INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Université d'Ottawa · University of Ottawa

ATM RendezView:
Decentralized Architecture for Multipoint Conferencing

by

Russell Pretty, B.A.Sc.

A thesis submitted to the
School of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Applied Science
in Electrical Engineering

Ottawa-Carleton Institute for Electrical Engineering
School of Information Technology and Engineering
Faculty of Engineering
University of Ottawa

May 1997

©1997, Russell Pretty, Ottawa, Canada

National Library of Canada

Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services

Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-22010-9

Canada

This thesis is dedicated to Silke Gramsch, the love of my life, who supported me continuously and patiently for two years while we completed this work.

Acknowledgments

Keith M. Smith was my invaluable partner in the design and development of the ATM RendezView prototype system. Keith's creativity, determination, and practical knowledge of Real Time processing and software development were crucial to the realization of the prototype. We colaborated on high level design based on an architecture that I proposed as a thesis topic in 1994 but split implementation; Keith implemented the audio conferencing subsystem while I implemented the video subsystem. Keith Smith's thesis, entitled "**ATM RendezView: Enabling Real-Time Multimedia to-the-desktop**", should be read as a companion to this thesis. Keith also created the name RendezView.

I would also like to acknowledge the support and assistance of many colleagues at Nortel Technology (formerly Bell-Northern Research) who enabled me to obtain leave from my regular duties to pursue graduate studies, and provided me access to lab facilities and information resources at Nortel Technology. Notable among these colleagues are: Karen Kobierski, Marek Wernik, Peter Carbone, Irving Ebert, Jack Dyment, Jane Stewart, Patricia Dazé, Peter Ashton, Eric Livermore and Khalid Ahmad.

The University of Ottawa selected me for an NSERC Industrial Postgraduate Research Scholarship and other university scholarships without which I could not have afforded the time to pursue graduate studies.

Finally, I would like to thank my research supervisor Professor Nicolas D. Georganas for his support throughout my studies.

Abstract

This thesis describes a novel architecture for a multipoint multimedia conferencing system and describes the operational prototype of such a system dubbed "ATM RendezView". It eliminates service specific centralized Multipoint Control Units (MCU) by instead utilizing generic ATM network multipoint Switched Virtual Connection bearer services. This achieves functionality and performance improvements over MCU based conferencing systems, eases service deployment and reduces equipment costs.

Today's commercially successful and emerging standard conferencing systems are built upon an architecture that was appropriate for Time Division Multiplexed networks. Emmerging packet networks (Frame Relay, Asynchronous Transfer Mode and IP) can support higher quality, broader functionality video conferencing more simply, on less expensive and more easily deployable decentralized systems.

An operational prototype of such a videoconferencing system was developed by Keith Smith and myself. ATM RendezView utilizes Point-to-Multipoint Switched Virtual Connections to achieve a hybrid switched presence & continuous presence conference system with more user control over presentation than is available with current MCU based systems. Furthermore, RendezView enables full-motion 24-bit colour video and lip-synched high fidelity audio with end-to-end delay on the order of 100 mS. This is significantly better than existing commercial conferencing systems.

Table of Contents

List of Figures

1. Organization of this Thesis

This thesis describes a novel architecture for a multipoint multimedia conferencing system and describes the operational prototype of such a system dubbed "ATM RendezView". It eliminates service specific centralized Multipoint Control Unit (MCU) by instead utilizing generic ATM network multipoint Switched Virtual Connection bearer services. This achieves functionality and performance improvements over MCU based conferencing systems, eases service deployment and reduces equipment costs.

Section 2 (Multipoint Conferencing Today) provides background on how commercially successful video conferencing is accomplished today using ITU-T H.320 Suite conferencing standards. H.320 systems were developed for use over point-to-point Time Division Multiplexed (TDM) channels, such as those provided by ISDN and leased digital lines. Hence, H.320 utilizes centralized Multipoint Control Units to achieve multipoint operation. So called Px64 (H.320) systems also utilize Inverse Multiplexor equipment to aggregate multiple (P) switched 64 kbps subchannels into a single wideband channel. Audio and video quality on these systems is limited by bandwidth limitations on TDM networks and high long distance tariffs per 64 kbps channel set to protect revenues for traditional voice telephony. Substantial delay is introduced in terminals to achieve sufficient video compression to meet these bandwidth constraints. In addition, channel aggregation and MCUs introduce significant additional end-to-end delay that seriously impedes natural human-to-human dialogue. Finally, the requirement for video conferencing specific network equipment (MCUs and I-MUXs) adds to the cost of video conferencing , reduces the ease of service introduction and growth, limits scalability and complications administration and management of conferencing systems.

Section 2 also introduces emerging technologies and standards which will result in significant growth in the use of video conferencing over the next decade. A wide suite of H.320 variants were standardized by the ITU-T in 1996 that will enable conferencing over Plain Old Telephony Service (POTS) networks, IP-based packet networks and ATM networks. The current size of the equipment market and key market players is summarized along with market forecasts which predict significant growth in desktop conferencing due to the emergence of the new H.3xx standards. The basic architecture of the H.3xx systems still maintains a fundamental limitation of H.320, the use of point-to-point connections and hence MCUs to facilitate multipoint operation. This architectural limitation will continue to constrain market growth due to high costs, complex service introduction, limited scalability and limited performance, especially delay performance.

Section 3 (Multipoint Services on ATM) provides background of the fundamental building blocks of a new multipoint video conferencing system architecture. Alternatives of using Point-to-Point, Point-to-Multipoint or Multipoint-to-Multipoint connections to achieve multipoint conference connectivity are compared. Multipoint-to-Multipoint connections are the simplest solution in terms of control complexity but introduce problems in the user data plane due to the cell interleaving of AAL-5 frames on the Multipoint-to-Point "merge" direction of the connections. Furthermore, current signalling standards do not support establishment of Multipoint-to-Multipoint connections leaving Point-to-Multipoint connections as the preferred building block for multipoint conferencing. These multicast connections ensure minimal network delay and minimal use of network resources.

Section 4 (Decentralized Multipoint Conferencing Proposal) describes the key design goals that motivated this new examination of a conferencing system architecture and summarizes their ramifications on system architecture.

Section 5 (Implementation Reality: ATM RendezView)describes, the operational prototype videoconferencing system developed by Keith Smith and myself. ATM RendezView utilizes Point-to-Multipoint Switched Virtual Connections to achieve a hybrid switched presence & continuous presence conference system with more user control over presentation than is available with current MCU based systems. Furthermore, RendezView enables full-motion 24-bit colour video and lip-synched high fidelity audio with end-to-end delay on the order of 100 mS. This is significantly better than existing commercial conferencing systems. Section 4 provides implementation details of the video sub-system design that, when read along with source code, will be a useful reference for further development of ATM RendezView.

Section 6 (Conclusions and Recommendations) summarizes conclusions reached, some practical lessons learned and some ideas for further work to enhance such a system.

Appendix A (Initial Design Approach: H.261 Video Multiplexing for Continuous Presence) describes the authors initial, and ultimately abandoned, design approach to extending the capabilities of H.320 multipoint conferencing to enable continuous presence conferencing via low-level processing of coded video streams without decompression in MCUs.

Appendix B (H.3xx Recommendations and Protocol Stacks) provides figures of the latest H.3xx conferencing suites from the ITU-T and tables listing all the relevant protocols in the suites.

Appendix C - (Comparison of Variable Bitrate ATM Adaptation Layers) provides technical comparison of AAL3/4 and AAL-5. In the context of multipoint communications the most significant portion of this comparison are the problems with cell interleaving of AAL-5 frames on ATM merge connections.

Appendix D (Routers - ARP, Inverse ARP, Reverse ARP) provides some background on Address Resolution in router based networks. ARP has been a key factor in the standardization of multicast connections since data networking is preceding multimedia services as a driver for the use of ATM networking.

2. Multipoint Conferencing Today

This chapter provides background on how video conferencing is accomplished today using ITU-T H.320 Suite conferencing standards and introduces emerging technologies and standards which will result in significant growth in the use of video conferencing over the next decade. The current size of the equipment and service markets and key market players are summarized along with market forecasts which predict significant growth in desktop conferencing due to the emergence of the new H.3xx standards. Limitations on the video conferencing system imposed by legacy networks they were developed for are summarized.

2.1. Multimedia Networking Applications

The term multimedia was first applied to computing systems when they began to support information media beyond text and still graphics. Now media including natural images, audio and video are commonplace in consumer computing.

A multimedia network application supports the transfer of any of these information media across a network. Focusing on those that support both audio and video (AV), the following general classes of multimedia network applications are defined in the following subsections:

- Video Conferencing
- Video Telephony
- Broadcast Video and Video on Demand
- Multimedia Database Access

Video Conferencing is the most demanding of these applications and constitutes a superset of all their functionalities. This thesis focuses on video conferencing but the reader should be reminded that architectural design concepts herein are applicable to supporting any of the above services.

Current computing and network architectures readily support the transfer text and still graphics or images. Support for temporal media (audio and video) networking has, until very recently, been realized with entirely overlay networks and service specific terminal equipment. Voice telephony is the most pervasive media specific network. Broadcast video and cable TV also use dedicated transmission and switching infrastructure. High quality video conferencing today requires leased lines and expensive dedicated equipment (e.g. CODECs, MCUs, I-MUXs).

A number of critical technologies are now evolving to the point where high quality AV services can be supported in more cost effective and ubiquitous manner. Computing hardware performance continues to grow rapidly. Software development environments, languages, operating systems and applications increasingly support AV. Network transmission and switching technologies operate orders of magnitude faster than just a few years ago. AV compression technology has advanced to mature standards such as MPEG-2 that ensure multivendor interoperability and diminishing costs for CODECs (COder/DECoders).

Recently, low quality video conferencing and VoD (Video on Demand) services have become available on the Internet, utilizing Reflector servers, Modem speed access and software CODECs. While they are not yet ready for prime time, they represent significant advances in video conferencing system architecture and technology. Once the networking infrastructure and access speeds catches up with these new designs, there will be an explosion in the use of these services.

Video Conferencing

Video conferencing provides the ability to conduct a "face-to-face" meeting among two or more geographically distant conferees. Ideally this means being able to see and hear all distant conferees in real time (i.e.: essentially instantaneously for natural dialogue), either simultaneously or on a dynamically switched basis. In addition to the real time AV of participating conferees, whiteboards, text chatting and remote pointers and controls require loose real time support. As well, conferencing systems should permit the sharing of stored multimedia documents such as presentation slides (animated or still), shared work documents , images and audio/video clips. A more descriptive name for Video Conferencing is Multimedia Conferencing, but the prior is used herein.

The system infrastructure required to support video conferencing is, functionally, a superset of the requirements of the other multimedia network applications.

Videoconferencing is the focus of this thesis but it should be remembered that subsets of the videoconferencing system architecture are applicable to other multimedia applications.

Video Telephony

Video Telephony is a subset of video conferencing where only two participants exchange real time AV for a face to face meeting.

Video telephony is available today via a limited installed base of video conferencing systems and some analogue phone line terminals (that haven't enjoyed much commercial success). Relatively low cost Basic Rate ISDN CODEC cards from companies such as PictureTel are the best quality of these options. However, even these 128 kbps systems introduce significant end-to-end delay that impedes natural dialogue. Furthermore the frame rate and resolution of video on these low-end systems is certainly insufficient for the, perhaps small but certainly important, market segment of non-hearing users that communicate visually with their hands and face using sign language.

Beginning in 1997, low cost ITU-T standard H.324 Video Telephony over POTS line terminal products are expected to hit the marketplace. 1997 may well be the year where ubiquitous use of video telephony begins.

Broadcast Video and Video on Demand

In broadcast applications, there are multiple recipients of a single stream of audio/video. This is logically equivalent to a conference where one participant is the source and the rest operate as receive only observers. A broadcast service could be like cable TV today in that users can "tune in" channels that are broadcast on a pre-determined schedule. A Video on Demand (VoD) service adds an interface for users to browse a menu of available videos and initiate the playing of the selected video. A hierarchical (disk and tape) video storage server is the media source. Commercial VoD and broadcast services demand high-resolution full motion video and hi-fi stereo audio. End-to-end delay is of low importance in VoD systems unless VCR like controls are provided to the user.

Multimedia Database Server Access

VoD is a specific instance of a Multimedia Database Access service. A multimedia database service permits access to stored multimedia documents. In VoD these documents are movies. In general, hyper-media documents can include text, image, graphics, animation, audio and video. Applications of multimedia databases are numerous including computer based training and multimedia library access.

2.2. The Evolution of Video Conferencing

The first videophone was demonstrated by AT&T in 1964. These videophones utilized POTS lines for networking. Immature technology, high costs and poor quality kept these early devices from having any commercial success.

In the early 1980's, proprietary digital video conferencing systems became available from several companies including Compression Labs, PictureTel and GPT[1]. Compression Labs Inc. (CLI) achieved early market dominance in video conferencing systems with room systems using leased digital lines at 384 kbps to 2 Mbps operating in a point-to-point fashion. Video conferencing was at the time a high-end niche service; conference rooms generally costed on the order of $100,000 to equip and leased digital lines of these bandwidths were inaccessibly priced for most companies. Market growth was further constrained by the absence of standards and the necessity to buy from only one vendor, both for terminal equipment and Multipoint Control Units (MCUs) that facilitated three or more sites participating in a conference. Analog NTSC systems such as Datapoint's MINX also achieved some success during this period.

In the late 1980's, PictureTel began to obtain an increasing share of the video conferencing market due to their superior video coding technology that could operate at lower bitrates. By the end of the decade prices for their 128 kbps "roll-about" terminals dropped to the order of $50,000. Furthermore, the emergence of switched digital services (switched 56 and ISDN) facilitated on-demand conferencing rather than dedicated leased lines.

In 1990 the CCITT approved the basic elements of the H.320 suite of video conferencing standards. The H.261 video compression standard, coined Px64 since it used "P" 64 kbps channels, was central to H.320. All of the major video conferencing equipment vendors quickly came out with products that supported an H.320 mode of operation and several new players emerged including VTEL and British Telecom. True multivendor interoperability to some time to achieve but the presence of the H.320 standard and falling prices helped the rate of the overall market growth. The standards also allowed increasing competition an new companies emerged such as VideoServer to sell standard MCUs for multivendor multipoint conferencing.

Numerous proprietary systems currently exist for low quality video conferencing over LANs and the Internet. The most notable of these is White Pine Software's CU-SeeMe which originated at Cornell University. In addition to lack of multivendor interoperability, limitations in Internet performance and access speeds have kept these systems from being widely used.

In 1996 the ITU-T approved numerous other standards for video conferencing(see Appendix B). H324 for conferencing and video telephony over POTS lines using Modems, H.323 for conferencing over non-guaranteed QoS LANs and packet networks and H.310 for conferencing over ATM networks using AAL-5 or AAL-1. The T.120 standard for multipoint real-time data augments the video conferencing application with features such as whiteboards and pointers. The average cost for a seven port MCU is over $60,000 but CODECs can now be purchased for on the order of $10,000.

Activity in ITU-T SG-XVI continues at a fervoured pace in 1997 on further extensions to these standards suites, driven by numerous companies coming out with products in 1997 to support the much anticipated explosion in desktop video conferencing and consumer use of videophones.

2.3. Conferencing over N-ISDN and Leased Lines

Video Conferencing is still in an early phase of market development despite significant growth in the past few years. The most commercially successful segment of the market has been for so-called "P x 64" systems based on the ITU-T H.320 Standard Suite. Such systems aggregate from 2 to 30 DS0 (64 kbps) channels on switched or leased-line TDM networks into a wideband channel (128 kbps to 2 Mbps) to transport Audio, Video, Data and Control in a point-to-point manner. Multipoint Conferencing (3 or more sites)is achieved through a centralized MCU which typically mixes audio and multicasts the single current speaker to all sites. Video switching can be automated, based on speaker audio activity, or controlled manually by a chairperson.

ITU-T H.320 series standards enable interoperation between "P x 64" terminals and MCU equipment from dominant market players (PictureTel, CLI, GPT, Video Telecom, Video Server etc.) although each of these companies also support proprietary CODEC modes of operation that provide better performance both in quality and features.

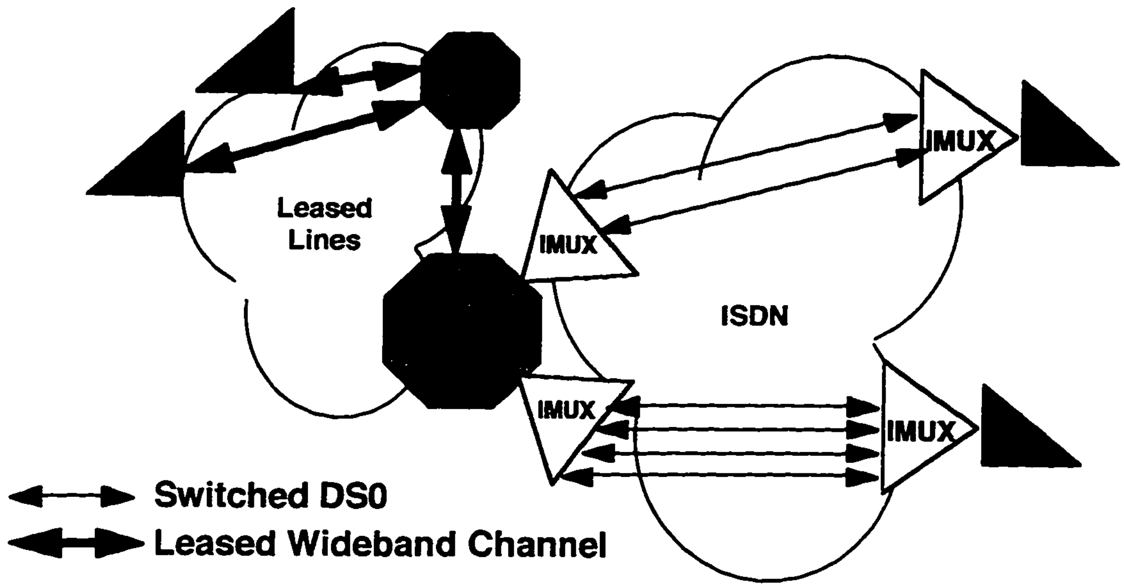


Figure 2-1 : Example Configuration of Network Components for Px64 Conferencing

More rapid growth of this market has been hampered by both high equipment costs and high service costs. In addition to the cost of Video CODEC terminal hardware, equipment costs are exacerbated in current networks by the need for I-MUXs (Inverse Multiplexors) to aggregate switched DS0 circuits (due the absence of wideband channel switching) and MCUs (Multipoint Control Units) (due to the absence of multicast switching) as illustrated in Figure 2-1. Service costs have been kept high due to bandwidth based tariffs needed to protect revenues from voice telephony. High bandwidth, high quality conferencing is still a luxury rarely afforded.

In addition to high costs, these highly complex systems are fraught with technical difficulties. Industry users report that multipoint calls are difficult to establish and maintain and that 80% of multipoint calls experience some technical problems[1]. Continuous presence conferencing is rarely available and performs poorly. Unless these technical problems are overcome videoconferencing will remain a small industry.

2.3.1. Multipoint Conferencing With H.3xx

All of the ITU-T H.3xx video conferencing suites rely on the same basic architecture to achieve multipoint conferencing: point-to-point connections to an MCU. With H.320 over ISDN or H.324 over POTS, pt-pt channels is an implicit limitation of the network technology. With LAN/IP or ATM networking (H.323 or H.310) the networks could utilize multicast connections but the standards do not support this. In 1997 work is being done on H.Loosely-Coupled which will allow some used of multicast to passive receivers in an H.323 conference. Aside from this limited use of multicast (suitable for lectures and public addresses) all the standard systems demand an MCU. The following sections introduce what functionality constitutes an MCU.

MCUs are a centralized "bridging" entity involved in coordinating how information is shared among more than 2 terminals. In the 'data' world of LANs, this type of functionality is typically handled in a distributed fashion. MCUs provide centralized bridging of audio, video and data. Coordination of events is relatively easy to handle with localized intelligence and signalling but scalability is limited. To enable larger conferences, multiple MCUs can be chained together.

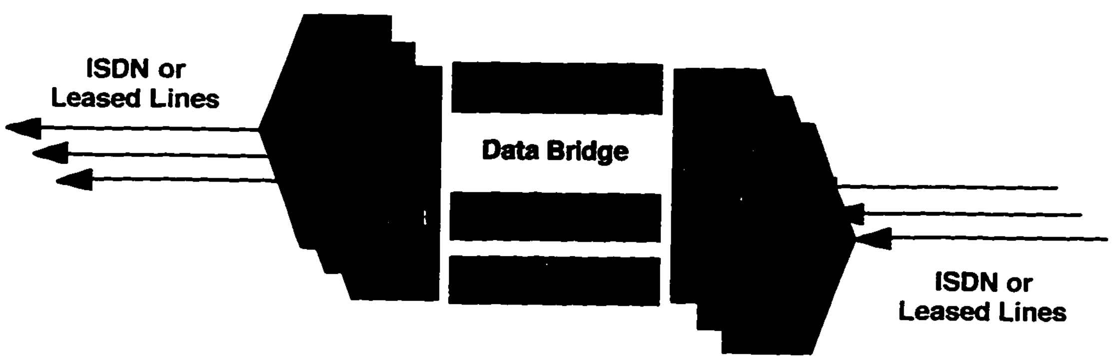


Figure 2-2 : Functional Model of an MCU

H.221 is a standard for time division multiplexing the various information media and signalling onto a single communication channel. At the conference bridge these components must be demultiplexed and handled separately by MCU sub-systems. These subsystems must be coordinated however. For example, in switched presence video conferencing, switching the current speaker 'on' could involve coordination with explicit manual controls or automatic audio triggered (volume of speaker) controls.

The MCU can have a range of capabilities including:

• Switching: select current speaker either automatically based on audio activity and/or manually selection by conference chair.

• Multicasting: sending the current speaker video stream to all participants except the current speaker who sees, typically, the previous speaker.

• Rate Adaption: Various terminal in the conference may be connected to the MCU at various line rates. The MCU may need to decode and recreate each CIF (Common Interchange Format) image and re-encode based on receiving terminals' line bandwidth (P=1 to 30). The alternative to a network service is to force all terminals operate at the level of the least capable terminal. This would result in poor quality for all regardless of local capabilities. One might want full T1 quality for all local locations but lower rates for international locations due to tariffing.

• Image Format Interworking: Not all CODECs need be capable of decoding full CIF (Common Interchange Format) images. Interworking between a CODEC transmitting at full CIF and another receiving at Quarter CIF could either be done by forcing both CODECs to operate at the least demanding level (QCIF) or by providing a network service to decode and recreate the CIF picture and re-encode based on receiving terminals' video CODEC capabilities(QCIF). In the opposite direction (QCIF to CIF) no interworking would be required assuming that CODECs that are full CIF capable are also QCIF capable as a subset of that functionality and re-encoding a QCIF signal into full CIF will not result in quality improvements.

• Audio Interworking: Conversions can be required between various audio coding schemes.

• Audio Summation: Audio mixing permits 'continuous presence' for multiple speakers; everyone can hear everyone else. Mixing is complicated by the fact that each persons audio should be omitted from the mixed stream sent back to them.

• Video multiplexing: Up to 4 QCIF combined into a single full CIF stream for continuous presence conferencing of up to 5 participating sites.

The conference bridge (MCU) is connected to the switch via one or more the ISDN PRI, or T1 trunks. Each PRI interface consists of 23 B + D channels. To conduct a multipoint conference, the D channel terminates on the switch and is used for conference setup, signalling and control. The B channels are used for audio, video, or data, in circuit switched mode between the participants and the conference bridge.

The network transport can be provided by non-switched connections (e.g., leased line) or switched connections (today almost). Inverse MUX capability uses independently routed B channel ISDN connections to transport p x 64 kbps stream. The Video channel is transported over multiple network connections (typically up to 6 B channels can be combined into one transmission channel by the Inverse MUX).

To handle a multipoint conference, a MCU requires one port for each location involved. Typical current MCU products can support up to 24 ports (usually 4 - 16).

2.3.2. Audio Bridge

The purpose of the audio bridge is to sum together the speech/audio signals from all the participants and to distribute this sum back to all participants. In practice, the bridge does not do a simple sum of all incoming signals. Other functions it performs include:

1) Echo Avoidance : A participants own voice should not be transmitted back to them in the summed audio signal as this will cause objectionable echoes. Hence the sums sent to speakers differ from that sent to passive listeners and the sums sent to different simultaneous speakers (if this is permitted) also differ from each other.
2) Noise Avoidance : Also to avoid noise buildup, not all incoming signals are summed. Often only the loudest speaker or a weighted sum of, say, the two loudest speakers is summed. Alternatively, all speakers exceeding a certain threshold are summed still avoiding accumulation of noise from idle channels.
3) Automatic Gain Control is often used to equalize speech volumes of distant & near or loud & quiet speakers.
4) Code conversions : m-law A-law coding conversions are already commonplace today for international telephony. A conference bridge conforming to H.221 must also be capable of understanding G.722 Sub-Band Adaptive Differential Pulse Code Modulation encoding and G.728 (16 kbps compression) DPCM encoding. In addition, it is interesting to note that even m-Law and A-Law PCM encoding is not

linear and signals must be linearized before they can be summed in a bridge. Possible encryption also complicates audio bridging.

