful multimedia applications will develop hand in hand with the deployment of MCUs.

References

1. "Narrow-Band Visual Telephone Systems and Terminal Equipment, CCITT Recommendation H.320, International Telegraph and Telephone Consultative Committee, Geneva, Switzerland, 1990.
2. "Video Codec for Audiovisual Services at P×64 kb/s," CCITT Recommendation H.261, International Telegraph and Telephone Consultative Committee, Geneva, Switzerland, 1990.
3. "Frame Structure for a 64 to 1920 kb/s Channel in Audiovisual Services," CCITT Recommendation H.221, International Telegraph and Telephone Consultative Committee, Geneva, Switzerland, 1990.
4. "Frame-Synchronous Control and Indication Signals for Audiovisual Systems," CCITT Recommendation H.230, International Telegraph and Telephone Consultative Committee, Geneva, Switzerland, 1990.
5. "System for establishing communications between audiovisual terminals using digital channels up to 2 Mbit/s," CCITT Recommendation H.242, International Telegraph and Telephone Consultative Committee, Geneva, Switzerland, 1990.

(Manuscript received June 18, 1992)