2.4. Today's Market

Rationale for Use of Videoconferencing

Multipoint Conferencing provides users with a communications tool that enables many applications including cooperative work, distance education, presentations to large numbers of recipients and visual telephony for the hand-signing deaf. The most

Common arguments made in favour of investing in conferencing systems include:

Improved Productivity

- participants get down to business and reach decisions more quickly
- participants are motivated to be better prepared
- reduced large travel expenses
- savings in travel time and contract negotiation time

Provides Opportunities

- improves the availability of key people and resources
- expedites critical response time to changing market conditions
- increases flexibility in frequency and timing of business communication
- enhances centralized managerial control

Cost Advantages

- shortens the business cycle
- minimizes costly delay
- costs can be accurately forecasted
- contains and may reduce travel expenses

In 1979 industry analysts predicted that corporate videoconferencing would reduce travel by as much as 25%. This has not proved to be the case. People still love to travel and feel that face-to-face meetings are still the best way to build trust and team skills for long distance collaboration. Even in companies that have used executive mandate to force people to use video conferencing whenever possible have achieved no better than a 10%

reduction in travel[1]. Instead of travel reduction, the primary benefits of videoconferencing have been shown to be:

- Improved communication among working groups.
- Perceived personal productivity enhancement.
- Increased executive reach and presence.
- More timely communication.
- Ability to reach and involve people who would otherwise not have participated in important group or decision making.

The most significant applications driving growth in the videoconferencing market are [4]:

Business Conferencing: Increasingly companies are selling to global markets and establishing joint ventures and affiliates around the world. With increasing interaction between the North American and Asian countries in particular, high travel costs and time are becoming more significant than in the past.

Telecommuting: There is an increasing social change towards work-at-home and smaller satellite offices which also increases the potential demand for videoconferencing.

Telemedecine and Distance Learning: Videoconferencing enables medical service providers, schools, universities and scientific research centers to provide services to remote locations.

In 1995, the market for videoconferencing equipment and services totalled $2.94 billion[4]. Of this, roughly 80% was for network transport services (switched and leased TDM services) and the remaining 20% for equipment (terminals and MCUs). Data for 1992 to 1995 and forecasts up to the year 2002 are shown in the figures below.

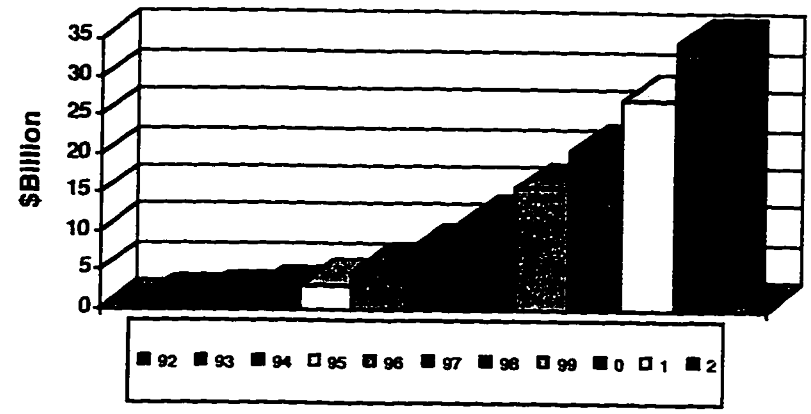


Figure 2-3 : Total U.S. Videoconferencing Systems and Service Market Revenue Forecasts 1992 to 2002 [4]

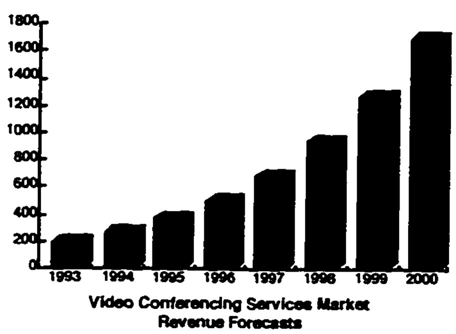


Figure 2-4 : Video Conferencing Services Market Revenue Forecasts, 1993 to 2000

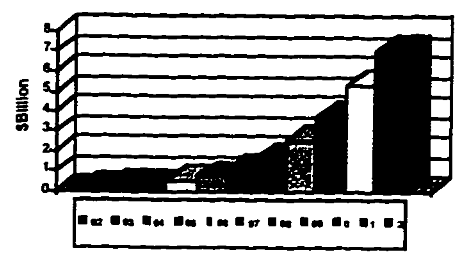


Figure 2-5 : U.S. Videoconferencing Equipment Market Revenue Forecasts, 1992 to 2002

Although the report [4] containing the above figures does state that with increased competition in the telecom marketplace prices of leased and switched digital lines continue to drop, the report does not consider the possible impact of emerging high speed data services, which may have flat monthly tariffs, may have on this forecast split of 80% services and 20% equipment. Once packet networks (e.g. Internet) have sufficient access speeds, core capacity, and integrated services support, the sales of equipment may well take a much higher percentage of the overall market.

In 1994, the North American video conferencing equipment market was dominated by room-based H.320 group systems from PictureTel, Compression Labs Inc. and VideoTelecom [2]. In 1995, this market was further fragmented suggesting a trend that will continue with the maturity of videoconferencing standards and deregulation of the telecommunication service industry.

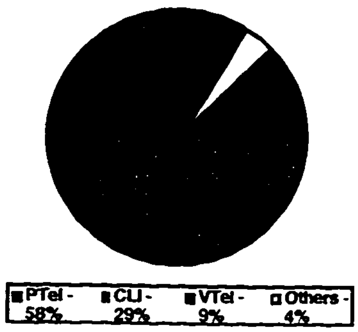


Figure 2-6 : 1994 Market Split for Group VC Systems (22,400 units) [2]

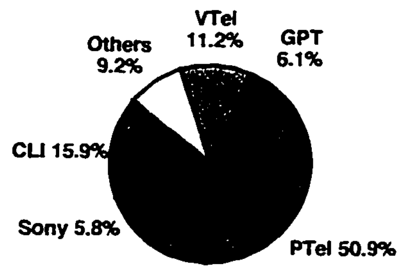


Figure 2-7 : 1995 Market Split for Group VC Systems [4]

In January of 1997 VTEL bought CLI for $80 Million in shares and the combined company will have revenues of $200 Million and a market share of approximately 25%.

PictureTel also dominated the MCU equipment market in 1994 as shown below along with VideoServer [2]

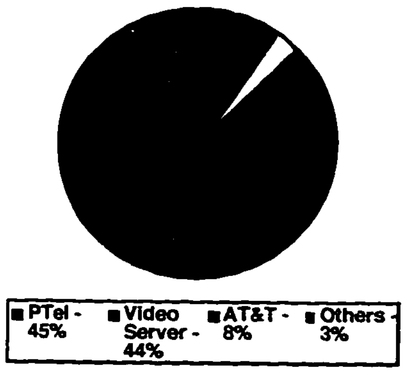


Figure 2-8 : 1994 Market Split for MCUs (900 units)

A more recent report [4] gives quite different figures for PTel's position in the 1995 MCU market; complete elimination from the market. This rather unlikely change draws some doubt about the figures tabulated in the report.

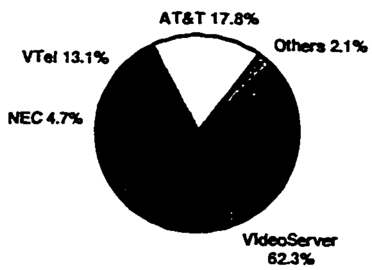


Figure 2-9 : 1995 Market Split for MCU Revenues

The 1995 video conferencing equipment market totaled $866 Million worldwide [3] ($510 Billion in the U.S.[4]) and is predicted to grow to $3.64 billion by the year 2000 [3]. This anticipated growth is not fueled by the currently dominant room based "group" conferencing systems but by dramatic increases in desktop conferencing , videotelephony and multipoint control units as illustrated in the following table.

	1995	2000
Group Videoconferencing	635	405
Desktop Videoconferencing	120	1,632
Gateways/MCUs	44	377
Videophones	37	582
Data Conferencing	30	646
Totals ($Millions)	866	3,642

Figure 2-10 : Worldwide Teleconferencing Market [3]

Figure 2-11 : Growth in Multipoint and Desktop Conferencing Equipment & Services

2.5. Emerging Video Conferencing Standards

By the end of 1996 the ITU-T had approved numerous H.3xx standards for video conferencing over various network technologies including Plain Old Telephony Service (POTS) and packet networks (LAN, IP and ATM networks). In addition, the T.120 standard for multipoint real-time data augments the video conferencing application with features such as whiteboards and pointers.

The existence of these standards will facilitate more widespread use of video conferencing through the development of interoperable products from many vendors for many network environments. Most significantly, low cost (low quality) desktop conferencing products will be tested in the marketplace.

All of these new systems maintain the same system architecture of the H.320 (Px64) systems which are commercially dominant today. Namely, they use point-to-point connections and centralized MCUs for multipoint operation. In addition, interworking units can be employed to enable conferences with terminals of different H.3xx types.

ITU-T Draft Recommendation H.24i describes requirements for interworking between the various H.series Multimedia Terminals including H.323, H.320 H.324 and H.310 and H.324Mobile terminals (Call Control, System Control and Media Flow) using Interworking Units. The areas where interworking is required can be seen by looking across the rows in the following table.

Network	POTS	N-ISDN	Guaranteed QoS LANs	Non-guaranteed QoS LANs	ATM (B-ISDN, ATM LANs)
Channel capacity	up to 28.8 kbps	up to 1536 or 1920 kbps	up to 6/16 Mbps	up to 10/100 Mbps	up to 600 Mbps
Characteristics	• ubiquitous	• circuit based • existing	• similar to N-ISDN	• packet loss prone	• future basic network
Total system (date of the first approval)	H.324 (96/03)	H.320 (90/12)	H.322 (96/03)	H.323 (96/11)	H.310 (96/11), H.321 (96/03)
Audio coding	G.723.1	G.711, G.722, G.728	G.711, G.722, G.728	G.711, G.722, G.723.1, G.728	G.711, G.722, G.728, ISO/IEC 11172-3
Video coding	H.261, H.263	H.261	H.261	H.261, H.263	H.261, H.262
Data	T.120 etc.	T.120 etc.	T.120 etc.	T.120 etc.	T.120 etc.
System control	H.245	H.242	H.242	H.245	H.242 (for H.321), H.245 (for native H.310)
Multimedia multiplex and synchronization	H.223	H.221	H.221	H.225.0, TCP/IP etc.	H.222.0, H.222.1
Call setup signaling	National standards	Q.931	Q.931	Q.931, H.225.0	Q.2931

Figure 2-12 : Audiovisual Communication Systems for Various Networks.

Figures showing the protocol stacks for the various H.3xx terminals and an exhaustive list of ITU-T Recommendations for these systems can be found in Appendix B.

2.6. Limitations of Today's Network Technologies

- Wideband Switching (eg. ≥ 384 Kbps) Unavailable
  - TDM Networks
    - 56-64 Kbps (DS0) channels ($ Services)
    - $ Channel Aggregation Hardware for N x 64 Kbps
  - Packet Networks (Internet, LANs, FRS, SMDS, X.25,...)
    - Unpredictable Bandwidth and QoS
- Lacking Multipoint Transport
  - Circuit Switched networks:
    - point-to-point
    - Multipoint Control Unit hardware cost $ 100K
  - Packet networks:
    - LAN multicast/broadcast must be used cautiosly
    - WAN multicast services not widely & readily accessible
    - protocol development is immature

Cost Barriers to Ubiquitous Use of Video Conferencing Significant

Figure 2-13 : Key Limitations of Today's Networks

ISDN: Limitations & Costs

ISDN has no wideband (>64 kbps) switching. H.261 requires at least 384 kbps to achieve bearable quality and more bandwidth to approach VHS videotape quality. Circuit switched networks such as the ISDN support dynamic dial-up 64 kbps channels while only provisioned leased-line networks can provide static channels of Px64 kbps (N = 1 to 24). On demand video conferencing requires multiple (P) ISDN 64 kbps (DSO) channels. This has two significant cost impacts. First of all, tariffs on DSO channels are lower bounded by tariffs for voice telephony. Service providers can not lower these tariffs without impacting their bread and butter voice telephony revenues. Users can have a difficult time justifying that a video conference is P-times more valuable than an audio conference. Secondly, users must purchase Inverse Multiplexors (I-MUX) to eliminate delay differential between the P independently routed DSO channels and create a single contiguous Px64 kbps channel.

ISDN's channels operate point-to-point. There is no support for more complex multipoint connection topologies involving three or more terminals. Hence, to facilitate multipoint conferencing a centralized Multipoint Control Unit (MCU) must be purchased. MCUs can cost on the order of $100K. The two dominant suppliers are PictureTel and VideoServer.

The following table illustrates the impact of narrow bandwidth and MCUs on delay. This data is from a network using leased lines; further delay would be incurred for channel aggregation in a switched network. Delay values are one-way delay.

Bandwidth	Point-to-Point	Multipoint
128 kbps	700mS	1000mS
384 kbps	340mS	380mS

Figure 2-14 : Delays for H.320 Conferencing

As can be seen, even at 384 kbps, round-trip delay is significant and certainly impedes natural dialogue.

Conferencing on the Internet and Desktop Video

In addition to ISDN/H.320 conferencing, there is a new low-end market developing for desktop video conferencing. These products either use Basic Rate ISDN (2x64 kbps) such as a new PictureTel/AT&T/Intel product, or packet networks (LANs or the Internet). Popular examples of Internet desktop conferencing systems are:

- MBONE: a free Internet Audio Visual conferencing tool using software based decoding.
- InSoft Communiqué: a commercial conferencing application using motion JPEG hardware compression/decompression.
- CU-SeeMe: Originally developed at Cornell University, this software-based video conferencing system has been commercialized by White Pine Software. It offers the most user friendly user interface of any system available today.
- Isabel: Another motion JPEG-based system developed in Europe.
- Emerging H.323 (see next section) products for video conferencing over packet networks.

None of these systems support adequate quality for natural dialogue. Delays are significant (usually over 1 second round trip), frame rates & spatial resolution of images are low and lip-synchronization is largely unsupported. The packet network (Internet Protocol) based systems are even more constrained than the ISDN based systems due to unpredictable network bandwidth and Quality of Service (QoS).

With the growing ubiquity of the Internet, numerous distributed desktop conferencing systems have emerged in the past few years. MBONE has been used to broadcast IETF meetings across the Internet using coding such as Sun CellB, CU-SeeMe or NV. Commercial products such as Insoft's Communiqué have enjoyed some success and other research systems such as Isabel also have used Parallax motion JPEG compression hardware to achieve multipoint conferencing. None of these Internet systems are very enjoyable to use yet; quality is low in terms of frame rate, resolution, audio fidelity and delays. Effectively these systems are all constrained by inadequate network access rates and unpredictable QoS. The MBONE system and it's separation of audio and video flows has many parallels to ATM RendezView but its usability is terribly constrained by todays IP networking. However, these systems do represent a significant rethinking of the architecture of a video conferencing system. These new architectural concepts will be ready when the Internet and Intranets evolve to higher speed access and on-demand connection oriented services with QoS guarantees.

At modem access rates, transferring still images alone is a challenge and can take minutes per image. Despite this, “live” (slow animation) video broadcast and conferencing systems are in some use on the Internet. VDOnet Corporation's VDO Live and White Pine Software's CU-SeeMe are examples. Products from these companies include client applications, video servers and MCU -like “Reflectors”. These provide slow frame rate and low fidelity audio using software CODECs. Services already available via these systems include news, speaker presentation broadcasts, movie industry advertisements, etc.

In addition to QoS constraints of today's Internet, the architectures being used by these AV systems still have some limitations resulting from centralized reflectors that become congestion points for multiple unicast streams. Multicast on the Internet is still in its infancy with a scattering of Multicast Routers comprising the MBONE (Multicast Backbone).

2.7. Evolving Technologies

2.7.1. Evolving Image and Video Coding Technologies

Numerous proprietary, platform specific or defacto standard techniques for video and image compression , storage and transfer are pervasive on the Internet today including Compuserve's GIF (Graphics Interchange Format), Microsoft's AVI, Apple's QuickTime, Sun Video, VDOnet's VDO and White Pine Software's CU-SeeMe. While all of these systems demonstrate what is feasible with image and video over the Internet, multivendor standards from the ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) represent the future of digital image and video with hardware support for JPEG and MPEG. JPEG stands for Joint Photographic Experts Group and MPEG stands for Motion Picture Experts Group. Current commercial Px64 video conferencing systems use ITU-T H.261.

JPEG compresses images to smaller size files using Discrete Cosine Transform (DCT) based compression to remove two dimensional (spatial) redundancy in an image. This results in some loss of image quality depending on the compression factor used. JPEG compression and decompression can be done in software but hardware compression cards are available for a large number of platforms thanks to chips developed by companies like C-Cube since around 1990. Hardware JPEG Coder/Decoders (CODECs) facilitate Motion JPEG video at up to 30 image frames per second. Because of the early availability of affordable JPEG CODECs like those from Parallax, JPEG has been used extensively for prototypes and early-products for digital video appiications. However, motion JPEG will be superseded by MPEG which achieves higher levels of compression by further removing temporal redundancies.

Image Dimensions	Uncompressed Data Rate	Quantization Factor = 25	Quantization Factor = 100	Quantization Factor = 400
640x480	221 Mbps	18 Mbps	7 Mbps	3.4 Mbps
320x240	55 Mbps	5 Mbps	2.5 Mbps	1.2 Mbps
160x120	14 Mbps	2 Mbps	720 kbps	360 kbps

Figure 2-15 : Measured Data Rates for 24-bit Motion JPEG at 30 Frames per Second

The acceptability of video quality at various compression factors is subjective and application dependent.

MPEG achieves a higher level of compression than JPEG by removing redundancies that occur between video frames. For example, a stationary background region need not be retransmitted or a motion vector can be sent for a region of the image that is moving. MPEG still encodes JPEG-like DCT coded frames periodically so the further compression is not achieved for every frame but instead for a large fraction of them, the number decided in compression options. Typically, MPEG can result in compression two to six times greater than Motion JPEG. MPEG-1, widely used on CD-ROMs supports a 352x288 pixel image size generating 1.15 Mbps. MPEG-2 extends MPEG-1 to provide more capabilities to compress various image dimensions for various applications.

Application	Resolution	Bit Rate	Quality
HDTV High Definition TV	1920x1152/1080 60 fps	20-30 Mbps	Awesome
EDTV Enhanced	960x576 25/30 fps	8-10 Mbps	Better than Studio
SDTV Studio NTSC/PAL/SECAM	720x576 25/30 fps	4-7 Mbps	Studio Quality PAL/NTSC/SECAM
Broadcast NTSC/PAL/SECAM	540x480 30 fps	2-6 Mbps	Broadcast TV
LDTV Low	352x288 30 fps	1-2 Mbps	VHS VCR
H.261	QCIF 176x144 or CIF 352x288 1-30 fps	128 kbps to 2 Mbps	Good Quality Video Conferencing
Miscellaneous Internet AV Systems	less than QCIF up to seconds per frame	10 to 64 kbps	Low Quality Video Conferencing

Figure 2-16 : Typical bit rates for MPEG-2 and H.261.

The examples in Figure 2-16 are illustrative only. Actual values are implementation dependent.

Hardware support for MPEG compression is imperative, though some software based decompression is achievable. Real time MPEG-2 encoders are still relatively expensive, starting at the order of $10K and hence are not yet commonly available to the consumer computing market.

Due to tight constraints on end-to-end delays for interactive dialogue video conferencing services, it is not feasible to obtain the same level of MPEG compression for conferencing applications that is feasible for broadcast and database applications. High levels of compression require the buffering and processing of multiple video frames. Each frame buffered results in 33 mS of delay (or more for frame rates less than 30 fps).

2.7.2. Audio Coding

The following table summarizes the characteristics of available standards for audio coding.

Standard	Encoding Technique	Analog Audio Bandwidth	Digital Bandwidth	Quality
G.711	8-bit Logarithmic PCM	3.4 kHz	64 kbps	Voice Telephony (μ-Law , A-Law)
G.721	Adaptive Differential PCM		32 kbps	Equivalent to G.711
G.722	sub-band ADPCM	7 kHz	48, 56 or 64 kbps	Better than G.711
G.728		3.4 kHz	16 kbps	Inferior to G.711
CEL-P Code Excited Linear Prediction	Sub-band and Vector Quantization		4.8 kbps	Slightly Inferior to G.711
Federal Standard 1015 LPC Linear Predictive Coding	Sub-band and Vector Quantization		2.4 kbps	Intelligible but Artificial
CD	16-bit Linear PCM	20 kHz	705.6 kbps mono 1.4112 Mbps stereo	Home Hi-fidelity audio
Sound Studio Quality			twice CD	
MPEG Layer 2 (MUSICAM)			192 to 256 kbps stereo	
MPEG Layer 3			64 kbps	near CD Quality

Figure 2-17 : Standards for Audio Coding

2.7.3. IP Evolution to Multiservice

The current Internet supports best effort connectionless delivery of data from point to point. Data is subject to unpredictable losses and long queuing delays. Multicast IP is supported only on the Multicast Backbone (MBONE), a sparse collection of multicast routers interconnected by "IP Tunnels" through conventional routers. Multipoint services must, in general, use centralized reflector servers to operate.

Because of these factors, the Internet does not yet support real-time audio/video communications with adequate performance or reliability. These problems have been acknowledged in the Internet Engineering Task Force (IETF) community which has several parallel efforts underway aimed at the evolution of the Internet to support multiple integrated services. These extensions to the Internet will be built upon advances in fiber optic transmission and high-speed switching. The key control elements being developed are support for multipoint connections and QoS guarantees through resource reservation. RSVP (Resource Reservation Protocol) is the "signalling" protocol proposed for the planned Integrated Services Packet Network.

Most of the features desired for the Internet parallel those of ATM (Asynchronous Transfer Mode) networking which was developed for the public Broadband ISDN. With demand for Internet services exploding and mature products supporting IP over ATM, ATM is beginning to be deployed for upgrading Internet backbones. This increases the likelihood that wide area ATM infrastructure will be available one day for improved conference services. ATM is also being deployed into corporations' Intranet backbones. The ATM Forum MPOA standard will play a significant role in encouraging the deployment of hybrid IP/ATM networks.

MPOA (Multi-Protocol Over ATM) is a standard from the ATM Forum for supporting high-performance, scalable IP networking over ATM networks. The MPOAv1 specification is scheduled for completion in mid 1997. MPOA builds upon existing mature ATM and IP standards and products enabling network operators to build multivendor campus, MAN and WAN private, public and Virtual Private networks.

MPOA complements the capabilities of Emulated LANs (LANEv2) to support large IP networks. LANE enables LANs to be emulated over ATM networks permitting existing network applications to operate unchanged. The size of a LAN (emulated or not) is constrained however, and IP routing is required to interconnect LANs and build large networks that can have worldwide geographic distribution. MPOA provides this routing and, furthermore, improves the performance available with traditional routers.

MPOA splits the forwarding and routing functions traditionally supported within an IP Router between MPOA Clients and MPOA Servers respectively. This facilitates setting up ATM connection shortcuts for IP flows that can then bypass intermediate IP router hops, reducing the load on the routers and improving the performance for IP transfers both in terms of bandwidth and delay.

Not only does MPOA improve performance for existing data networking applications but it provides an evolution path to support emerging multimedia applications that involve continuous flows of voice and video that require Quality of Service (QoS) and bandwidth guarantees. ATM was designed specifically to support these types of multimedia applications.

In the following sections, the impact of Asynchronous Transfer Mode (ATM) networking is explored. ATM is beginning deployed by telecom service providers for use as backbone transport infrastructure for all services. This provides an opportunity to re-examine multipoint conferencing techniques to design systems that will perform better and at a lower cost by utilizing capabilities unavailable in previous networks. Some ATM vendors already have video CODECs for conferencing over ATM networks but these are sold only in very high quality small niche markets (e.g.. NT's DV45 and Newbridge/MPR motion JPEG) and require 10's of Mbps bandwidth capacity.

3. Multipoint Services on ATM

While the concept of switched broadcast (one-to-all) and multicast (one-to-many) network transport has been around for many years, it is not widely supported or used on today's networks. The telephone network (POTS or ISDN) offers a unicast (point-to-point) service exclusively. The installed base of the Internet and most Enterprise Intranets support only unicast datagram service aside from a limited logical overlay of the MBONE (Multicast Backbone) that supports some limited multicast via advanced configuration. Frame Relay networks support some multicast, at least on a PVC (Permanent Virtual Connection) but these capabilities are not widely used. Even on LANs, multicast and broadcast are used sparingly for applications such as ARP (Address Resolution Protocol) to avoid unnecessary loading on all attached terminals that would result from wider use.

In tomorrow's Integrated Services networks, point-to-multipoint connections will be offered as a generic on-demand network service. Ubiquitous availability of such a capability permits a dramatic rethinking of architectures used to support multipoint network services. In particular, centralized service-specific network equipment can be eliminated while simultaneously achieving optimal service performance with maximal network efficiencies.

The question of what the standard and ubiquitous service and programming interface will exist to tomorrow's network is still open. Whether it is to be based on "raw" ATM U-plane and C-plane protocols via WinSock, or IPv6, RTP & RSVP, or perhaps some yet proposed interface is moot to be basic architectural impact that such capabilities will have.

This section describes how Asynchronous Transfer Mode (ATM) layer bearer service connections can be used to support multipoint services. ATM technology remains today's most likely basis for tomorrow's network, independent of the service interface presented to the user. Understanding the generic capabilities at the ATM layer reveal what functionality is available to the distributed application architect.

3.1. ATM Basic Basics

This thesis makes no attempt to fully describe ATM networking. The reader is expected to be familiar with ATM networking fundamentals.

The ATM bearer service provides for unassured ordered delivery of ATM cells which stream along a path which is fixed for the duration of the connection. Cells are 53 octet packets; a cell has a 5 octet header and a 48 octet payload. The header contains a connection identifier field structured in a 2-level hierarchy of a VPI (Virtual Path Identifier) and a VCI (Virtual Channel Identifier).

Switches use either the VPI only or both VPI/VCI concatenated to switch streams of cells, depending on the service requested. At the User-Network Interface (UNI) there are 256 VPIs (2^8) each containing 64K VCIs (2^{16}). Network-Node Interfaces (NNI) support 4,096 VPI (2^{12}). Users can either obtain a VP connection or VC connection service from the network.

These connections can be unicast (point-to-point) or multicast (point-to-multipoint). ATM connections can be PVC (Permanent Virtual Connections), manually configured via network management , or SVC (Switched Virtual Connections), dynamically configured via ITU-T Q.2931 signalling.

Independent of connection topology or dynamics, each connection has QoS (Quality of Service) and bandwidth parameters associated with it. This enables deterministic or statistical allocation of network resources (transmission lines and switch buffers) by the network operator in order to provide a "pipe" of the right size, shape and reliability to the user to suite the particular application.

- Performance
  - wide bandwidth switched channels available
  - admission control for QoS gurantees
  - excellent error performance on fiber optic networks
  - media specific adaptation
- Flexibility
  - selectable bandwidth and QoS
  - multipoint connections on demand

Reduced Costs

- Broadband economies of scale
- no IMUX or MCU
- service-based tariffing

Figure 3-1 : ATM's Advantages over ISDN

Bandwidth capacity of ATM networks is orders of magnitude higher than existing networks. Interfaces operate at 34 Mbps (minimum access speed) to 1.2 Gbps (or higher as state of the art in fiber optics evolves). Wideband/broadband capacities (such as those required to carry video) can be allocated to individual connections, eliminating the need for DS0 channel aggregation via I-MUXs (Inverse Multiplexors).

Existing circuit switched Integrated Services Digital Network (ISDN) and provisioned leased-line networks provide only single-point to single-point channels for the flow of voice, video or packet data. This necessitates centralized bridging, the MCU for conferencing or the router for data.

ATM networks, on the other hand, have the fortunate characteristic of also supporting more complex connection topologies than Point-to-Point (Unicast). A spanning tree of ATM connections can support Point-to-Multipoint (Multicast) and Multipoint-to-Point (Merge) flows of information as well. Such connections offer optimal performance and minimal consumption of network resources. This affords ATM many advantages for supporting multipoint video conferencing.

3.2. Multipoint Applications & Multicast Services

There are two primary reasons to support multicast services.

1) Bandwidth Efficiency: Multiple Unicast transmissions (source replication) results in inefficient use of link bandwidth. In a unicast service environment, UNI; bandwidth constrains the product of the number of destinations and the rate at which information can flow to them. Using distributed network replication over a spanning tree, minimal resources consumed and minimal delay is introduced.

2) Application and Terminal Simplicity: Certain applications are made simpler to develop if the network can support multipoint connectivity. For example, the number of Virtual Connections a terminal must manage can be reduced and terminals (i.e.: Routers) can self-configure by using ARP. When used in Multipoint conferencing, Multipoint connections simplify the process of terminals joining or leaving the conference by eliminating the need for control interaction with all other conferees.

3.2.1. Video; - Key Driver for Multicast

Video distribution is an application for which point-to-multipoint connectivity is clearly required. Presuming video bandwidth demands are a relatively high portion of UNI capacity, it is infeasible to replicate a video stream at the source to unicast to multiple destinations. Even where this is feasible for a small number of destinations, it is extremely inefficient use of UNI bandwidth. Video & Audio distribution, also required in conferencing, require multicast services especially because they involve a continuous flow ("real-time") of information.

3.2.2. Data Applications Requiring Multicast

The multipoint capabilities of ATM have been driven in their early phases by data applications as described below. As a result these facilities are available for use in conferencing as well.

Multipoint connections will be demanded for data applications such as router Address Resolution Protocols (ARP) (see Appendix D). This was made evident by the proactive work done by router vendors to ensure a multicast service was defined within the Frame Relay Forum for precisely this application. The Closed User Group (CUG) of ATM interconnected routers that constitute a customers Virtual Private Network (VPN) should have the capabilities of a typical LAN-based subnet including multicast connectivity used for ARP. While it is technically feasible for a router to perform multiple unicast ARPs within a limited size CUG, such a requirement would necessitate software development by router vendors to support this different paradigm of operation on the WAN port.

Hence, to support immediate market needs for LAN interconnection, multicast ATM connections should be supported at least for ARP applications. The primary benefit of this is to terminals (routers) which can use existing ARP implementations developed for shared medium LAN subnets. The absence of multicast capability in either FR or ATMbearer services would encourage users to use the Switched Multimegabit Data Service (SMDS; I.364), a network service that does support multicast & ARP.

In addition to simplifying terminal application design and implementation, as multicast application bandwidth increases and the number of group members increase, the requirement for multicast service will become more absolute. Multiple unicasts become infeasible for the same reasons as with video when the aggregate traffic generated consumes a large percentage of link (UNI) capacity. Multimedia and shared-space applications are already beginning to be developed such as the MBONE audio visual conferencing performed on the Internet during Internet Engineering Task Force (IETF) meetings.

Furthermore, using multipoint connections significantly reduces the complexity of a terminal joining or leaving an active multipoint conference. The network need only perform a point-to-point route selection from the terminal to the existing multipoint tree and add a branch to that tree.

3.3. Building Blocks: ATM Connection Topologies

Establishing connectivity from a group of user terminals engaged in a multipoint conference can be achieved in a variety of ways using the basic ATM bearer channel types. The following subsections describes the basic ATM connection topology types: Point-to-Point, Point-to-Multipoint and Multipoint-to-Multipoint. The use of Multipoint-to-Multipoint connections (Bidirectional Multipoint Virtual Path) was considered for multipoint conferencing but ultimately Point-to-Multipoint was chosen for the RendezView Prototype.

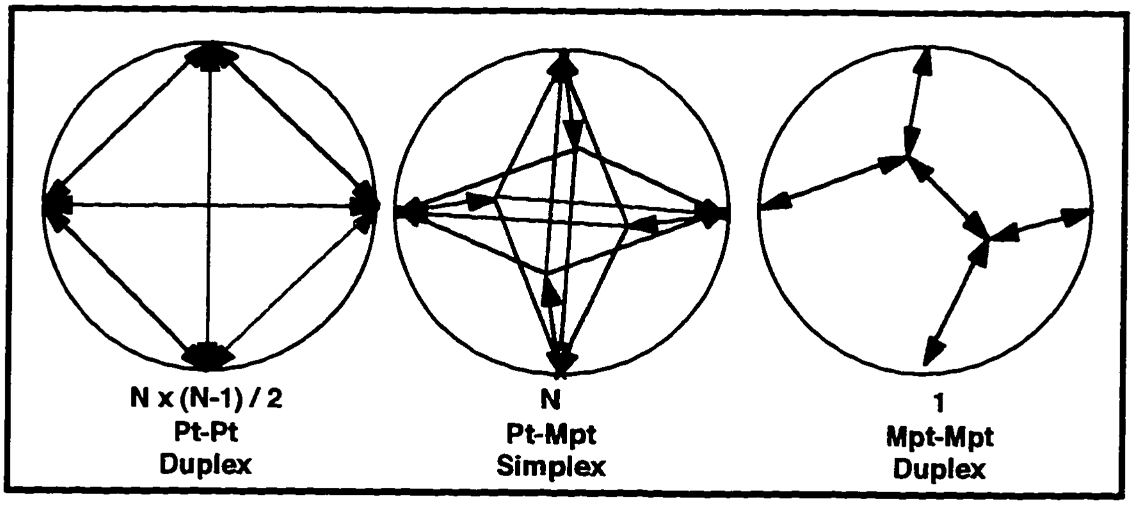


Figure 3-2 : Options for Achieving Multipoint Connectivity

3.3.1. Point-to-Point (Unicast)

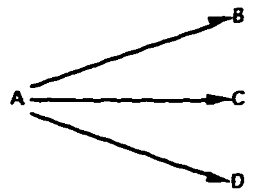


Figure 3-3: Three Unicast Connections

Unicast connections should be used whenever a single destination for a transmission is known by the source. Unicast connections can be viewed as a simple subtype of a general multipoint connection.

Asynchronous Transfer Mode Point-to-Point connections can be either simplex or duplex. When they are duplex, they follow the same path in both directions. They can be provisioned through the Management Plane (M-Plane) as static Permanent Virtual Connections (PVC), or dynamically through the Control Plane (C-Plane) User-Network Signalling as Switched Virtual Connections (SVC). Connections can be wither Virtual Paths or Virtual Channels. Switched Virtual Paths are not supported in the current phase of UNI signalling standards.

Complete multipoint connectivity (each to every) is potentially required for multipoint conferencing; any conferee can become the current speaker and their audio/video stream must be transferred to all others. When only point-to-point connections are utilized, either the expensive centralized MCU must be resurrected or a full mesh of N x (N-1) / 2 bidirectional point-to-point connections as shown in Figure 3-4.

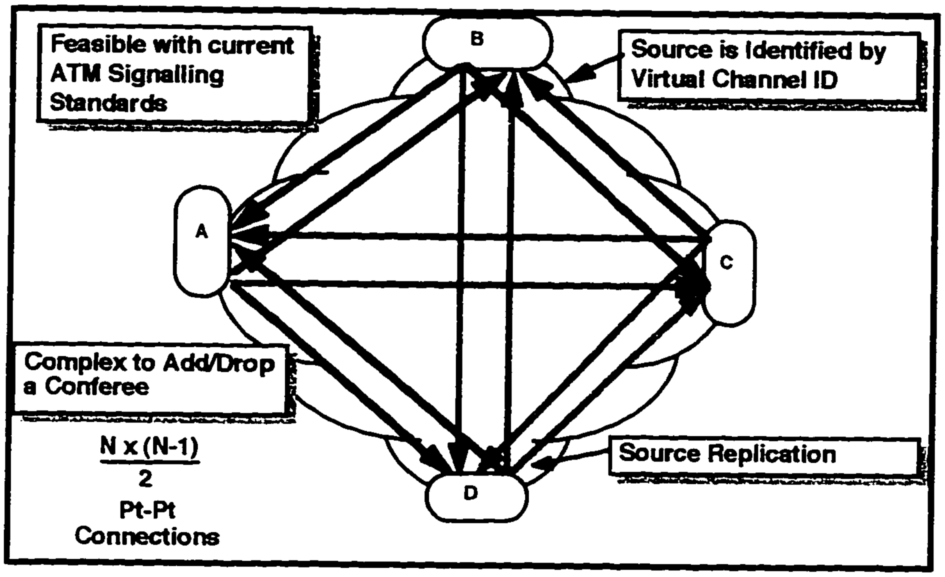


Figure 3-4 : Multipoint Connectivity Using Point-to-Point Connections

This is the technique utilized be current versions of InSoft's Communiqué Internet conferencing product. This architecture results in significant performance hits. Terminals and access links are excessively loaded with multiple and redundant continuous media streams. This adds substantial transfer delays, terminal loading and constrained scalability. With Communiqué, if N is greater than 2, the added delay is significant and lip synch is gone.

3.3.2. Point-to-Multipoint (Multicast)

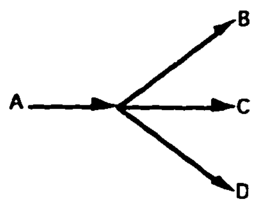


Figure 3-5 : One Multicast Connection

Multicast connections should be used whenever an multiple destinations for a transmission exist. Multicast connections are of critical importance for some applications, especially applications which both demand high continuous bandwidth and are real-time in nature. Using a shortest-path, point-to-multipoint spanning-tree for the connection minimizes consumption of both terminal and network resources and guarantees maximal performance (minimal delay).

- Video distribution or multipoint video conferencing services clearly fall into this category. Information streams must be delivered continuously at whatever bit-rate is demanded by the source to arrive more or less simultaneously at multiple locations.
- Some data applications which have traditionally run over broadcast shared-medium LANs will also benefit from multicast connections which improve efficiency, performance and simplicity for the source host. Interconnected remote routers can use multicast for Address Resolution Protocols and distribution of routing table updates. LAN emulation is another user of ATM multicast connections).

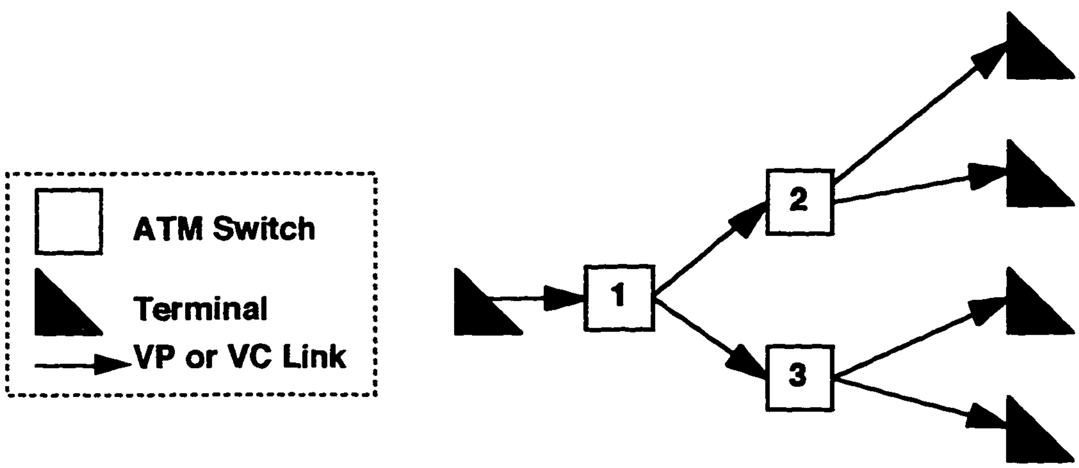


Figure 3-6 : Multicast Spanning Tree Connection

In the Figure above, terminals A through D belong to a Closed User Group (CUG) having multicast connectivity. The point-to-multipoint (pt-mpt) connection from A is illustrated. Similar connections with B, C & D as their roots also exist and are shown in the diagram below only to illustrate the connection complexity. These connections may or may not follow the same path as A's pt-mpt tree. N pt-mpt connections are required to support N terminals.

A pt-mpt spanning tree connection is ideal from a bandwidth efficiency & performance perspective. Replication of cells occurs at the last possible point in the network (at switches 1 & 2) so that only one copy exist on a given physical link.

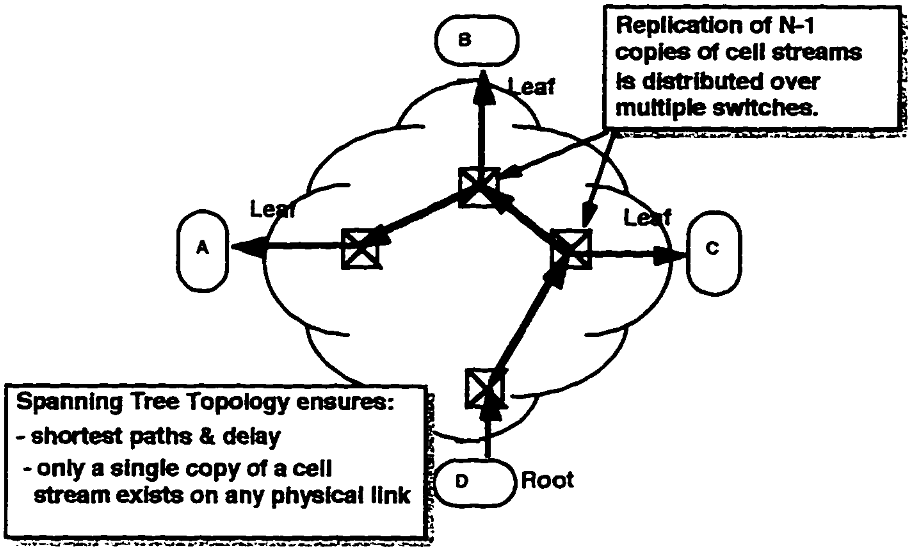


Figure 3-7 : Point to Multipoint Spanning Tree

ATM Point-to-Multipoint connections provide the user with a unidirectional multicast capability although the connection is established bidirectionally through the network as illustrated in the figure below. The Multipoint-to-Point direction of the connection is reserved for use by network management for OAM (Operations Administration and Maintenance) cells that can monitor connection performance or trace faults. Only management cells can travel along in this "merge" direction; zero bandwidth is allocated for user traffic.

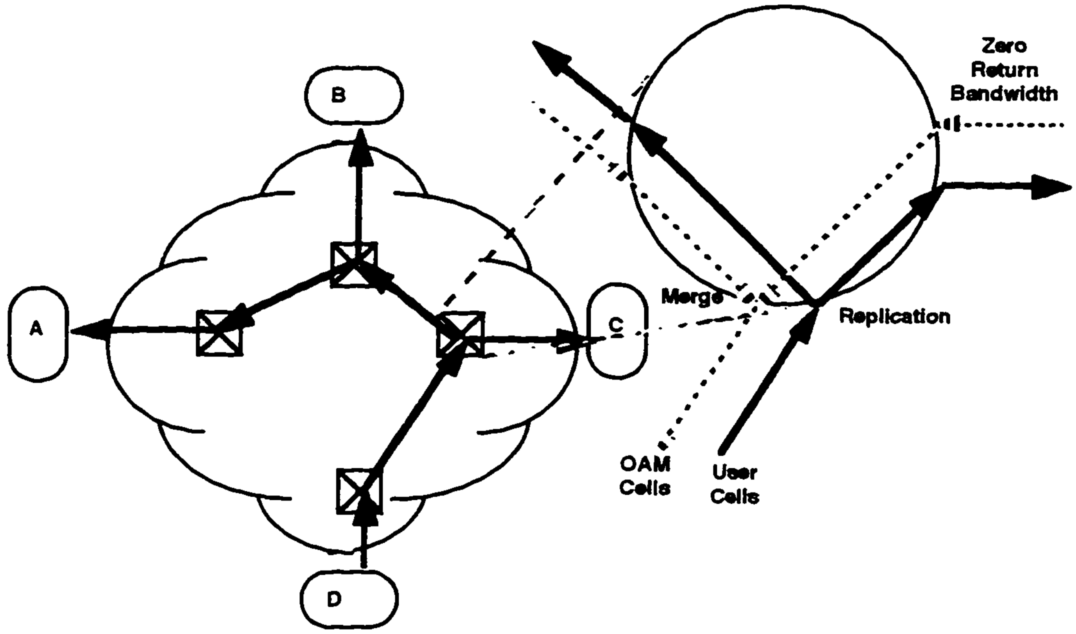


Figure 3-8 : Bi-Directional Multicast Connection Mapping in the Switch

ATM Point-to-Multipoint connections can also be established either by the M-Plane (Management) or C-Plane (Control signalling). Current standards support only unidirectional Point-to-Multipoint connections. Multicast connection trees are built one branch at a time; no atomic SETUP is supported to establish the entire tree via one request from a terminal. Either the Root of the connection spanning tree, or a Leaf can initiate joining/leaving the tree, thus adding/removing a branch on the tree. The following example illustrates the process for building a multicast connection.

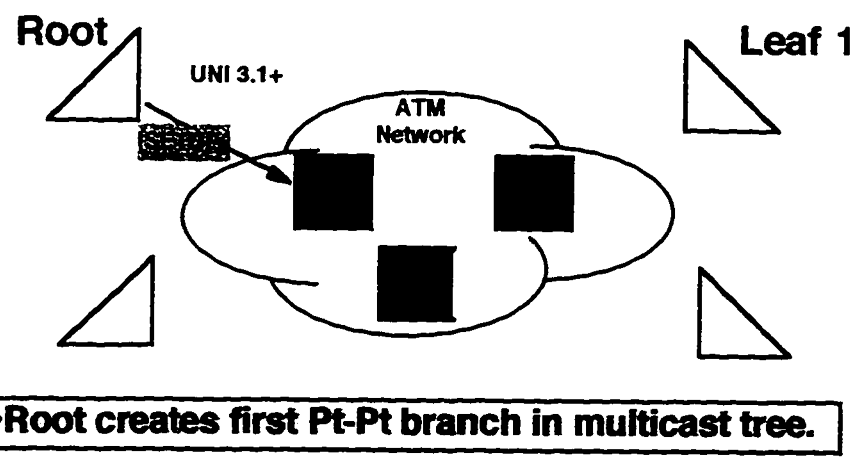


Figure 3-9 : Multicast Connection Control SETUP

With ATM Forum UNIv3.1 or greater signalling, a multicast connection can be initiated by either the root (source) or a leaf (sink) of the connection. The case where the root initiates the connection is shown above. The Root issues a SETUP message indicating that the connection will be multipoint, the ATM address of the first Leaf, and QoS & bandwidth parameters for the connection. Alternatively, the Leaf could have initiated the same event by issuing an LEAF_SETUP_REQUEST message.

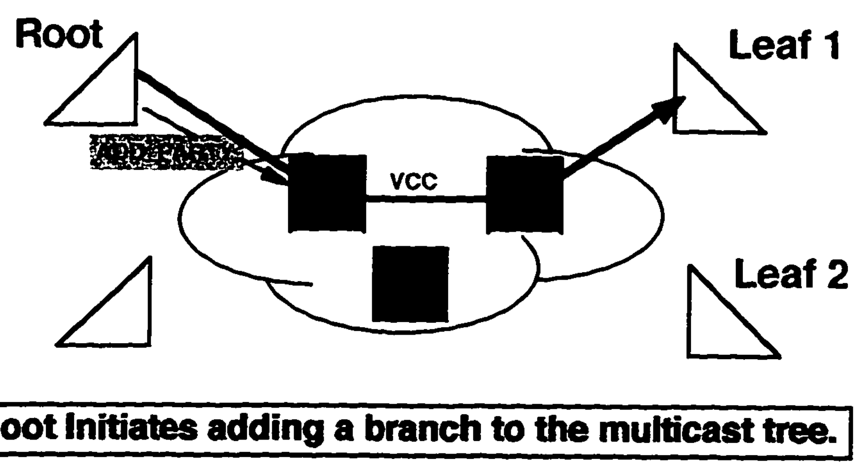


Figure 3-10 : Multicast Connection Control ADD_PARTY

In response to the initial SETUP message, the network establishes the first branch of the Virtual Channel Connection illustrated by the thick black arrowed lines above. The Leaf had to accept this connection for it to be established. Subsequently, the Root can now add another leaf to the existing connection by issuing an ADD_PARTY message as shown above.

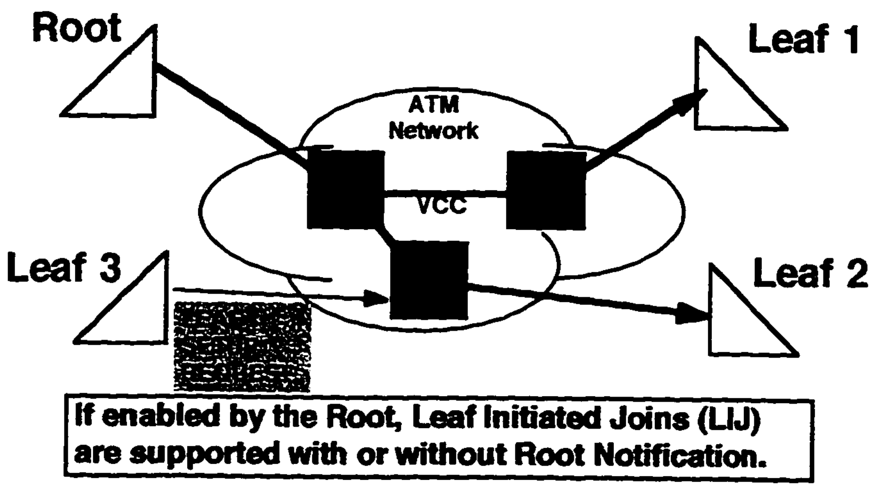


Figure 3-11 : Multicast Connection Control LEAF_SETUP

In response to the ADD_PARTY, the network determines the point on the original connection to branch from and establishes the branch to Leaf 2, with that users acceptance of the connection request. The Root is not the only user that can initiate another branch on the connection. The figure above shows Leaf 3 issuing a LEAF_SETUP_REQUEST to be added to the existing connection tree. The network can either add the branch to Leaf 3 with or without involving the Root with accepting the request, depending on options chosen by the Root in it's initial SETUP request.

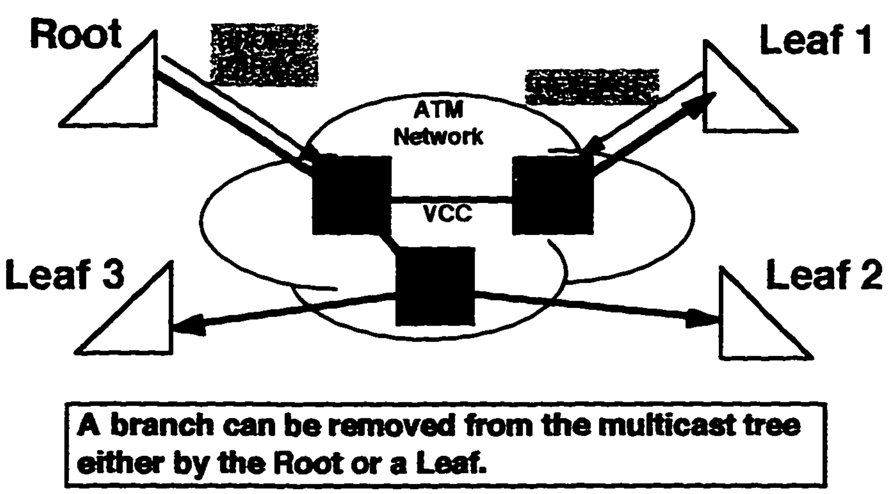


Figure 3-12 : Multicast Connection Control RELEASE/DROP

At any point in the existence of a multipoint connection, participating recipients (Leafs) can also be dropped from the existing at the initiative of either the Root (DROP_PARTY) or the Leaf (RELEASE) as shown above. The remainder of the connection tree remains intact after the branch to the particular leaf is removed.

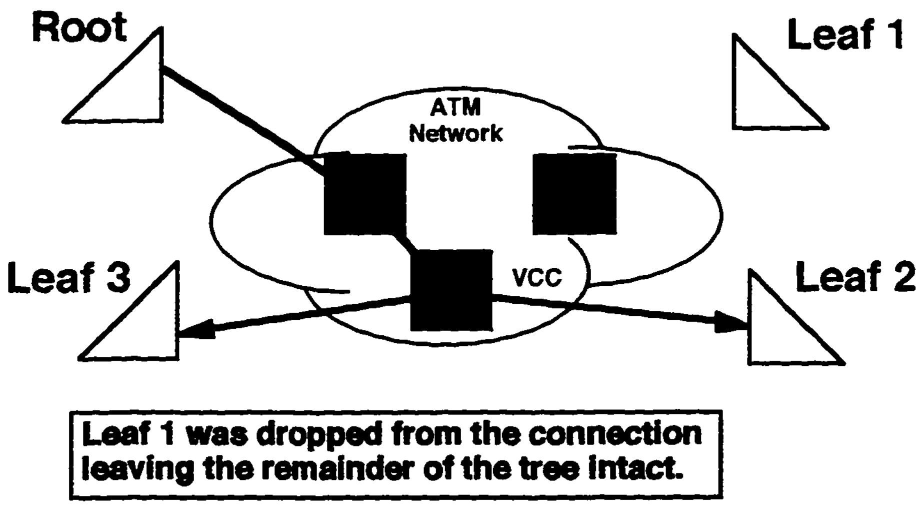


Figure 3-13 : Multicast Connection Control - Final State

To achieve full multipoint connectivity using point-to-multipoint connections, a mesh of N unidirectional spanning tree connections can be used.

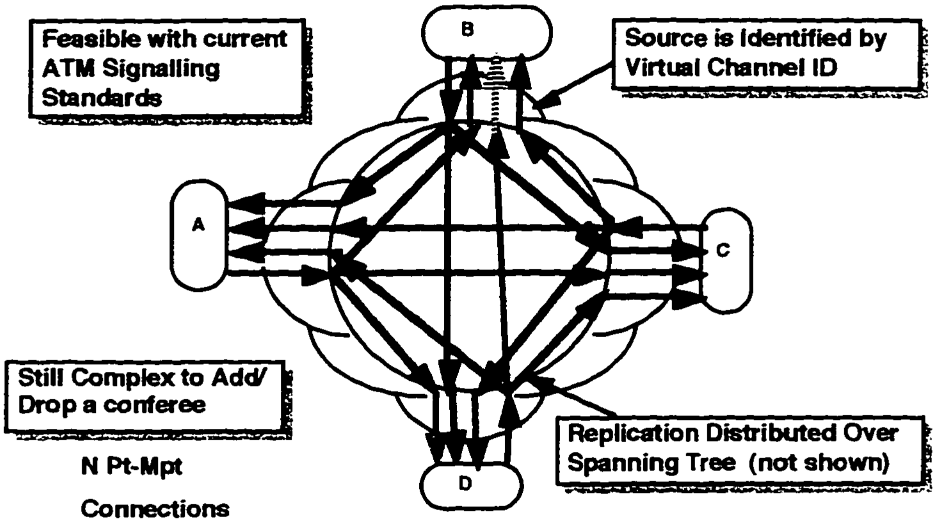


Figure 3-14 : Multipoint Connectivity Using Point-to-Multipoint Connections

Compared with a full mesh of pt-pt connections, using point-to-multipoint connections reduces the load on both the source nodes, that need only transmit a single stream, and the network, that can replicate a stream as close as possible to the recipient and ensure only a single stream on any given link. Pt-Mpt thus inherently improves delay performance and minimizes network resource loading.

Fully mesh connectivity of this kind does not ensure unlimited scalability. As the number of participants in the multipoint service increases the load for recipients increases linearly (N-1) streams to receive). Furthermore, when ever a conferee wishes to join or leave an ensuing conference, every other terminal must interact with it, signalling to SETUP or TearDown connections. Scalability of a multipoint application is ensured by not demanding this full mesh connectivity. Recipient terminal leafs must be able to control how much they receive and from whom. Leaf initiated join and drop facilitates this.

Illustrated following a physical topology through a network of switches, rather than the functional model above, the same full connectivity is shown below.

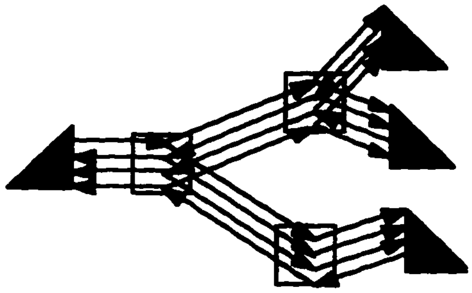


Figure 3-15 : Four Multicast Connections in a Four Terminal Conference

Establishing a spanning tree connection is significantly more complex than a pt-pt unicast connection from a routing & signalling perspective. A link-state type routing algorithm could construct a multicast tree based on a knowledge and comparison of delays and capacities available on various alternate paths available. Adding another terminal to the multicast group involves modifying all existing connections from A, B, C & D to add another branch to E, plus establishing an additional pt-mpt tree from E. All of this should be performed without interruption of service. The ATM Forum's P-NNI (Private NNI) specification is the most advanced standard on how to accomplish such routing decisions on an ATM network.

Connection performance & fault management is also quite complex. Each multicast connection will have a merge connection in the reverse direction for OA&M cells. This could facilitate fault isolation and performance testing via "ping" cells. However, this is clearly much more complex than in a pt-pt connection.

3.3.3. Multipoint-to-Point (Merge)

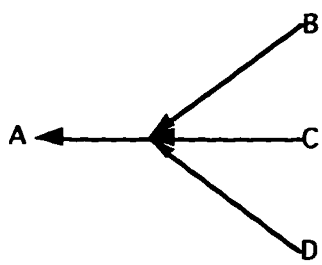


Figure 3-16 : Merge Connection

The need for Merge (Multipoint-to-Point) connections is less obvious but exist. A Merge connection minimizes the number of active receive Virtual Channel/Path connections the receiving Root needs to maintain; if point-to-point connections are used, a receiver in a group of N terminals requires N-1 connections to receive from the others in the group. This requires interaction with the Root whenever a terminal joins or leaves the group. Both these interactions and the number of connections to manage constrain the scalability of using Point-to-Point connections.

A reliable Multipoint Transport Protocol (MTP) is an example of an "application" which can use Merge connections. The function of an MTP is primarily reliable distribution of "data" from one root source to numerous leaf destinations via a Multicast connection. Destinations may rarely communicate with the source to communicate exception events (i.e.: request retransmission). A Merge connection for this sporadic return-path communication minimizes the number of active VCI connections the source root would need to maintain.

Another application requiring a merge connection is B-ISDN OA&M cell flows on the return path of a multicast connection. These management cells are used for connection tracing & fault isolation, loopback tests, etc.

A key design issue for using Merge connections is that each cell requires a unique identifier if either the source must be recognized or the higher layer Protocol Data Unit (packet/frame) to which the cell belongs must be identified for reassembly in an ATM Adaptation Layer (AAL). This means that either frames of length less than 2 cells must be used or an AAL such as AAL-3/4 must be used. AAL-5 is the most widely supported AAL available on most ATM NICs (Network Interface Cards). Unfortunately, AAL-5, formerly known as SEAL (Simple Efficient Adaptation Layer), unlike AAL-3/4 does not provide per multiplexing. This is a strong disincentive to considering use of Merge type connections for multipoint applications.

Switched (on demand through the C-Plane) Merge connections are not currently supported in B-ISDN signalling standards. Provisioned Merge connections are currently feasible on ATM switch products, and are, with point-to-multipoint, the basis for Multipoint-to-Multipoint connections.

3.3.4. Multipoint-to-Multipoint (Broadcast Bus)

A Multipoint-to-Multipoint (Mpt-Mpt) connection provides each terminal in the connection with a single duplex connection through which they can either multicast to the entire group or receive from anyone else in the group. This is essentially a higher level of abstraction for the connection provided to the user, while underlying it is a mesh of Multicast & Merge connections. Logically the connection behaves like a shared medium or broadcast bus. Any terminal can send at any time unless higher layer session controls prevent this.

A functional schematic of data flows in a multipoint to multipoint connection is shown below. Transmitted cell streams are Multicast to Merge points where they are interleaved as they asynchronously arrive.

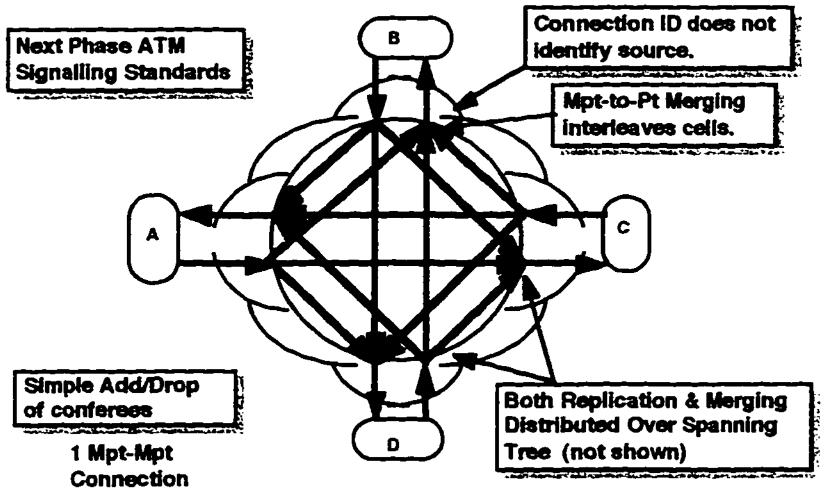


Figure 3-17 : Multipoint Connectivity Using a Multipoint-to-Multipoint Connection

The figure below illustrates a physical spanning tree topology for the multipoint-to-multipoint connection shown functionally above. A conferee can be added or removed from the spanning tree without interactions with any other conferees (interaction with the Root is optional). This is a significant simplification for the terminals over using point-to-multipoint connections.

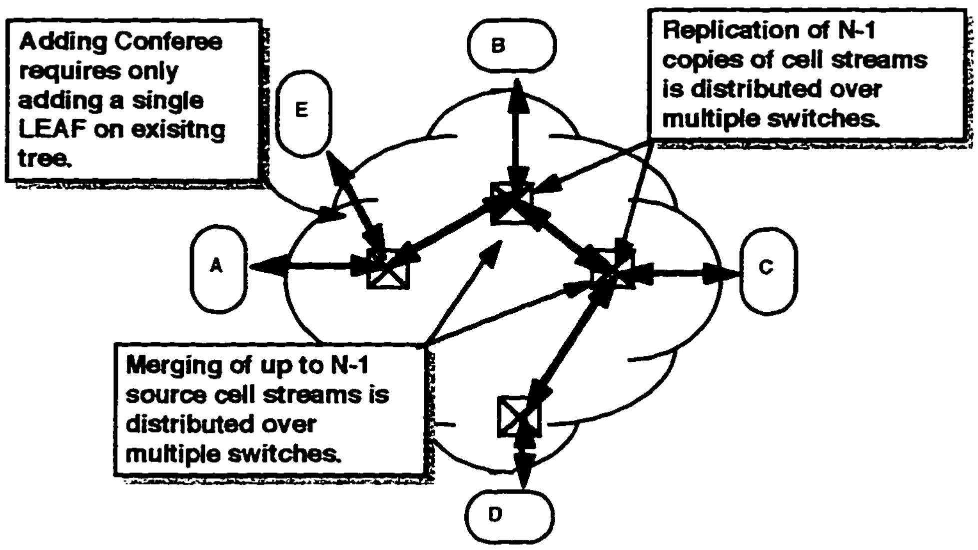


Figure 3-18 : Multipoint-to-Multipoint Spanning Tree

Each switch point the Mpt-Mpt tree above has relatively complex merge and multicast connection mapping to do internally, as shown below.

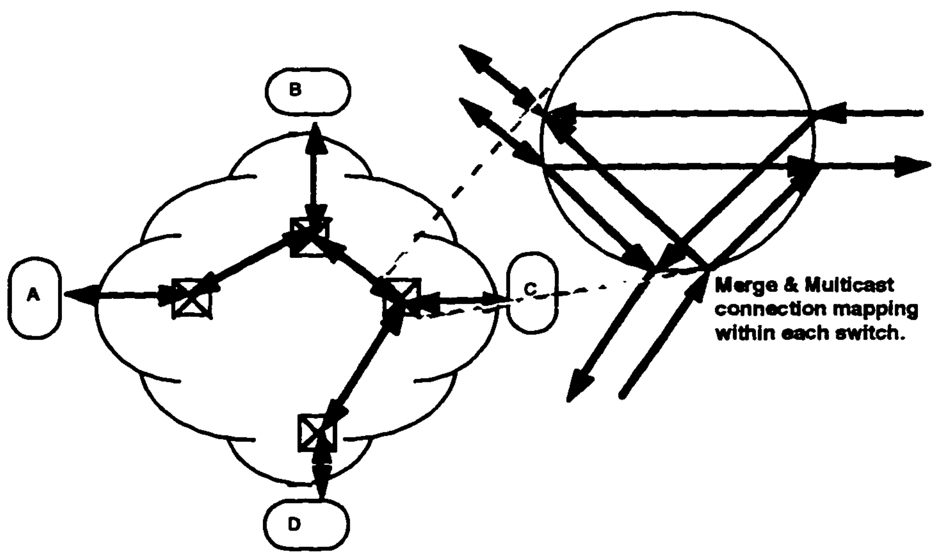


Figure 3-19 : Distributed Multipoint-to-Multipoint Switch Points

To support multiple media and control flows, a set of M Bidirectional Multipoint connections can be used.

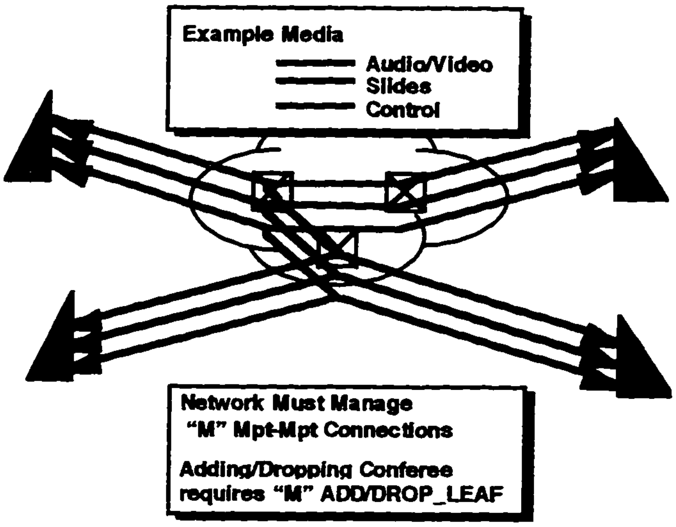
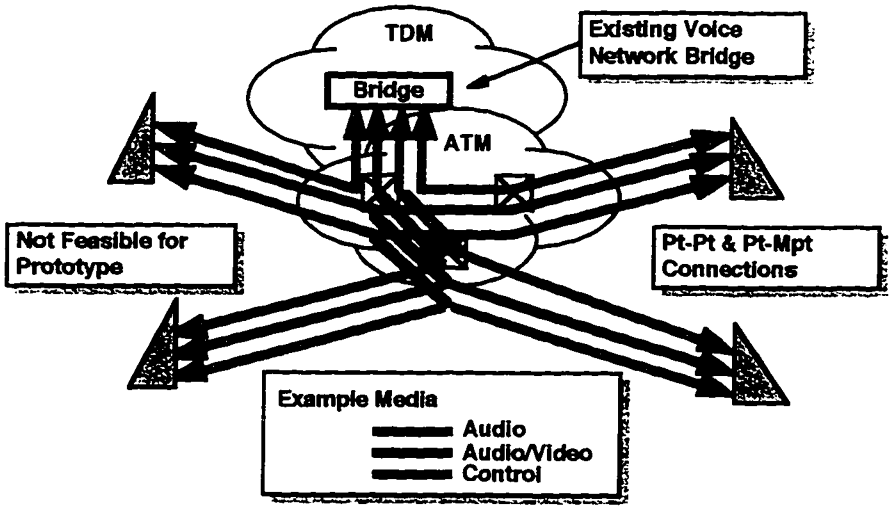


Figure 3-20 : Multiple Connections for Multiple Information Media

To enable continuous presence audio bridging (being able to hear all conferees simultaneously) without undue terminal loading a hybrid architecture that uses point-to-point connections to a network based bridge could be envisaged as shown below



Multipoint-to-multipoint (AKA. bidirectional multipoint) connections have Merge points which result in Cell interleaving if multiple sources are transmitting concurrently.

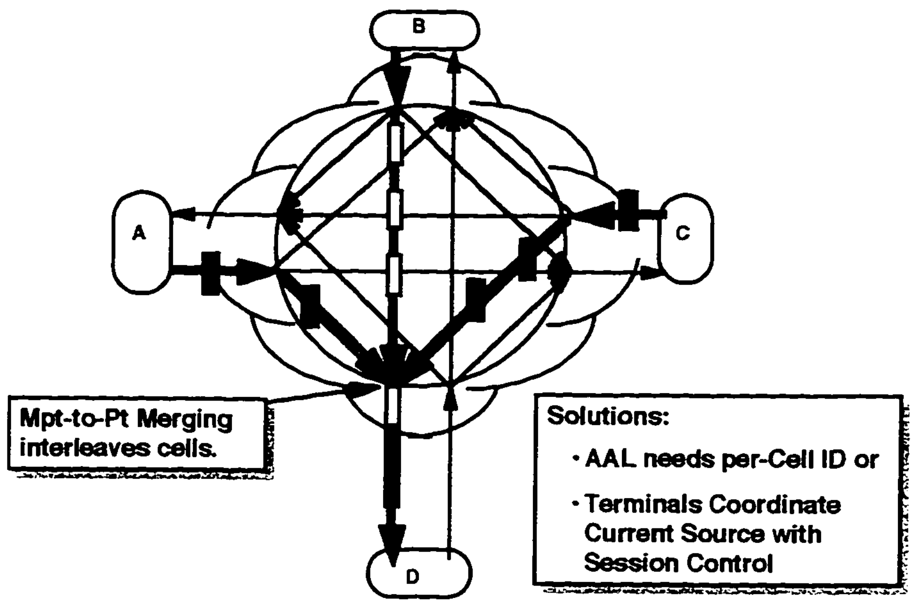


Figure 3-21 : Cell Interleaving: Merge Direction on Mpt-Mpt

This is disastrous for the most popular packet data ATM Adaptation Layer, AAL-5. AAL-5 has no frame identifier carried in each cell. Cell interleaving of AAL-5 frame results in certain corruption and failure to reassemble the packet. Hence, if AAL-5 is used, a single terminal must be assured mutually exclusive privilege of transmitting. This can be solved by either:

- running a distributed session protocol to coordinate mutually exclusive transmit access to a connection or
- the use of Multipoint-to-Multipoint Virtual Paths, each which contain multiple Virtual Channels.

A Bidirectional Multipoint VP (BMVP) can be used in conjunction with source terminal specific VCI to support multipoint communications. This technique reduces the complexity of the connections to be established and managed by the network. The network provides a BMVP service only, while the terminals can manage source discrimination by VCI allocation.

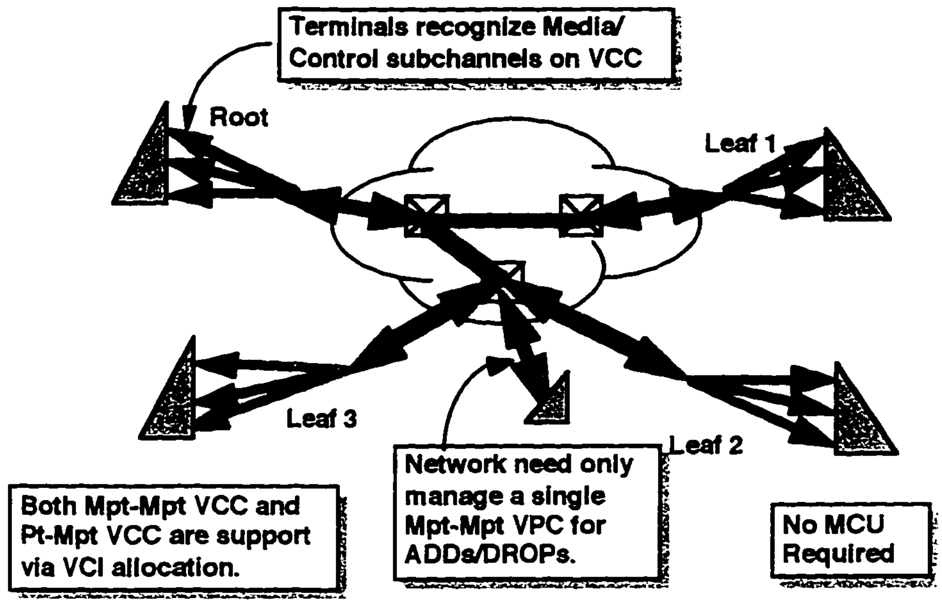


Figure 3-22 : Mpt-Mpt Virtual Path for MMC

If all terminals can transmit on the same Virtual Channel within the Virtual Path, the virtual channel is a multipoint-to-multipoint one. If unique Virtual Channel Identifiers (VCI) are allocated to each source then they effectively become point-to-multipoint connections. Either can be utilized as appropriate, given the nature of the media specific AAL being used.

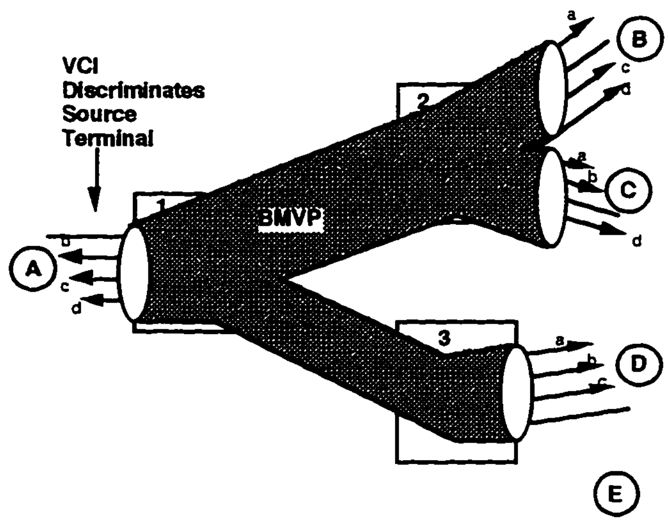


Figure 3-23 : Bidirectional Multipoint VP Service with VCI Used to Discriminate Source

4. Decentralized Multipoint Conferencing Proposal

4.1. Motivation for Developing ATM Conferencing

The past chapters have argued that there is a significant pent up demand for Video Conferencing services: conferencing is useful for personal video-telephony, business meetings in increasingly multinational partnering world, an aid to increasingly popular telecommuting, distance education to new markets in small remote locations, presentations to large audiences by executives or subject matter expert conferees, telephony for the signing deaf, entertainment video on demand, etc. Despite all these potential applications, video conferencing is not widely used today. Simply said, it costs too much and is difficult to set up, maintain and use.

In the world of computing and telecommunications, video conferencing is still a relatively small market. However, market researchers predictions do correlate at least in optimism for dramatic growth, driven largely by videophone and desktop computer video conferencing products expected to hit the market in 1997 based on 1996 ITU-T standards.

Aside from the POTS based videophones however, the network infrastructure is not yet in place to support ubiquitous on demand use of multipoint conferencing. The penetration of ISDN is still relatively low and the Internet does not yet have the capacity nor the capabilities for adequately supporting interactive conversational use of audio and video. Typical 10 Mbps LANs can only support a few simultaneous video calls and, even if they could support more, the typical Enterprise network wide area leased lines can not. Furthermore, all of the ITU-T standard systems are constrained to require expensive service specific network equipment, MCUs, to support multipoint conferencing. The requirement for MCUs increases costs, adversely impacts performance and feature functionality, and slows deployment of multipoint conferencing equipment and services to keep pace with the deployment of service specific logical overlay equipment. This historic constraint has its origin in the inherent limitation of switched TDM networks supporting only point to point connections.

Public and private networks are undergoing revolutionary change today with the maturation of broadband switching and transmission equipment and its increasing deployment and feature richness. Plans are underway to evolve the Internet to a true integrated services network that will deterministicaly support audio and video QoS and bandwidth requirements and dynamic topology multipoint connections.

The future availability of a switched wide-area broadband transport infrastructure and continuing deregulation of telecommunications worldwide provides an environment for an explosion in competitive service providers. The availability of multipoint connections in particular permits a dramatic rethinking of the architecture of a multipoint video conferencing system.

This chapter describes a novel architecture for a multicast packet network based multipoint video conferencing system . The architecture is built upon a packet network which supports:

- multicast connections
- wideband data rates (>384 kbps)
- deterministic QoS
- geographic wide areas

The design herein specifically uses ATM networking, but when IP based networking (e.g. Internet) evolves to support the same features it could be used instead of ATM.

4.2. Design Goals and Implications

The following subsections summarize the key design goals and their ramifications on system design.

4.2.1. Eliminate Video Conferencing Specific Network Equipment

Eliminating Multipoint Control Units (MCUs) and Channel Aggregation Inverse Multiplexors (I-MUXs) from the video conferencing system will reduce equipment costs, ease service deployment, and reduce performance and functionality constraints imposed by this equipment.

I-MUXs are eliminated by using ATM's inherent capability to support channels of various bandwidths.

The key requirement for the network to facilitate the elimination of MCUs is that multipoint connections can be configured dynamically by terminals. This can be accomplished today on ATM networks supporting ATM Forum UNIv3+ signalling. The network-based generic multicast capability then replaces the MCUs function of replicating video streams to be seen by multiple recipients.

Since the MCU performs audio mixing as well, this function needs to be moved either to the terminals or to network based audio conferencing equipment such as that widely used for voice conferencing today.

4.2.2. Maximize Audio/Video Quality While Minimizing Delay

Compression technology has advanced tremendously over the past 10 years to the point where it is feasible to have usable H.320 (or proprietary) video conferencing using only two 64 kbps channels. These current systems are severely limited in their quality however, even at 384 kbps.

The high levels of compression used in these systems come at costs in CODEC complexity and audio/video quality and delay. To reduce bandwidth demands, the spatial resolution and frame rate of video is kept low. Audio is compressed to 8 or 16 kbps. The result is audio and video that are intelligible for conversation but not pleasing (natural). Furthermore, the service application of the system is constrained. For example, full motion entertainment video, videophones for the signing deaf or high fidelity stereo music can not be supported on such systems. One of the most notable impacts is end-to-end delay; temporal compression of video involves buffering of multiple video frames at both the source and (less so) the receiver. Each frame buffered introduces another frame time of delay (e.g. 33mS at 30 fps or 66mS at 15 fps). This coding delay is the most significant element of end-to-end delay that makes conversations unnatural; users adapt to the system via chairperson controls and social protocols for waiting for a turn to speak. Video conferencing today is not very much like conferees being in the same room.

For conversation applications like conferencing, imperceptible end-to-end delays are imperative for natural conversational interaction between conferees. One way delay must not exceed the order of 100 milliseconds. Delay is less of an issue for broadcast applications.

ATM eliminates these constraints by simply providing higher bandwidth. This allows using simply spatial video compression, and hence potentially less that 1 frame time of delay for compression (depending on the power of the CODEC). Hi-fi audio can also be supported via superior quality coding.

Hardware based CODECs, simple protocol stacks with large packet sizes to minimize processing load in terminals, and synchronization mechanisms that discard delayed or corrupted AV samples are also implied to meet delay constraints when using a general purpose (i.e.: non-real-time) computing platform for a terminal.

The elimination of the MCUs and I-MUXs from the network, and the use of multicast connections also minimize network introduced delay.

4.2.3. Enhance Multipoint Functionality Beyond Switched Presence

Multipoint conferencing systems today typically provide a *switched presence* service. This means that all participants see a single current speaker (except the current speaker who sees the previous speaker). Under chairperson or voice activated controls, the MCU selects the current speaker and multicasts their video stream to all sites.

Continuous presence systems also do exist where up to 4 remote participants can be seen continuously. 4 QCIF images are mapped into 4 quadrants of a full CIF image and this image is sent by the MCU. However, such systems are expensive and introduce significant delay due the need to decode and multiplex video in the MCU prior to re-encoding for transmission.

By using ATM multicast SVCs, a more versatile conferencing system can be achieved. Users can individually control who they see on their screen. This can be 1 or more remote speakers up to the capacity of the terminal to process multiple incoming video streams.

Audio must be "continuous presence" in a general purpose conference system; all participants should be able to be heard. Since the terminals will not necessarily be able to process all of the video streams sent by remote conferees, audio must be transmitted separately from video to allow continuous presence audio.

4.2.4. Tolerate Network Losses

Using a packet network (ATM) as opposed to a TDM network implies that congestion and losses can occur. This can be eliminated by an over provisioned network but the general traffic management philosophy popular in the telecom industry is to maximize network utilization at the cost of some probability of cell loss. In a delay sensitive conversational service there is no time for retransmissions to ensure a lossless service.

Hence, both the audio and video receiver must be tolerant to lost information. For example a lost video frame can be ignored and the next one received displayed.

4.2.5. Ensure Audio/Video Lip Synchronization

Since audio and video are carried on separate connections, it is the terminals responsibility to ensure they are synchronized for display. This involves delaying the media stream that arrives "early". A simple mechanism for this is the use of time stamps in audio and video data.

Lip-synchronization errors appear to be unobjectionable for video-to-voice lags if they are in the range of -90 to +120 milliseconds.

4.2.6. Secure Access Control

A basic security requirement for a video conference is ensure that only welcome participants have access to the conference. This type of functionality exists in MCUs in the form of passwords demanded on connection.

A *Registrar* server can be used to register participants for a conference. Users connect to this Registrar and provide information to validate their identity before being provided with information on conference participants and the addresses of their offered media streams. This is one centralized function of the MCU that is not readily distributed.

4.2.7. Encoding Independence and Extensibility

The conference system should not be locked into using specific media encoding algorithms or formats, nor should it be limited in the media types supported. Video compression/decompression algorithms for example should be implemented in a modular fashion such that they are independent of other components in the system (such as the network), and easily exchanged for new algorithms in the future. Additionally, the application should be structured so that new services such as VoD or collaborative work environments become simple extensions to the existing application structure. These new services may require new network functionality such as a guaranteed multipoint transport protocol, which should be implemented with little difficulty. This enables the system to be extensible as new technologies emerge.

4.3. System Architecture

4.3.1. Terminal Architecture

The proposed system was designed to include audio/video functionality along with a slide presentation tool, all utilizing native ATM technology such as point-to-multipoint (multicast) connections and switched virtual circuits (SVCs). The functional components of the terminal are shown in the following figure.

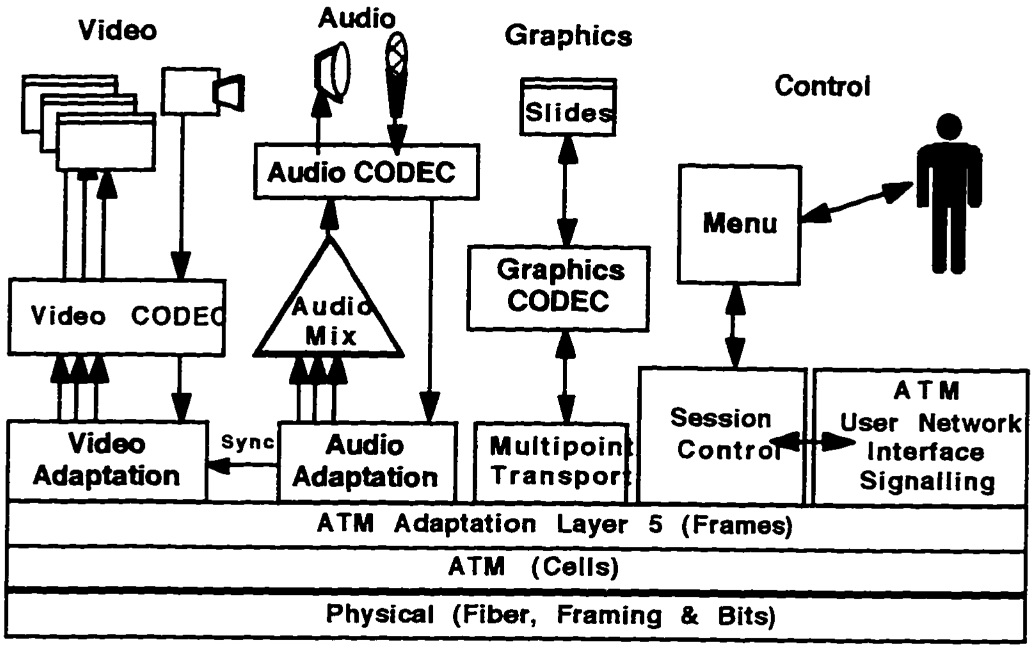


Figure 4-1 : Terminal Architecture

Each vertical stack operates autonomously from the others in the terminal (i.e.. they are separate concurrent processes). They all rely on AAL-5 adaptation and ATM UNI Signalling. The system could be extended to have additional subservices (beyond audio, video and graphics) the addition of other vertical stacks.

4.3.1.1. Graphical User Interface

The Graphical User Interface (GUI) provides the interface between the end-user and all of conference system features. In essence, it is the heart of the system since it takes on the role of command and control center. Its functionality enables a user to:

- Register to the local environment
- Apply to the Registrar for membership within a particular conference
- Obtain a dynamic view of the conferees within a conference
- Selectively receive media streams being offered by other conferees
- Selectively advertise and offer media streams to other conferees
- Leave a conference

Based on received configuration and state information from the Registrar, as well as conferee selections within the GUI, control messages are sent to the appropriate components. For example, assume that the Registrar has conveyed the existence of audio and video streams being offered by conferee xyz to all other conferees. The local conferee may choose to receive these two streams and may indicate this via the GUI. The GUI would in turn issue a control message to the Video Display component of the form open(xyz,video), and another control message to the Audio Mixer component of the form open(xyz,audio). These two components would then take it upon themselves to establish the appropriate ATM connections to enable the receipt of these media streams. Upon reception of the streams, these components would also be responsible for properly displaying these streams in a synchronized fashion. In a similar fashion, the GUI controls the function of all other components.

4.3.1.2. Session Control

Session Controller provides bi-directional information exchange between the Registrar and the terminal GUI. The Session Controller can queue incoming messages from the Registrar and communicate them to the GUI using IPC when available.

4.3.1.3. Video Capture

The Video Capture component's main function is to code (digitize and compress) a live video signal from a camera or other video device and distribute it across the network to any conferees who request it. The coding process uses Motion JPEG for compression/decompression of video streams. Once this component is launched by the GUI, it is able to work relatively autonomously requiring no control from the GUI. The component has the added responsibility of monitoring the network for incoming LIJ requests from other conferees who wish to receive this video stream. It then assists in negotiating the required ATM connection. Video data units are time-stamped and numbered in order to facilitate synchronization with associated audio upon playback by remote conferees.

4.3.1.4. Audio Capture

The Audio Capture component's operation is similar to that of the Video Capture component. Its function is to capture and digitize a live audio signal from a microphone or other audio source and distribute it across the network to any conferees who request it. Once this component is launched by the GUI, it is able to work relatively autonomously requiring no control from the GUI. The component has the added responsibility of monitoring the network for incoming LIJ requests from other conferees who wish to receive this audio stream. It then assists in negotiating the required ATM connection. Audio data units are time-stamped and numbered in order to facilitate synchronization upon playback by remote conferees.

4.3.1.5. Video Display

The function of the Video Display component is to playback a video stream to the local display in a synchronized fashion as the data is received from the ATM network. As with other components, this one is launched and configured appropriately by the GUI. The component must first submit an ATM LIJ connection request to establish a connection between itself and the conferee's end-station which is the source of the video stream. It must then begin receiving the video stream and display it in a synchronized fashion. Assuming that an audio stream is currently being received which is associated with the video stream, the Video Display component must communicate with the Audio Mixer component to achieve inter-stream synchronization. The GUI will launch a separate process/thread to handle each individual video stream being received by the conferee. Therefore, there can be multiple Video Display components communicating with the Audio Mixer Component at any one time.

4.3.1.6. Audio Mixer

The function of the Audio Mixer component is to playback all of the incoming audio streams to the local audio device in a synchronized fashion as the data is received from the ATM network. As with other components, this one is launched and configured appropriately by the GUI. Unlike the Video Display component, only one Audio Mixer component can ever exist. This is due to the fact that audio streams must be mixed together before playback. The GUI sends a message to the Audio Mixer to indicate when new audio streams are to be received and mixed or removed. The Audio Mixer is then responsible for submitting an ATM LIJ request to negotiate the required connections with the audio stream sources and subsequently add these streams to the pool of audio streams to be mixed. Synchronization is performed based on the received time-stamps and sequence numbers associated with media data units.

4.3.2. Operation Overview

Connectivity among ATM conferees to enable the receipt of media streams requires a conferee's membership within a native ATM multicast connection. An ATM multicast connection, as defined under the ATM Forum's User Network Interface (UNI) standards documentation versions 3.1 and later, employ the tree analogy to produce the concept of a point-to-multipoint connection. There exists one *root* which is the source or sender of any data on the multicast connection, and multiple *leaves* which are the receivers or sinks of data on the connection. A multicast connection is owned by the root who not only determines the QoS details of this point-to-multipoint connection, but also determines who is allowed to become a leaf. Additionally, the root may optionally empower the network to accept anybody's request to become a leaf of its point-to-multipoint connection without being notified. To become a leaf, a user must signal their intentions at the UNI. The network is then responsible for accepting or rejecting this request. If the leaf is accepted, the network then becomes responsible for any data replication which must occur. It should be emphasized that multicasting is a native ATM network based service, so the root of the connection is not responsible for routing or duplicating data streams.

An important advantage of native ATM point-to-multipoint connections is that they are dynamic, hence the number of leaves may change throughout the lifetime of the connection. This is extremely useful in a video conference where a switched presence environment is desired and bandwidth may be at a premium. Assume that each participant is multicasting their video stream into the network and hence are each the root of a unique point-to-multipoint connection. Other conferees wishing to see only a subset of these video streams may dynamically add or remove themselves from any of the multicast connections. Conferees may easily and efficiently implement a policy such as "view the video of the current speaker only" or even "view the video of the three most recent speakers", without the need for an expensive and complex conference bridge. This type of service demonstrates the potential value of ATM's native multicast capabilities.

One important component of ATM multicasting is the method in which leaves indicate their desire to be added to a multicast connection. Two methods exist today namely Leaf Initiated Join (LIJ) (implemented in ATM Forum's UNI v4.0) and Root Initiated Join (RIJ) (supported in previous standards such as FORE Systems' SPANS and ATM Forum's UNI v3.1). LIJ involves a leaf signaling their intention directly to their ATM user network interface (UNI) whereas RIJ requires the leaf to use some external out-of-band method of informing the root of their desire and having the root perform the network signaling on their behalf. LIJ is the most useful and functional since it doesn't need to involve the root.

Understanding the role of the various terminal components is made more clear by explaining the operation of a system of multiple terminals operating over a network. An example in the following figures is used to illustrate.

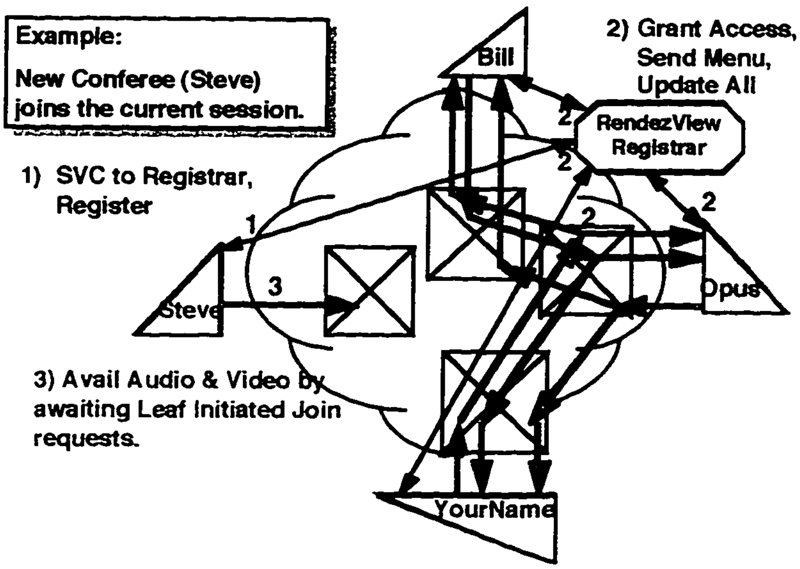


Figure 4-2 : Conference Registration

A global registration server called the *Registrar* implements access control, security and all mechanisms enabling potential conferees to gain access to the virtual conferencing environment. Additionally, the Registrar may be employed repeatedly throughout the lifetime of the conference as new conferees join or others leave. The Registrar enables the guaranteed distribution of state or configuration messages specific to a conferee which would enable others to become globally aware. The message content includes personal information such as a conferee's name, mugshot, email address etc., of a media-specific information such as media types, encoding style, data rates etc., and network information such as end-system addresses and media stream IDs. Irrespective of the actual information being distributed, the Registrar is necessary to guarantee the initial and on-going operation of the conference.

Steve must first connect with the Registrar in order to obtain state information about other users. In step 1, Steve establishes a point-to-point connection to the Registrar located at a globally known address. After completing a security validation, the Registrar returns conference state information to Steve describing the current state of the virtual conference environment (step 2). This state information provides personal details on each of the three conferees (name, address, email address, mugshot etc.) as well as media specific information on each conferee (media stream types offered, encoding specifics, ATM multicast addresses etc.). In step 3, Steve is globally aware of his conference environment and can now issue LIJ requests to the ATM network in order to be added to any of the three conferees' existing multicast connections to receive audio, video or other media. At this point, the network will establish the appropriate connection to add Steve as a leaf. Any information subsequently transmitted by the root onto these multicast connections will be duplicated and sent to Steve as illustrated below.

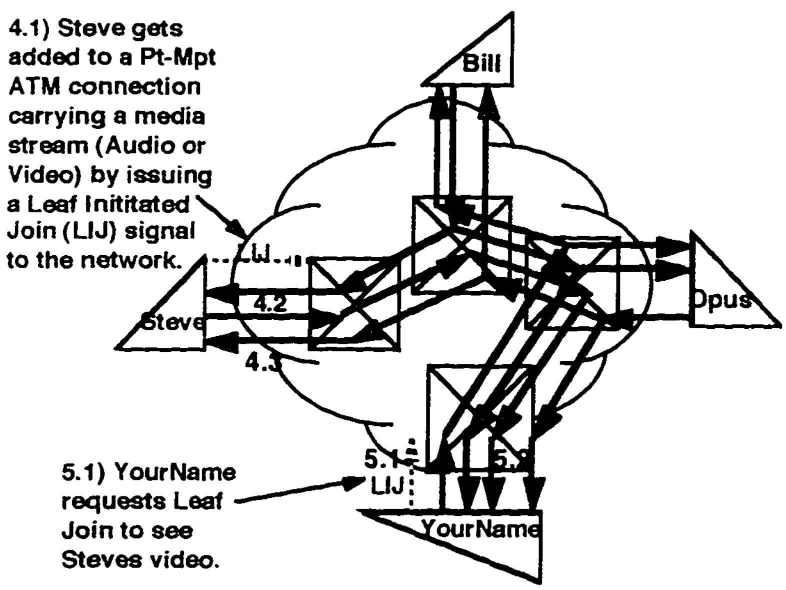


Figure 4-3 : Joining Connections

Once Steve has been granted access to the conference environment and has received complete topological information about the virtual conference environment, he must then issue LIJ (4.1) requests to begin receiving audio or video from other conferees. In the figure, Steve would like to receive Opus' video and so issues an LIJ resulting in being added as a leaf on the connection from Opus (4.3). Similarly, Steve begins to receive Bill's media stream (4.2) and other participants begin to receive from Steve (5.2).

ATM's multicast connections are extremely versatile since they may be altered dynamically while the connection is in use. Leaves may be added or removed on the fly and the network guarantees the QoS contract initially negotiated by the root for every leaf. LIJs are the most efficient and natural connection method. The native point-to-multipoint facility offered by ATM demonstrates the advantages of operating a video conference in a decentralized mode as opposed to a centralized approach where expensive MCU hardware and multiplexors would typically be required.

5. Implementation Reality: ATM RendezView

ATM RendezView is the prototype multipoint video conferencing system built following the architectural principles outlined in the previous chapter. The system was first demonstrated in Nortel Technology's COBRAnet lab in October of 1996. RendezView was designed and built in partnership with Keith Smith.

RendezView demonstrates the most demanding elements of the distributed multipoint conferencing architecture; conference establishment and configuration is completely dynamic via Leaf Initiated Joins to media streams prompted by button presses on the Graphical User Interface (GUI). Audio and video are hi-fidelity, lip-synched, and delivered with low end-to-end delays on the order of 100mS.

The video subsystem receives more attention than the audio in this thesis. More detail on the audio subsystem can be found in [5].

5.1. Development Platform:

ATM RendezView was developed using the C programming language because of its large installed base, versatility and speed. Motif/XWindows was used to produce the GUI and a POSIX.4 compliant library along with Solaris Threads was employed to support real-time functions. The hardware employed was more selectively chosen for the following reasons:

Sun SPARC 20 workstation running Solaris 2.4 OS: This workstation configuration was chosen for several reasons which lend well to the development of real-time multimedia applications. The Sun SPARC's SBus is supported by many third party hardware vendors. Additionally, the SPARC workstation sports some impressive processing and bus bandwidth figures which are inherently required in any multimedia application. Solaris 2.4 has fairly complete support of IEEE's POSIX.4 real-time programming standard. Additionally, it has support for multithreading facilities which provide the equivalent of the POSIX.4a standard. Solaris 2.4 also defines a real-time processing class which enable processes to obtain real-time service and scheduling in an otherwise non-deterministic scheduling environment.

Native Sun Audio CODEC: Because Sun's native audio device is bundled with the computer and is fully-featured, it was essentially a shoe-in. The CODEC possesses both D/A and A/D converters which are able to operate at a multitude of encoding rates and formats ranging from basic μ-law up to CD quality audio. The device contains microphone and line-level inputs as well as headphone and line-level outputs.

Parallax PowerVideo Card: Parallax's PowerVideo Card was chosen because of its on-board Motion JPEG CODEC. It has a video development environment including a complete set of X/Motif API widget libraries which allow for full control of the CODEC hardware thus off-loading any compression/decompression functions from the main CPU. The video hardware is capable of providing the encoding and decoding of a Motion JPEG video stream of 640 x 480 pixels, 24 bit color at frame rates in excess of 25 fps in real-time. As video resolution is decreased, additional video streams may be processed in real-time.

FORE SBA-200 ATM NIC: The ATM RendezView prototype is dependent on many features offered by the ATM Forum's UNI v4.0 standard. These include the negotiation of QoS parameters on a per-connection basis as well as the dynamic establishment and modification of point-to-multipoint ATM connections using SVCs. The advantage of employing FORE Systems ATM NIC was that it offered an API library interface to much of this functionality via its proprietary SPANS signaling protocol. This signaling functionality is instrumental to the success of the ATM RendezView prototype.

5.2. Network View of Operation

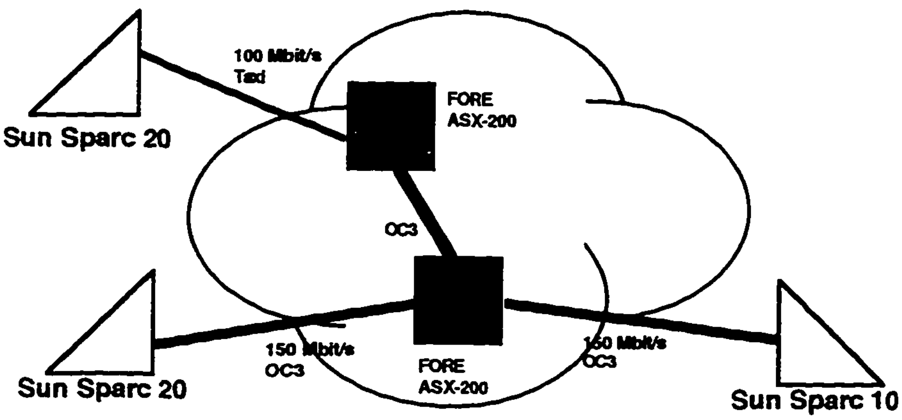


Figure 5-1 : Network for ATM RendezView Prototype

ATM RendezView utilizes Sun Sparcstation based terminals interconnected by an ATM network consisting of two FORE Systems ASX-200 switches. The workstations are running a Solaris 2.4 operating system. RendezView applications software was written in C and utilizes X-Windows for its GUI.

The design objective with this distributed multipoint conferencing system was to use UNIv4.0 Leaf Initiated Join (LIJ). Conferees obtain a menu of other participants and select which media streams they wish to receive.

UNIv4.0 signalling was not supported in commercial products at the time of RendezView's implementation. Hence, FORE's "SPANS" proprietary signalling protocol was chosen because it appeared to have equivalent functionality to LIJ with a concept FORE called Groups. Joining a SPANS Group was believed to force automatic leaf branch setup, and tear down of the branch when a terminal left the Group. It turned out however, that the Group Addressing concept did not operate in this fashion.

In order to accomplish the goal of dynamic control by the receiver, an extra control path was added to send Leaf Join Requests (LJR) to the media source (Root). The Root then signals the network to add the leaf branch to the requester. This two step procedure is illustrated by an example in the following figures.

Any newly arriving participant to the conference awaits Leaf Join Requests (LJR) before transmitting Audio or Video into the network.

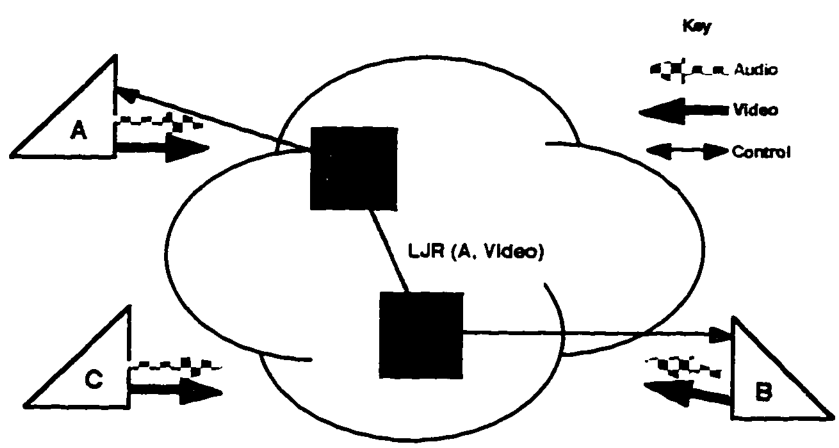


Figure 5-2 : Signalling Leaf Join Requests

Any terminal wishing to received Audio or Video issues Leaf Join Request to the source Root. the LJR includes an indication of which media stream is requested and the ATM address of the requester. LJRs are sent by a UNIX socket which uses a point to point TCP/IP over ATM connection to the Root. In the figure above, terminal A sends a LJR to B, requesting B's video.

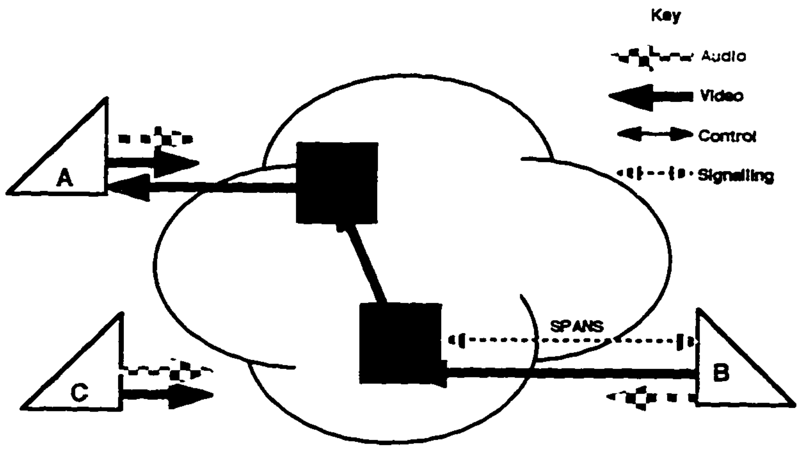


Figure 5-3 : Root Initiated Join Response to LJR

The Root (B) responds to the LJR (from A) by issuing an `atm_connect` SPANS signal to the network. The network then establishes the branch on the multicast connection which appears initially as a point-to-point connection as illustrated above.

The above procedure is repeated under receiver control to start or stop receiving any media streams offered by remote participants. RendezView was demonstrated with three terminals.

5.3. RendezView Terminal Components

ATM RendezView is an implementation of the audio and video media subsystems and dynamic connection control subset of a generic multipoint multimedia conferencing system. Features such as whiteboards, pointers, and transfer of presentation slides could be readily added since the architecture is based on autonomous parallel processes for each media stream that is carried across a dedicated ATM Virtual Channel Connection (VCC). Processes are represented by parallelograms in Figure 5-4. Instead of using a network Registrar server, apriori knowledge of the other participants is provided to each terminal for the RendezView prototype.

The video CODEC utilized in RendezView is a Parallax Power Video card, which facilitates Motion JPEG (M-JPEG) video through an X-Windows/Motif Application Programming Interface (API). Offloading video compression and decompression to hardware support permits simultaneous transmission of one video stream and reception of multiple video streams. A default mode of 320x240 pixels at 15 frames per second is used in RendezView

The Audio CODEC used is the defacto standard Sun Audio card. Linear PCM coding is used to minimize process for coding and mixing of audio streams.

A FORE systems ATM Network Interface card support AAL-5, ATM and signalling functions through a socket-like file descriptor programming interface.

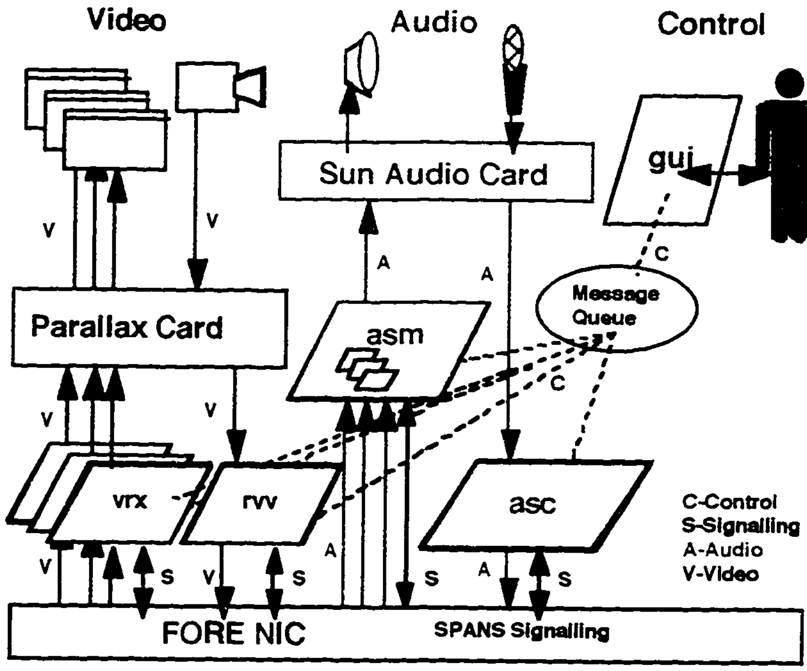


Figure 5-4 : Terminal Architecture for ATM RendezView Implementation

Each parallelogram in Figure 5-4 represents a separate process in the ATM RendezView program. Three hardware assist cards are used shown as rectangles: A Parallax motion JPEG CODEC, the standard Sun Audio card and a FORE System's ATM Network Interface Card.

Briefly, the processes that constitute RendezView are:

GUI: The Graphical User Interfaces provides the user with buttons to turn on/off the transmission of audio and/or video, and selection of the audio/video streams from remote conferees to be received. The GUI process launches the necessary processes and controls them via Unix Messages sent via a message queue.

RVV The RendezView Video transmitter periodically prompts the Parallax Card to compress a still image and then it transmits this image across the network using the FORE NIC. RVV also provides a video configuration GUI to set video parameters including frame rate, spatial resolution, compression factor, colour balance, brightness etc.

VRX The Video Receiver process reads a stream of motion JPEG images (one at a time) from the FORE NIC and then prompts the Parallax Card to decompress and display them in a window. A separate instance of VRX is created for each video stream being received.

ASC The Audio System Capture process reads PCM audio samples from the Sun Audio card, packetizes them, and transmits them via the FORE NIC.

ASM The Audio System Multiplexor process receives one or more audio streams from the FORE NIC and mixes them so that all can be heard simultaneously. The singular ASM process utilizes separate Threads for each incoming stream

5.4. Process Scheduling Philosophy

Unix is not a real time operating system; it is not possible to deterministically reserve processing resources for a given process. Control of resource allocation is limited to relatively crude priority mechanisms that determine which processes will get the worst level of service if and when the system is overloaded.

Audio quality is widely agreed to be paramount in conferencing systems. Low delay, high fidelity, continuous presence audio is the basic requirement for effective and intelligible conversations.

Hence the audio processes were given the highest priority available in the Solaris operating system, so-called Real Time Class. Furthermore, the ASM (audio mixer) process utilized Threads to ensure maximal efficiency (minimal system calls) processing multiple incoming audio streams. More details about the audio system can be found in [5].

Since the Real Time Class is reserved for audio in RendezView, the video subsystem was designed with the ability to adapt to diminished processing resources as the number of conference participants grows. The result is that when the system gets overloaded, video frames are dropped. However lip synchronism with the audio is maintained and subjectively the reduced video frame rate is not unpleasant (at least not as unpleasant as audio clipping or silence).

5.5. Video Subsystem

Motion JPEG is used for video compression in RendezView. Motion JPEG was chosen primarily due to the availability of affordable hardware at the time of RendezView's implementation. A hardware CODEC card off-loads the processing of full-motion video from the CPU, otherwise it would be impossible to compress and decompress multiple video streams simultaneously, even on a high-powered Sparcstation. This ability is imperative for continuous presence multipoint conferencing.

A Parallax Power Video card was used in ATM RendezView. The Parallax card becomes the driver for the computer's monitor so that it can directly insert video into displayed windows. Parallax specifications state that the card can digitize, display and compress a single video at 30 frames per second with a 640x480 pixel resolution and 24bit colour depth. The unknown entering into development of RendezView is how that processing power would divide to support multiple lower fidelity video streams, compressing one and decompressing several.

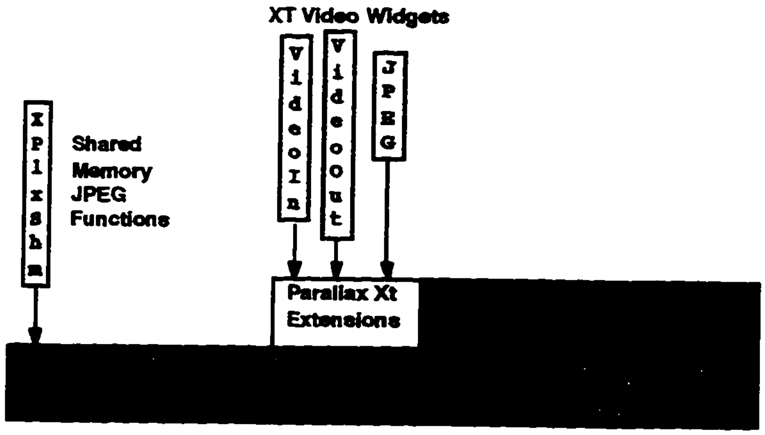


Figure 5-5 : Parallax and X-Windows

Parallax uses X-Windows/Motif as the Application Programming Interface for their Sun cards. The video digitizer is viewed programatically as an Xt level widget called VideoIn. The compressor/decompressor is viewed as the JPEG widget. The widget for outputting video on an NTSC interface (VideoOut) was not used in RendezView.

The RendezView video transmission subsystem is structured as shown below. Everything within the large parallelogram is within the RVV process.

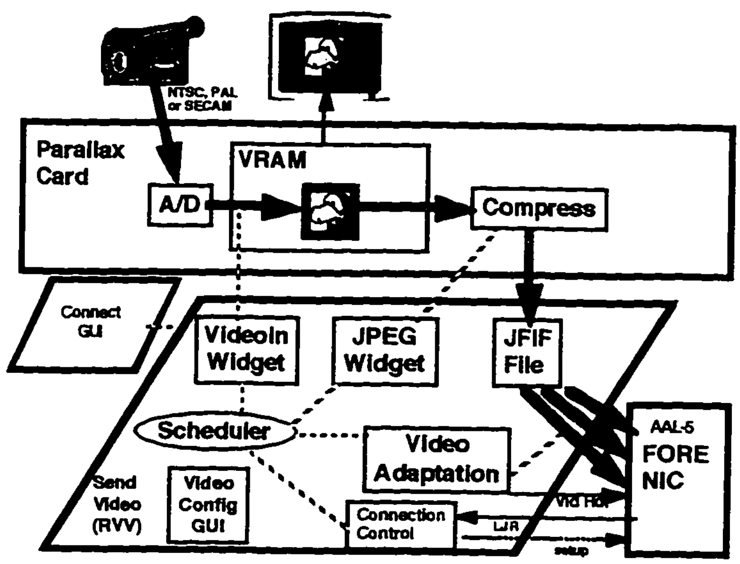


Figure 5-6 : Video Transmission Subsystem

The Connection GUI is the primary GUI for the RendezView application. When the user presses "Send Video", the RVV process is launched and it awaits Leaf Join Requests (LJR) from distant conferees.

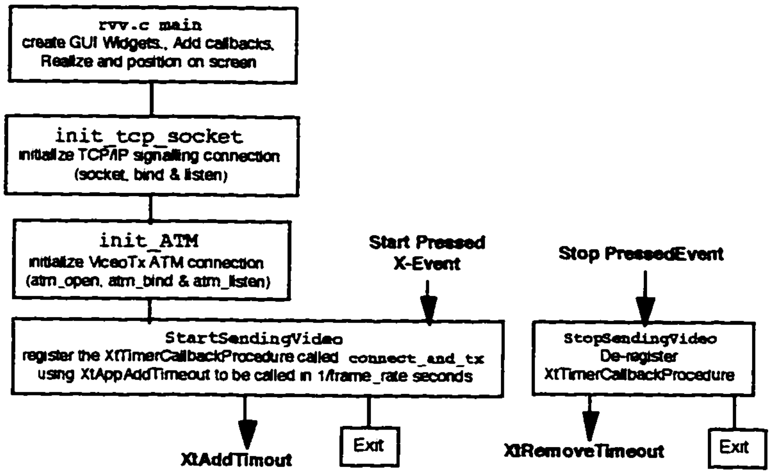


Figure 5-7 : Flow of rvv.c Mainline

At any time, the user may interact with the RVV process via the Video Config GUI to set frame rate (Scheduler), resolution and compression factor (JPEG Widget) and various colour and aesthetic options for the video(VideoInWidget).

Scheduling is done by an Xt timing facility called an XtTimerCallbackProcedure. A call is made to Xt requesting a one-shot callback to a named procedure. The time period requested is one video frame time (default is 66mS for 15 frames per second). At best, this is when the procedure will be called. However, if another process is running at that time or the X-system is processing another event, this procedure will be called less often than 15 frames per second. This makes it important that the programmer ensures that a button press does not lock up the X system with a long or blocking procedure. Again, the audio processes have the highest priority class to ensure they have enough resources. Video frames can be dropped occasionally or delayed a small fraction of the frame time delayed without perceptible subjective quality degradation. The first thing the XtTimerCallbackProcedure does before processing the next frame of video is to register itself to be called back in one frame_time.

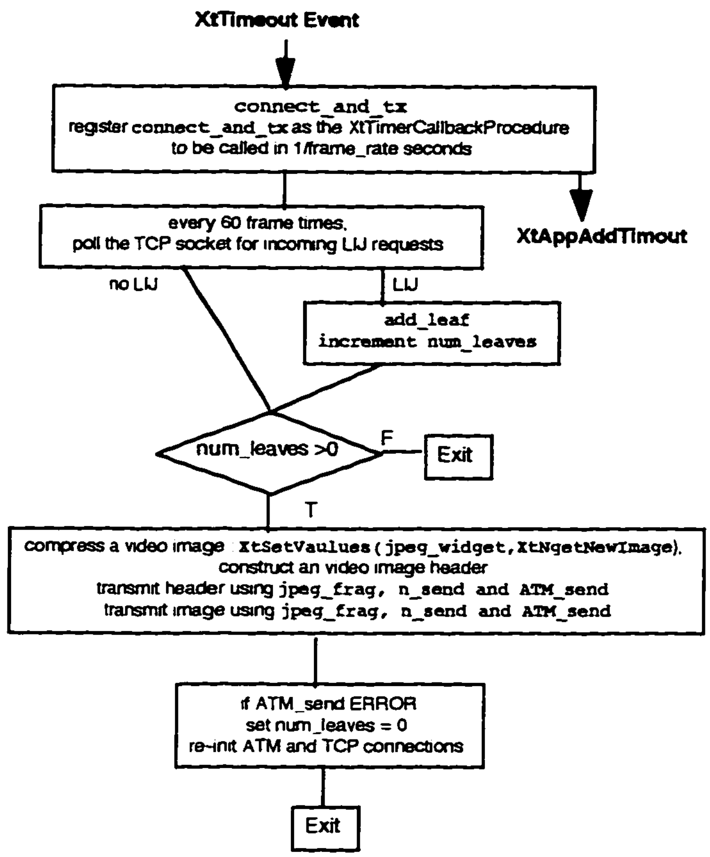


Figure 5-8 : Connect_and_Tx XtTimerCallbackProcedure in rvv.c

The TCP socket is polled via Connection Control, every 60 frame times currently, for any incoming Leaf Join Requests. This means that it can take over 4 seconds to respond to such a request but keeps processing for most frame times minimal. Upon receipt of an LJR, Connection Control issues an `ATM_setup` to add the branch to the new leaf recipient of this video stream. This terminal is the Root of that connection and all recipients are Leafs.

The parallax digitizer runs continuously updating the local display with full resolution 30 fps video unless it is told to freeze frame via the Video Config GUI and VideoIn widget. The JPEG widget on the other hand compresses a single image at a time under program control. Every time the XtTimerCallbackProcedure is entered, a single image is compressed into a JFIF (JPEG File Image Format) file that is stored in RAM (Shared Memory). At this point, the Video Adaptation function takes over, creating a "Video

Header". The video header ensures the receiver (VRX process) can interpret the incoming video stream.

Header Delimiter
JFIF Size
Sequence Number (RSN)
Time Stamp
Changed Flag
Width
Height
Q-Factor

Figure 5-9 : RendezView Video Header

5.5.1. Video Adaptation and Fragmentation

After the video header is created, it is transmitted within a single AAL-5 PDU. At the receiving end of the connection, the VRX process can search incoming AAL-5 PDUs for one carrying a Video Header (identified by Header Delimiter). The receiver is thereby synchronized with the video stream structure of video header with length field, followed by a JFIF file of that length. Furthermore, it has sufficient information to lip-synchronize it with a parallel audio-stream through remote time stamps or sequence numbers

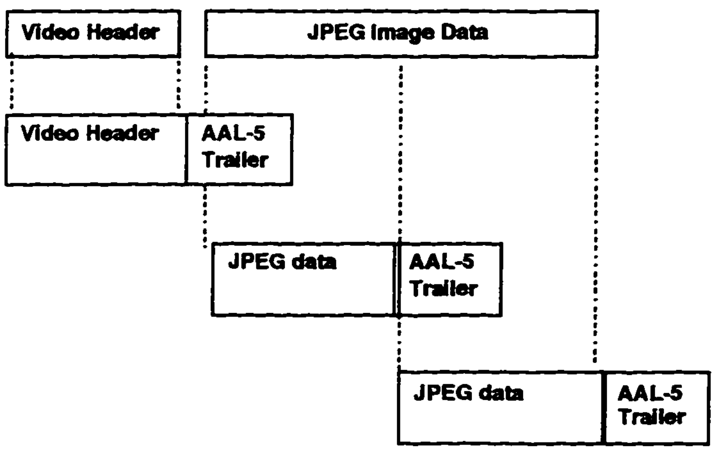


Figure 5-10 : Video Fragmentation into AAL-5 Frames

While the video header always fits within a single AAL-5 PDU (in fact a single ATM cell), the video image data size depends on various settings chosen in the Video Config

GUI including pixel resolution, compression factor and colour or black-and-white. Hence the video data can not in general fit into a single AAL-5 PDU. The FORE ATM_SEND interface permits the transmission of any data unit up to the system MTU (Maximum Transmission Unit) size. The MTU size used in RendezView is 9188 octets so any image data files greater than that size are fragmented by the jpeg_frag function.

The RendezView video reception subsystem is illustrated in the figure below. The source code for VRX.c is considerable shorter than for the transmitter (RVV.c) since VRX does not include a Video Config GUI; the source has total control over video coding.

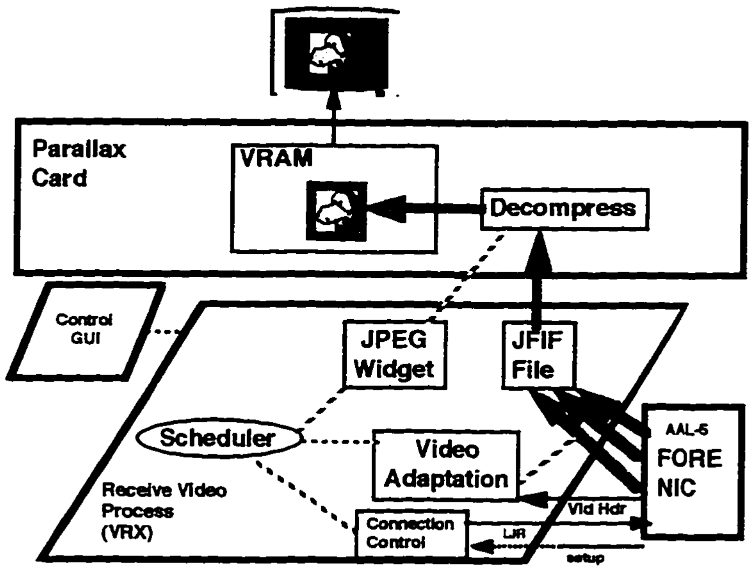


Figure 5-11 : Video Receiver Susbystem

VRX sends Leaf Join Requests to video sources (Roots) via a TCP socket. It then receives an ATM setup form the network when the Root adds this terminal as a Leaf on a point-to-multipoint connection.

Scheduling in the video receiver is not tied to a system timer, unlike the transmitter. The receiver does a busy wait and displays images when they arrive. An X-Systems facility called XtAppWorkProc is used to achieve this busy-wait.

When the first video frame arrives at the receiver, the difference between the timestamp in the Video Header and the local system time is calculated. This difference is called the STD (System Time Differential). The STD allows the timestamp contained in any video header to be compared to the local system time to determine if the video frame has arrived "on time". Since VRX does not have guarantee processing resources, it can, when the computer is overloaded, get behind in processing video. To avoid buffer overflows and wasting of CODEC resources decompressing untimely data, any video frames arriving 50mS later than indicated by header time stamps are discarded.

The flow of the vrx.c process is illustrated in Figure 5-12.

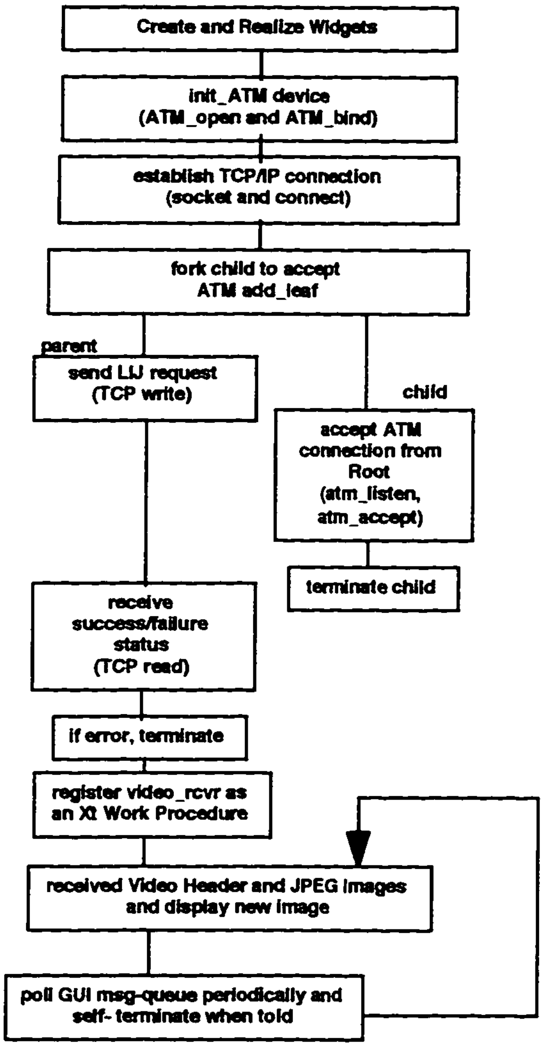


Figure 5-12 : Process Flow for vrx.c

5.5.2. RendezView Widget Hierarchies

"Widgets" are the X Windows programmatic objects that are realized as windows and buttons on the display. Parallax uses X-Windows for its programming environment, hence X-Windows programming constitutes most of the RendezView video subsystem source code. As a reference for possible future development of ATM RendezView, the structure of the widget hierarchies used is illustrated in the following figures.

Three processes within ATM RendezView contain X Windows Widgets: the connection configuration GUI (gui.c), the video transmitter (rvv.c) and the video receiver (vrx.c).

5.5.2.1. GUI.C Widgets

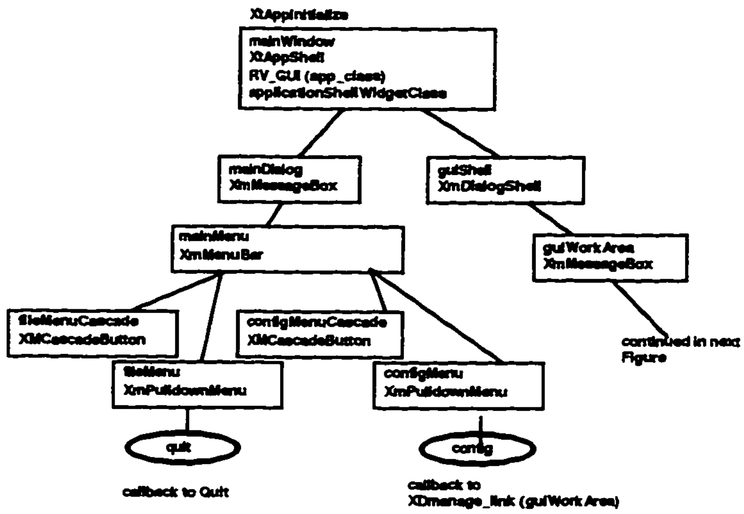


Figure 5-13 : GUI Widgets 1

The gui.c process has the widget tree hierarchy shown above. Pushbuttons with active callbacks are shown as thick ellipses.

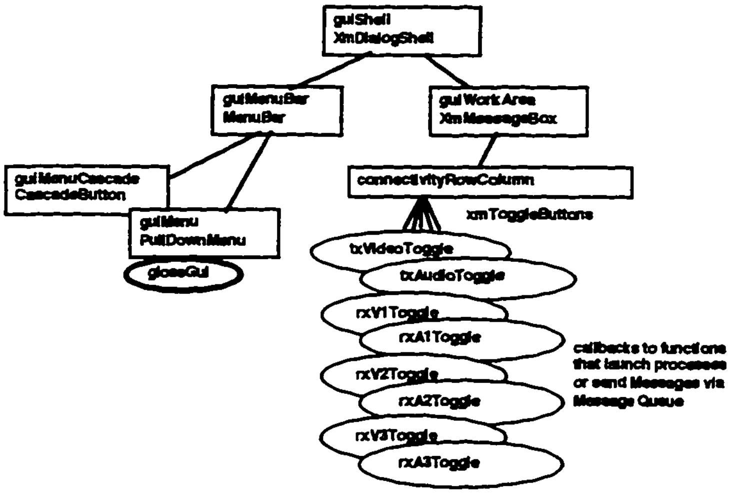


Figure 5-14 : GUI Widgets 2

5.5.2.2. RVV.C Widgets

The root structure of rvv.c is shown below, There are 4 root branches. The digitizer branch is shown in its entirety. The digitizerWidget is a Parallax defined widget responsible for displaying local uncompressed video (from a camera) in a window. The JPEG branch is also shown in its entirety. The jpegWidget is a Parallax defined widget that is responsible for compressing the images displayed by the digitizerWidget into JFIF files that are passed to the video adaptation layer. The two other branches require more figures to document.

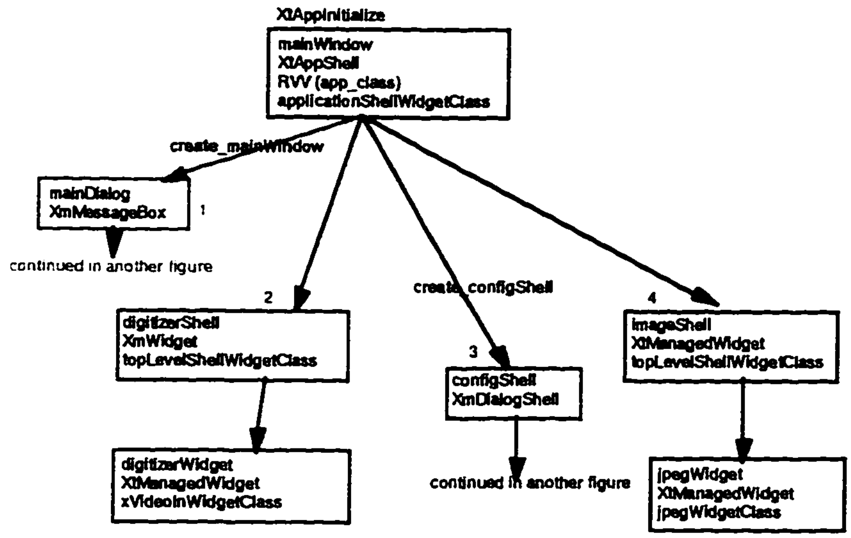


Figure 5-15 : rvv.c Root Widget Structure

The mainDialog hybrid widget has only two buttons with implemented callbacks. One is used to quit rvv.c. The other is used to pop up the video config GUI. Other buttons are provided for future enhancements to the program.

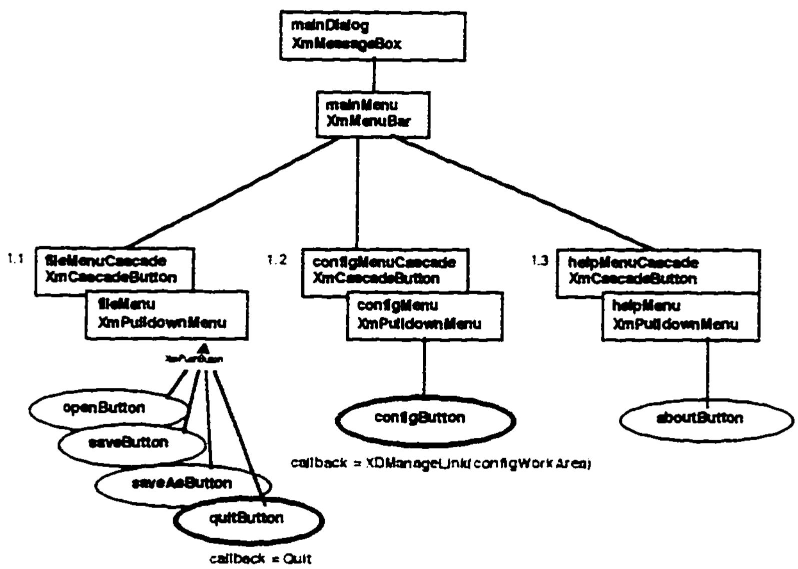


Figure 5-16 : Main Menu Widgets

The configWorkArea provides the various buttons and slider controls to establish parameters such as image resolution, frame rate, compression factor, brightness etc. Its structure is shown below.

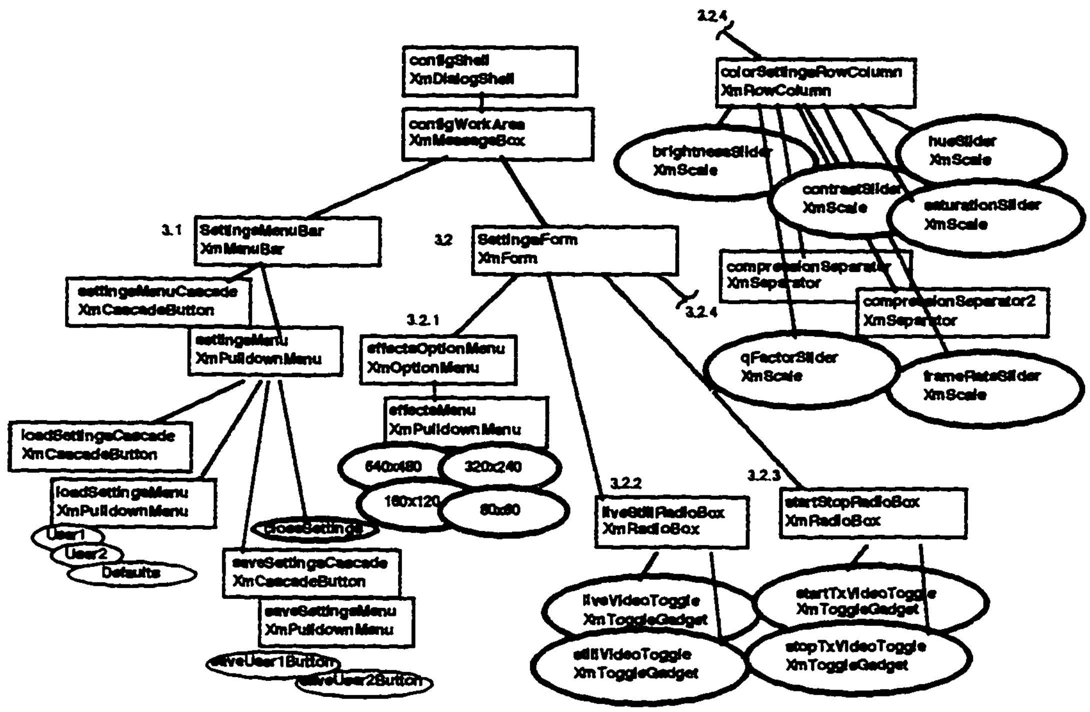


Figure 5-17 : Video Configuration Widgets

5.5.2.3. VRX.C Widgets

The video receiver process, vrx.c, has a very simple widget structure as shown below since it provides no GUI for user control. All video controls are within the transmitting root rvv.c process.

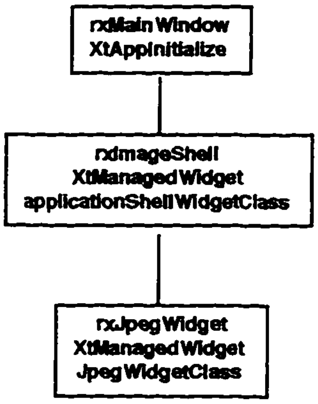


Figure 5-18 : Video Receiver Widgets

5.5.3. Data Rates with RendezView

JPEG compresses images to smaller size files using Discrete Cosine Transform (DCT) based compression to remove two dimensional (spatial) redundancy in an image. This results in some loss of image quality depending on the compression factor used. JPEG compression and decompression can be done in software but hardware compression cards are available for a large number of platforms thanks to chips developed by companies like C-Cube since around 1990. Hardware JPEG Coder/Decoders (CODECs) facilitate Motion JPEG video at up to 30 image frames per second. Because of the early availability of affordable JPEG CODECs like those from Parallax, JPEG has been used extensively for prototypes and early-products for digital video applications. However, motion JPEG will be superseded by MPEG which achieves higher levels of compression by further removing temporal redundancies.

Figure 5-19 is a table listing calculated data rates based on measured image sizes when using RendezView. The actual compressed image size depends upon the uncompressed image spatial frequency content.

Image Dimensions	Uncompressed Data Rate	Quantization Factor = 25	Quantization Factor = 100	Quantization Factor = 400
640x480	221 Mbps	18 Mbps	7 Mbps	3.4 Mbps
320x240	55 Mbps	5 Mbps	2.5 Mbps	1.2 Mbps
160x120	14 Mbps	2 Mbps	720 kbps	360 kbps

Figure 5-19 : Measured Data Rates for 24-bit Motion JPEG at 30 fps

6. Conclusions & Recommendations

ATM RendezView (Chapter 4) is a prototype of a video conferencing system that was built to demonstrate the feasibility of two key design goals:

- Lower equipment and service deployment costs by eliminating the need for Multipoint Control Units (MCUs) and Channel Aggregation Inverse Multiplexors (I-MUXs).
- Enhance service features by user control over modes of operation beyond switched presence conferencing (seeing the single current speaker) and improve service quality (audio/video fidelity and synchronization) beyond what is currently available.

ATM RendezView utilizes multipoint ATM Switched Virtual Channels and a decentralized architecture to achieve these goals. ATM provides significantly better performance and dynamic control flexibility than other networking technologies; ATM's support for dynamically re-configurable point-to-multipoint connections is the fundamental enabler of this novel architecture for multipoint conferencing. Although the current installed based of ATM infrastructure is minimal, the deployment of ATM is accelerating both in Enterprise Intranets and Public data networks including the Internet, driven by short term pragmatics of improving backbones for data networking and longer term expectations of multiservice support.

The following subsections summarize the key observations made during the process of developing RendezView.

6.1. Network Facilities and System Architecture

The ATM RendezView prototype demonstrates that high quality, continuous presence , multipoint video conferencing is feasible with generic packet networking (i.e.. ATM), without the need for centralized MCUs or service dedicated terminal equipment.

ATM's Point-to-Multipoint connections and UNIv4.0 Leaf Initiated Join signalling provide ideal infrastructure for such a service.

Deployment of a multipoint video conferencing service based on the RendezView architecture would be much less costly for a service provider than a system based on today's service specific Multipoint Control Unit hardware investment. RendezView requires only generic ATM (UNIv4.0) networking. Once ATM networks are in place, driven by this and/or other services, providing the service requires only capacity engineering and the addition of an access control server.

Using ATM's point-to-multipoint spanning tree connections optimizes performance for audio and video distribution to conferees, while minimizing the load on the source terminal and the network, thereby ensuring scaleability of conference size. ATM to the desktop would be a significant enabler for wider use of these services.

Use of Leaf Initiated Join signalling is viable and results in more flexibility in how a video conference is viewed by the user than with current MCU-based switched presence conferencing. Generic ATM network capabilities (UNIv4.0) are sufficient to support multipoint conferencing without service specific network equipment.

If it were available, Multipoint-to-multipoint Virtual Path SVCs would be an attractive alternative to Pt-Mpt Virtual Channels for multipoint conferencing. A new conferee would only need to issue a single leaf join request to join a conference and would then distinguish sources by the VC within the VP. Unfortunately, switched VPs are not yet part on ATM standards. Furthermore, containing 64K VC connections each, VPs are a scarce resource. An ideal solution to the conferencing problem would introduce a new level of connection multiplexing in ATM between the VC and the VP. It could be called a Virtual Bundle of, say 128, VCCs. Unfortunately, ATM ASICs available on the market today would not be able to support such a concept.

The systems being used on the Internet today for low quality video conferencing represent a beginning to a sound basic architecture for improved multipoint conferencing, but they are running on a yet inadequate network and still utilize reflectors for multipoint operation. Once network and access infrastructure is upgraded, a significant opportunity exists to exploit these architectures for widespread use of conferencing services. This is likely to occur first in corporate Intranets.

AAL-5 is sufficient for audio and video and it permits tuning of the packet size to be used as an engineering trade-off between the probability of network errors and minimization of processing load on the end station. ATM allows larger packet sizes than today's Internet and, together with reducing the number of protocol layers to be processed, can significantly reduce terminal processing loads.

On the various vintages of ATM Network Interface Cards used with RendezView, one had difficulty keeping up with the data rate when multiple video streams were being received. An ATM NIC should be able to keep up with the line rate.

6.2. Computing Technology

While it is feasible to accomplish video conferencing using currently available real-time facilities in end-systems/workstations, Real Time Operating Systems supporting deterministic allocation of computing resources to media streams would significantly ease design and improve performance.

Development of a standard Application Programming Interface (API) for ATM signalling and data transfers would greatly encourage the development of multimedia networking applications by making it easier to port code between platforms. The API should be event driven; the arrival of a data frame should awaken the process that needs the data. The API used in RendezView development burdened the processor required multiple polls and reads per frame.

Even when hardware CODECs are used, the sheer volume of data transferred in Audio/Video applications can load CPUs and computer bus immensely. Novel hardware architectures, such as a separate real-time bus for audio and video from the Network Interface Card to CODECs would free the rest of the computer for other applications that a user may wish to run in the background while conferencing.

Media stream autonomy (being carried on separate connections) eases software development by permitting concurrent processes to handle each media.

6.3. Audio Coding

Audio processes should be given the highest operating system priority since audio quality and continuity is critical to user perception of conference usability.

Use of Linear PCM for audio coding facilitates simple and efficient mixing and hi-fi quality in terminals at the price of some network bandwidth. In fact, delays were kept so low with RendezView that synchronization never had to be implemented. In a more general purpose application, intermedia synchronization can be accomplished via simple time stamp and delay-if-necessary processing.

Audio needs to be continuous presence (all can speak at any time). The use of Constant Bit Rate (CBR) audio in the RendezView prototype presents a scalability issue since, as the size of the conference grows, the number of continuous streams to be received and mixed increases. An intuitive solution is to use VBR audio with silence suppression (i.e.. no data is sent when the participant isn't speaking). Thus a audio mixer could be engineered to handle a moderate number of truly concurrent speakers while still permitting anyone (not everyone!) to interrupt at anytime.

6.4. Video Coding

High bandwidth is necessary for high-quality and low delay video. Video compression achievable for conferencing is less than for 1 way broadcast due to conversational delay constraints. While it is feasible to use Motion JPEG for video coding, hardware assisted MPEG-2 would be a preferred CODEC technology since it achieves higher compression. However, MPEG coding options must be chosen such that end-to-end delay is minimized for natural conversational interaction between conferees. This delay constraint thereby limits the level of compression that can be achieved with MPEG for real-time conversational services to less than that achievable for stored or one-way broadcast video. A reasonable balance may be to simply send DCT differences from the previous frame.

Layered coding, in which video sub-bands of various resolution are sent on different connections, would further enhance the conferencing service by permitting terminals of varying capabilities to observe the video streams in the quality they are capable of. It would also enable a user to choose to see more people (continuous presence) rather than just 1 or 2 at high quality.

Most CODEC cards sold today support a single full-motion (30 frames per second) compression or decompression. When several video streams are simultaneously compressed and decompressed, some frames need to be skipped in each stream, achieving increasingly jerky motion as the frame rate drops. For bi-directional multipoint conferencing it would be ideal to have a CODEC that could decode some small number (say 4) of full-motion streams while encoding another one . Costs for that much processing power are still inaccessible.

6.5. Commercialization

Further development of RendezView would be required for commercialization. Features such as the Registrar are necessary for access control security.

It's argued here that significantly higher bandwidth than is affordable today is necessary for high quality video conferencing. Service providers must be provided a way to tariff such a service at rates different than that for voice telephony bandwidth. In particular, the Registrar provides the service provider with control of who has access to the conference service and hence a mechanism to charge for conference connections at a tariff related more to the perceived value of the service than the bandwidth used.

The Registrar service further facilitates multiple Virtual Private Conferences to concurrently run over the same ATM network.

The Graphical User Interface for RendezView is crude and would require significant development to commercialize.

RendezView supports only audio and video. Incorporation of whiteboard and slide show capabilities would be necessary.

Commercial success also demands porting the system to a common computing platform such as the PC. Furthermore, IP may continue to be the point of interoperability between all network technologies and hence porting ATM RendezView to IP RendezView may be critical.

7. APPENDIX A - Initial Design Approach: H.261 Video Multiplexing for Continuous Presence

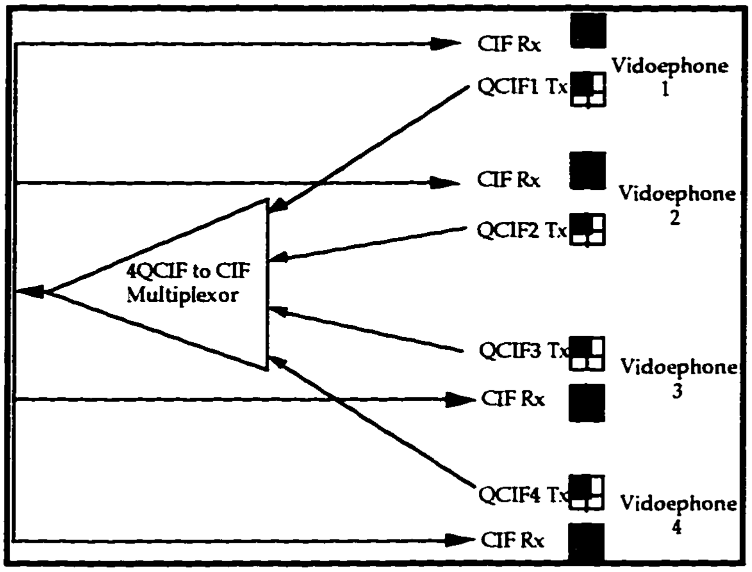
One characteristic of video conferencing systems is whether they are continuous presence or switched presence. Switched presence systems, where all participants see a single current speaker are the most commonplace. Continuous presence (seeing multiple site simultaneously) is more desirable from a user perspective, but significantly impacts MCU cost, demands higher transmission bandwidth and results in increased end-to-end delays.

Terminal requirements to permit continuous presence conferencing of Px64 video include:

- capability to transmit at QCIF and receive at full CIF.
- wideband (>4xDS0 or P=4) capable on the receive side.
- participating sites/terminals must transmit at the same frame rate.
- users must be willing to tolerate significant additional delay.
- special functions such as freeze frame or re-allocation of video bandwidth to Data 1 (image, facsimile, etc.) could be provided with additional complexity.

H.261 (Px64) can support continuous presence conferencing for up to 5 sites through the capability of multiplexing 4 Quarter CIF picture format source data streams (from 4 sites) into a single full CIF stream for transmission to the 5th site. At the 5th site, the observer will see 4 other sites in the 4 quadrants of the full CIF picture. 5 multiplexors would be required to perform this for 5 sites.

A simpler system, permitting only 4 sites in total but requiring only 1 multiplexor is shown schematically below. Each site will see the same full CIF picture with it's own video echoed back in one of the quadrants.



Clearly, the videophone terminals must be capable of receiving full CIF streams. The minimum compliance for a H.261 CODEC is only QCIF capability and clearly such low end terminals could not participate (on the receiving end) in such a conference. It should also be clear that sufficient bandwidth must be available on the receive line. If 20xDS0 (P=20) is available, then the maximum Tx rate for any source is 5xDS0 (P=5). This form of continuous presence conferencing could not be done for terminals having only 'narrow' bandwidth (Basic Rate ISDN access or lower).

Bandwidth could be conserved by, unlike the diagram above, not echoing an individual's image back to them (since they have it locally anyway). This would permit 5 participants in the video conference (if 5-choose-4 = 5 multiplexors were used).

The most performant (least delay) video would operate at the coded stream level, TDM multiplexing these streams without needing to decode them. The following section proposes the type of processing that would be required to do such multiplexing by examining the H.261 coded stream field formats. It is clearly revealed that TDM multiplexing of H.261 streams is far from trivial. In practice, MCU products perform continuous presence multiplexing of 4 video streams into 4 QCIF quadrants by decoding each stream in the MCU. Not only does this result in significant extra hardware cost but also introduces significant extra delay (100's of mS) in decoding and re-encoding. This added delay defeats the whole objective of continuous presence conferencing: natural dialogue.

The following section summarizes the initial design approach taken by the author to create a continuous presence conference bridge for H.320 conferencing systems. The H.261 coded stream is examined to see if the Group of Block data structures could be separated and interleaved placing 4 QCIF images into 4 quadrants of a CIF image stream. This approach was ultimately discarded in favour of the distributed architecture of ATM RendezView but is described here to illustrate the complexity and limitations of continuous presence conferencing with H.320.

transmitted first

PSC	TR	Type1	PEI1	Parity	PEI2	PSpare	GOB Data

PSC-Picture Start Code: frame delimiter

TR=Temporal Reference:mod32 time stamp corresponding to picture frame number at 30 Hz

Type1: split screen, document camera, freeze, release, QCIF or CIF

PE1=Picture Extra Insert Information 1: isparity active?

Parity: Odd Parity of bit planes for Y, CB and CR in the previous image

PE2/PSpare: optional usage

GOB=Group of Blocks Data: 3 fields for QCIF, 12 for CIF

H.261 Data Stream

If the Type1 field indicates that the picture was encoded in Quarter Common Input/Interchange Format (QCIF) then it may be possible to interlace up to 4 of these streams into a single full CIF format. This may be done without fully decoding the input streams (recreating the QCIF images/pictures in the multiplexor). Several conditions, such as the wideband capabilities cited above, must be satisfied.

Frame rates at the sources should be the same. The source coder clock should be synchronized to the network and all sources should generate frames with the same frequency (i.e. ~30Hz, ~15Hz, or ~10Hz etc.). The difference between successive Temporal Reference (TR) values in each stream will reveal the frame rate from that source. A difference of 1 for example indicates that frames are being generated at ~30Hz (2 for 15Hz, 3 for 10Hz, 4 for 7.5Hz, etc).

Even if the source CODECs are synchronized to the same network clock and sending at the same frame rate, there does not appear to be a method of synchronizing the frame starts among multiple CODECs. Hence, a minimum delay of 1 frame time (~33.4mS, ~66.8mS or ~100.2mS etc correspondingly) may be required at a 4QCIF/CIF Multiplexor to ensure that data is buffered for all 4 streams. This amount of delay is dangerously high for real time interactive communication.

The frame rates from sources could be different as long as the multiplexor can buffer and repeat old frames for the slower input streams and create an output stream at the rate of the fastest source (or ignore some frames from the fastest source(s) and transmit at the rate of the slowest). The combined output stream from the multiplexor will require appropriate coding of the TR field according to the output frame rate.

The multiplexor would need to be able to recognize a freeze frame signal in any incoming stream and buffer and repeat that frame. This, unlike in single point-to-point video telephony, would make it impossible to re-allocate bandwidth on the fly to a data channel for transmission of an image during a freeze frame on the video. To switch bandwidth from video to data transmission in this continuous presence scenario, the receiving terminals would need to be able to freeze frame one QCIF quadrant of their full CIF display and then along with the multiplexor recognize the incoming data stream (image, graphic, text,...) and treat it accordingly. Alternatively (and it would appear more simply) the entire full CIF frame could be frozen at all terminals while a single source sent data using 1/4 of the available bandwidth.

Assuming the above conditions are satisfied, the functions performed by the video multiplexor during a stable state of multiplexing 4 continuous streams is as follows.

1) Each stream is fed into a fifo buffer to permit synchronization of the 4 streams. A minimum delay of 1 frame time must be introduced to permit synchronization. At frame rates of 30Hz, 15Hz and 10Hz this corresponds to delays of 33mS, 67mS and 100mS respectively.

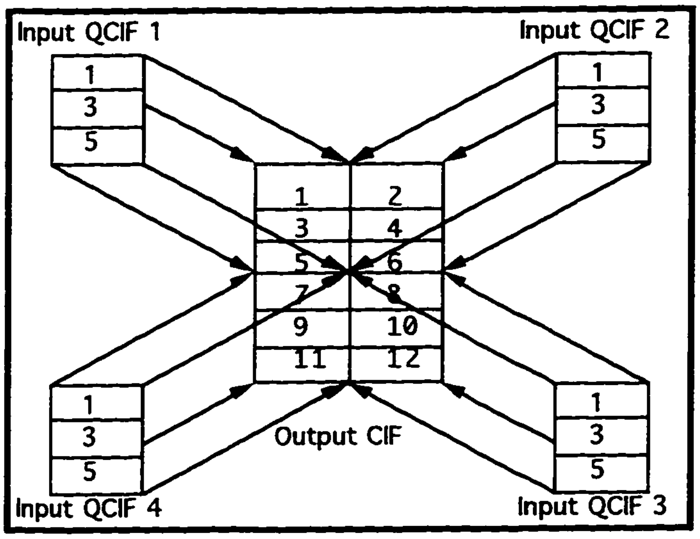
2) A decision must be made on which of the 4 quadrants of the full CIF each input QCIF stream will be mapped into. There could be a default order (dependent on, for example, the chronological order people called into the bridge) but this should ideally also be user programmable. (see fancy 3-D diagram a couple of pages further on)

3) The frame header for the output full CIF frame can be pre-calculated.

   - Picture Start Code is static.

- Temporal Reference increments with each subsequent frame depending on the output frame rate (by 1 for 30Hz, by 2 for 15Hz, by 3 for 10Hz, by 4 for 7.5Hz, etc.)
- Type1 in-band signalling information:
- Bit 1 - 0...split screen normally off
- Bit 2 - 0...document camera off
- Bit 3 - 0...freeze picture release (i.e.: used if image was frozen during switching in a new source QCIF stream or data transfer using video bandwidth was done)
- Bit 4 - 1...format - full CIF
- PEI1 bit = 0 normally indicating that no parity information is to follow. The parity information can only be calculated if the incoming 4 streams are decoded into PCM values for Y, Cr and Cb. This could be done but would eliminate the simplicity (and minimal hardware) required to simply multiplex the 4 streams.
- PEI2 bit = 0 indicating there is no PSPARE field (reserved for future use)
- Parity and PSPARE would then not exist (0-length).

4) The Group Of Blocks (GOB) data fields can then be inserted behind the frame header described above. There are 12 GOBs, 3 in each QCIF stream. The multiplexor must transmit the GOB data sequentially from GOB 1, 2, 3,... to GOB 12. This means that data must be read from input stream 1 then 2 (3 times) then input streams 3 then 4 (3 times) (see diagram below).



GOB Group Number mapping for QCIF Multiplexing

5) A CIF picture is divided up into 12 Groups of Blocks. Each GOB Data field has the following format.

transmitted first ←

GBSC	GN	Type2	Quant1	GEI	Gspare	MB Data

GBSC=Group of Blocks Start Code
GN=Group Number
Type2: Signalling bits
Quant1= Quantizer Information: indicates which Quantizer to use
GEI=GOB Extra Information Indicator
Gspare: GOB Extra Information
MB Data: MacroBlock Data

GOB Data Field Format

Each Group Of Blocks that has blocks that had significant changes during the last frame time is transmitted in the format shown above. At most, 12 GOBs are transmitted for a full CIF picture.

The fields that a 4QCIF-to-CIF multiplexor may have to process are:

- Group Number must be mapped as shown two diagrams above.

6) Each GOB is divided up into 33 Macroblocks. Each Macroblock is 16x16 pixels of Luminance (4 8x8 blocks) and two 8x8 blocks of Chrominance (1 each for Cr and Cb). Information is only transmitted for those Macroblocks having significant change during the last frame time. Each Macroblock Data Field transmitted has the following format.

transmitted first ←

MBA	Type3	Quant2	MVD	CBP	Block Data

MBA=Macroblock Address Number (1-33)
Type3:Variable Length Codeword
Quant2=Quantizer Indicator
MVD=Motion Vector Data
CBP=Coded Block Pattern
Block Data: DCT coefficients followed by End-of-block delimiter

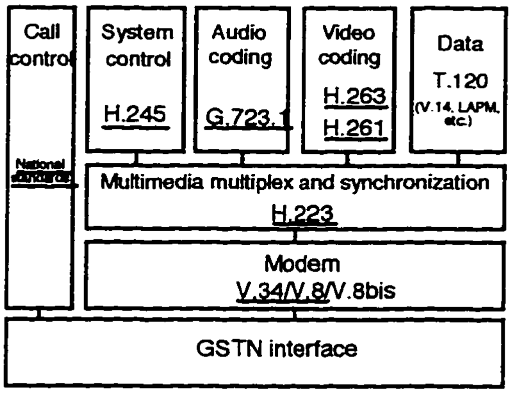
Macroblock Data Field

The video multiplexor need not process any of the Macroblock Data fields.

8. APPENDIX B - ITU-T H.3xx Recommendations and Protocol Stacks

8.1. Protocol stacks of the H-series audiovisual communication systems

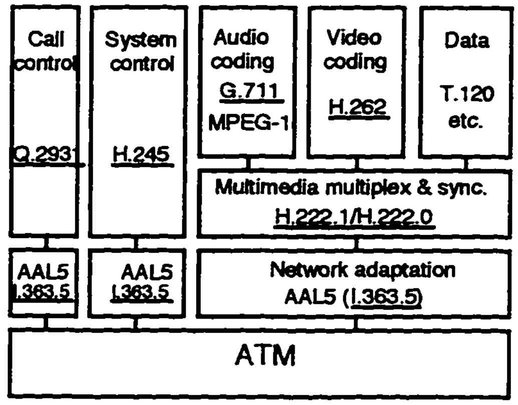
The following figures illustrate the protocol stacks of each H.3xx system. Underlined elements are mandatory, while others are optional. There are also multipoint and security related Recommendations for H.320 terminals; H.231, H.243 for multipoint and H.233, H.234 for security that are not shown.



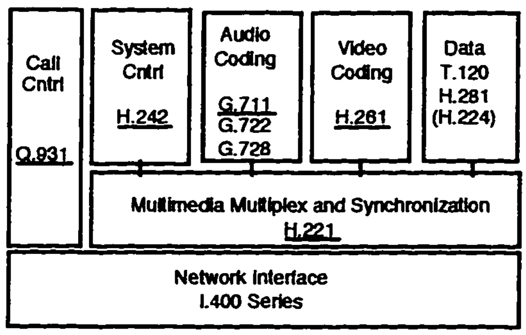
H.324 POTS Terminal Stacks

Manufacturers believe that standards will generate high production volume, thereby lowering costs below $1,500 range for earlier videophones. "Introductory" prices cited were typically in $350-$500 range, with Internet-based software products as low as $200 (with benefit of no long distance charges).

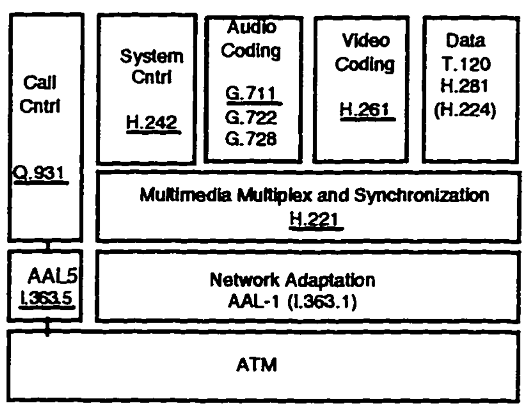
Videophones still have problem of low frame rates, which means jerky motion images and unrealistic audio synchronization. Internet-based products typically operate at rate of 2-6 frames per sec. (fps), compared with full-motion video's 30 fps. Telephone-based videophones claim rates of 15-16 fps, assuming that telephone line is absolutely clean and that little motion is being displayed. In real world, rates probably are closer to 8 fps, or with much motion, as low as 2-3 fps. Furthermore, multipoint operation still requires an expensive and performance degrading MCU.



H.310 ATM Native Mode Terminal Stacks



H.320 ISDN Terminal Stack



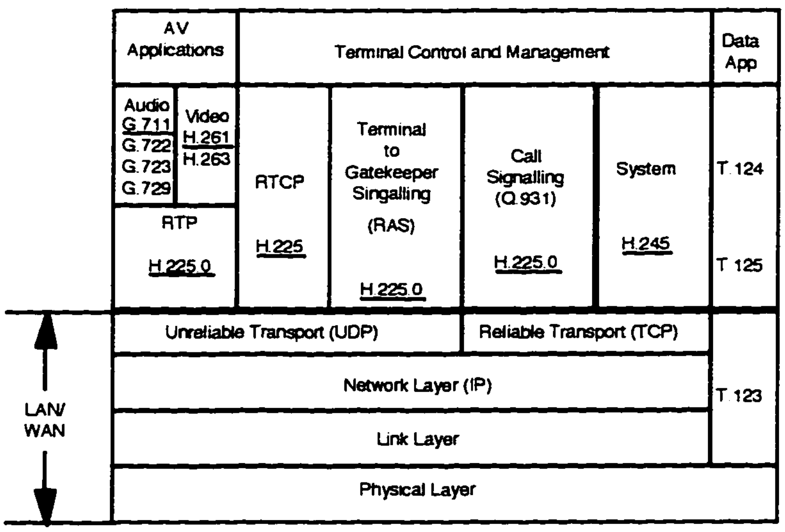
H.321 ATM Stack for Interworking with ISDN H.320

An MPEG protocol stack consists of a core physical layer, a network and transport adaptation overlay and several elements above network transport. Connection control and H.245 session-control signals reside contiguously with the transport stream. The H.222.0

transport stream resides below an H.221 program stream (multimedia multiplexing). Above the program stream reside digitized audio using ISO/IEC 13818-3 voice coding, digitized video utilizing H.262 video CODEC, and signals for user-to-user control.

The MPEG-2 system may also be used for MPEG-1. The differences between MPEG-1 and MPEG-2 comprise four areas:

- MPEG-2 has a higher resolution than MPEG-1, which results in compressed video of 3 to 8 Mbps, as compared to 2 to 3 Mbps for MPEG-1. MPEG-1 has a provision to encode at an even lower resolution to allow even lower transmission rates (1.5 Mbps) for certain applications such as CD-ROMs.
- MPEG-2 has a richer syntax.
- MPEG-2 has a transport stream that maps program streams into standard (constant) packet size for facilitating transport through various networks, including ATM. An MPEG-1 program can be mapped in an MPEG-2 transport stream.
- MPEG-2 allows for burstiness (variable bit rate, or VBR) of the encoded video, thus reducing delay compared to MPEG-1, which runs at a constant bit rate.



H.323 Terminal Stack for Non-Guaranteed QoS Packet Networks

H.230 is the recommendation for signals controlling the video and audio signals, such as muting audio, or activating and deactivating specific video sources. An example where activation/deactivation of a video source is utilized is the use of two cameras, with one projecting conference participants and the other projecting a writing tablet or similar object for display. Only one camera would be activated at a time to remain within the channel bandwidth and retain a good video quality.

H.233, the "confidentiality system," is an encryption spec. H.242 is a set of control signals used for handshaking between terminal equipment to establish compatible channel capacity, compression algorithms and such. H.243 consists of procedures for setting up H.242.

H.262 is the ITU-T video coding standard for MPEG systems, both 1 and 2. The MPEG-2 audio coding standard is ISO/IEC IS 13818-3. H.222.1 is a standard for multiplexing MPEG-2 program elements specifically for ATM networks, i.e., each element (video, audio, etc.) is assigned a different identifier which maps directly into a virtual-connection identifier.

H.222.0 is also a draft recommendation for MPEG-2 systems which is the same as ISO/IEC IS 13818-1. This includes definitions of the program stream, transport stream, the system syntax, and specific profiles. Profiles are pre-defined parameter sets (resolution, interoperation, and bit rate parameters) that can be selected by users to simplify systems interoperation.

H.245 is also a draft of a control protocol analogous to H.230 and H.242. Additionally, there is a standard for controlling digital storage media in the MPEG-2 systems, i.e., ISO/IEC IS 13818-6. The MPEG-1's standard number is ISO/IEC IS 11172, and it is composed of five parts: systems, video encoding, audio encoding, conformance testing and simulation software.

MPEG-4 is an evolving ISO standard aimed at filling the gaps in previous MPEG standards. It will support full multimedia conferencing, integration of natural and synthetic audio/video, simultaneous handling of audiovisual material to multiple destinations from multiple sources, coding tools for dynamic audiovisual objects, syntactic description of audiovisual objects to provide a formal method for describing the coded representation of objects and the method of encoding, error robustness and scalability. This standard is still in its early stages. The target schedule is to produce a committee draft in November 1997, which will become an international draft standard in July 1998 and an approved international standard by November 1998.

H.225.0 describes a method for combining audio, video, data and control information on non-guaranteed QoS LANs to provide conversational services via H.323 equipment. Topics include AV coding, control and signalling messages and methods for providing improved QoS in this environment. H.245 runs over UDP/IP or TCP/IP (call signalling and H.245 channel). H.323 Terminals and H.323 Gateways use H.225 over non-guaranteed QoS LAN environment. H.225.0 uses RTP (Real-Time Transport Protocol and RTCP (Real-Time Transport Control Protocol) for media stream packetization and synchronization for all underlying LANs.

8.2. ITU-T Recommendations for H-series Systems

Rec.	Title	Most recent approval
G.711	Pulse code modulation (PCM) of voice frequencies	1988
G.722	7 kHz audio-coding within 64 kbit/s	1988
G.723.1	Dual rate speech coder for multimedia communication transmitting at 5.3 & 6.3 kbit/s	1996
G.728	Coding of speech at 16 kbit/s using low-delay code excited linear prediction	1992
G.729	Coding of speech at 8 kbit/s using Conjugate Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)	1996
H.221	Frame structure for a 64 to 1920 kbit/s channel in audiovisual teleservices	1995
H.222.0ISO/IEC 13818-1	Information technology - Generic coding of moving pictures and associated audio information: Systems	1995
H.222.1	Multimedia multiplex and synchronization for audiovisual communication in ATM environments	1996
H.223	Multiplexing protocol for low bitrate multimedia communication	1996
H.224	A real time control protocol for simplex appplications using the H.221 LSD/HSD/MLP channels	1994
H.225.0	Media system packetization and synchronization on non-guaranteed quality of service LANs	1996
H.230	Frame-synchronous control and indication signals for audiovisual systems	1995
H.231	Multipoint control units for audiovisual systems using digital channels up to 2 Mbit/s	1996
H.233	Confidentiality system for audiovisual services	1995
H.234	Encryption key management and authentication system for audiovisual services	1994
H.242	System for establishing communication between audiovisual terminals using digital channels up to 2 Mbit/s	1996
H.243	Procedures for establishing communication between three or more audiovisual terminals using digital channels up to 2 Mbit/s	1996
H.244	Synchronized aggregation of ISDN B-channels	1995

H.245	Control protocol for multimedia communication	1996
H.261	Video codec for audiovisual services at px64 kbit/s	1993
H.262 | ISO/IEC 13818-2	Information technology - Generic coding of moving pictures and associated audio information: Video	1995
H.263	Video coding for low bitrate communication	1996
H.281	A far end camera control protocol for videoconferences using H.224	1994
H.310	Broadband audiovisual communication systems and terminals	1996
H.320	Narrow-band ISDN visual telephone systems and terminal equipment	1996
H.321	Adaptation of H.320 visual telephone terminals to B-ISDN environments	1996
H.322	Visual telephone systems and terminal equipment for local area networks which provide a guaranteed quality of service	1996
H.323	Visual telephone systems and terminal equipment for local area networks which provide a non-guaranteed quality of service	1996
H.324	Terminal for low bitrate multimedia communication	1996
H.331	Broadcasting type audiovisual multipoint systems and terminal equipment	1993
I.363.1	B-ISDN ATM Adaptation Layer (AAL) specification, Type 1 and 2	1996
I.363.5	B-ISDN ATM Adaptation Layer (AAL) specification, Type 5	1996
I.580	General arrangements for interworking between B-ISDN and 64 kbit/s based ISDN	1993
Q.2931	Broadband integrated services digital network (B-ISDN) - Digital subscriber signaling No. 2 (DSS2) - User network interface layer 3 specification for basic call/connection control	1995
Q.931	ISDN user-network interface Layer 3 specification for basic call control	1993
T.120	Data Protocols for Multimedia Conferencing	1996
T.121	Generic Application Template	1996
T.122	Multipoint Communication Service for Audiographic and Audiovisual Conferencing	1993
T.123	Protocol Stack for Audiographics and Audiovisual Teleconference Applications	1994
T.124	Generic Conference Control For Audio-Visual and Audiographic terminals	1995
T.125	Multipoint Communication Service Protocol Specification	1994
T.126	Still Image Protocol Specification	1995
T.127	Multipoint Binary File Transfer Protocol Specification	1995
V.8	Procedures for starting sessions of data transmission over the GSTN	1994
V.8bis	Procedures for the identification and selection of common modes of operation between data circuit terminating equipment (DCE) and between data terminal equipment (DTE) over the general switched telephone network and on leased point-to-point telephone-type circuits	1996
V.34	A modem operating at data signaling rates of up to 28 800 bit/s for use on the GSTN and on leased point-to-point 2-wire telephone-type circuits	1994

ITU-T Recommendations Relevant to H-series AudioVisual Communication Systems

Rec.	Content
H.24i	Interworking between different H.3xx systems
H.310 systems	Implementors Guide, Version 2 enhancements security, loose multipoint, VBR
H.320 systems	Inclusion of new audiovisual coding

H.323 systems	Implementors Guide, Version 2 enhancements for inclusion of new audiovisual coding, security, loose multipoint, medisMIB
H.324 systems	Extension to mobile environments
T.130 series	Real time audiovisual control for multimedia conferencing

Ongoing ITU-T Works Related to AudioVisual Communication Systems

F.710	1991	General Principles for Audiographic Conference Service.
F.730	1992	Videoconference Service - General : Operations and Quality of Service
F.740	1993	Audiovisual Interactive Services
H.120	1993	Codecs for videoconferencing using primary digital group.
H.130	1988	Frame Structures for use in the international interconnection of CODEC for videoconferencing or visual telephony.
H.140	1988	A Multipoint International Videoconference System.
H.200	1993	Framework for Recommendations for Audiovisual Service.
H.221	1993	Frame Structure for a 64 to 1920 kbit/s Channel in Audiovisual Teleservices.
H.230	1993	Frame-Synchronous Control and Indication Signals for Audiovisual Teleservices.
		H.320 Standard Suite for Px64 Video Conferencing
H.231	1993	Multipoint Control Units for Audiovisual Systems Using Digital Channels up to 2 Mbit/s. (see also Bellcore GR-1337-CORE, 1994, Multipoint Multimedia Conferencing Control Unit)
H.233	1993	Confidentiality System for Audiovisual Services.
H.234	1994	Encryption Key Management and Authentication System for Audiovisual Services.
H.242	1993	System for Establishing Communication Between Audiovisual Systems using Digital Channels up to 2Mbit/s.
H.243	1993	Procedures for Establishing Communications Between Three or More Audiovisual Terminals using Digital Channels up to 2 Mbit/s.
H.261	1993	Video CODEC for Audiovisual Services at p x 14 kbit/s.
H.281	1994	A Far End Camera Control Protocol for Videoconferences.
H.320	1993	Narrow-band Visual Telephone Systems and Terminal Equipment.
		Miscellaneous and Work In Progress
H.225	97 ?	Media Stream Packetization, Multiplexing and Synchronization on Non-guaranteed QoS LANs; closely tied to RFC1899 Real Time Protocol (RTP)
H.Security		Security

H.24i		Interworking
H.lcc		Loosely Coupled Conferences
H.262		MPEG-2 Video CODEC.
H.310 & H.321 Ann. A		ATM-based Terminals using AAL-1 or AAL-5 H.320 over AAL-5
H.323	97 ?	LAN & Internet Based Terminals
H.331	1993	Broadcasting Type Audiovisual Multipoint Systems and Terminals.
		H.324 Standard Suite for Audiovisual over POTS and Modems
H.324	1995?	Audiovisual System for use over POTS and Modems.
H.223		Multiplex for H.324 POTS/Modem Audiovisual Systems
H.245		Control for H.324 POTS/Modem Audiovisual Systems (H.245v3 also covers H.320 systems over ISDN interfaces using Q.931 "Call"/Connection Signalling
H.263	1996	Video Coding for low bitrate H.324 POTS/Modem Audiovisual Systems
G.723.1		Speech CODEC for H.324 POTS/Modem Audiovisual Systems
		T.120 Multipoint Data Standard Suite
T.122	1993	Multipoint Communication Service for Audiographics and Audiovisual Conferencing
T.123	1994	Protocol Stacks for Audiographicudiovisual Teleconferene Applications
T.125	1994	Multipoint Communication Service Protocol Specification

ITU-T Audiovisual Conferencing Standards

9. APPENDIX C - Comparison of Variable Bitrate ATM Adaptation Layers

9.1. Scope & Purpose

This appendix summarizes technical comparisons between the two ATM Adaptation Layers (AAL) that may be used to support variable length frame services. These AAL are:

• Common Part AAL-3/4 and
• AAL-5, the Simple Efficient Adaptation Layer (SEAL),

Discussion of non-technical (standards, market, etc.) and product implementation issues is intentionally omitted.

9.2. Conclusion

AAL-3/4 and AAL-5 provide fundamentally the same service, unassured delivery of frames with error detection. AAL-5 has some simplicity and efficiency advantages and AAL-3/4 has somewhat more functionality via MID submultiplexing. However, neither of these differences are "killer" advantages for one over the other.

AAL-5 has proven the most popular AAL and hence should be used when possible.

9.3. Comparison Summary

• MID Multiplexing & Address Space

AAL-3/4 can support more connections than AAL-5 via MID sub-multiplexing on VCC. However, no application requiring a connection address space size greater than that provided by VPI/VCI alone has been identified. VP address space appears the most scarce addressing resource and adding MID sub-multiplexing does not change this.

• MID Multiplexing & Multipoint Services

AAL-3/4 enables the use of multipoint-to-point (Merge) Virtual Channel Connections that will cell-interleave variable length frames from multiple sources onto one VC. However, no customer services requiring such connections have been identified. Multipoint services can be equally well supported via either AAL as long as only point-to-multipoint (Multicast) and point-to-point (Unicast) connections are used. OA&M cell flows, however, may require the use of Merge connections on the return path of a Multicast connection. If OA&M messages may be longer than one cell in length, some form of per-cell message identification must be used.

• **Frame-level Layer Management**

AAL frame level management functions such as Ping-test are, at present, absent from the AAL-5 definition. AAL-3/4 provides code points in its CPI field to facilitate such functions.

• **Location of Header/Trailer in BOM or EOM**

Unlike AAL-3/4, AAL-5 does not possess a header with frame length indication to facilitate dynamic buffer allocation.

If frame-level control information (equivalent to CPI in AAL-3/4) was adopted into the 2-octet Control field in AAL-5, it would be in the trailer. This would mean AAL management frames could not be identified until the last cell in the frame is received.

• **Error Robustness**

Neither AAL has a clear advantage over the other.

• **Efficiency**

AAL-5 is at least as efficient as AAL-3/4 for short frames and is always more efficient for long frames. (>485 Bytes).

• **Simplicity**

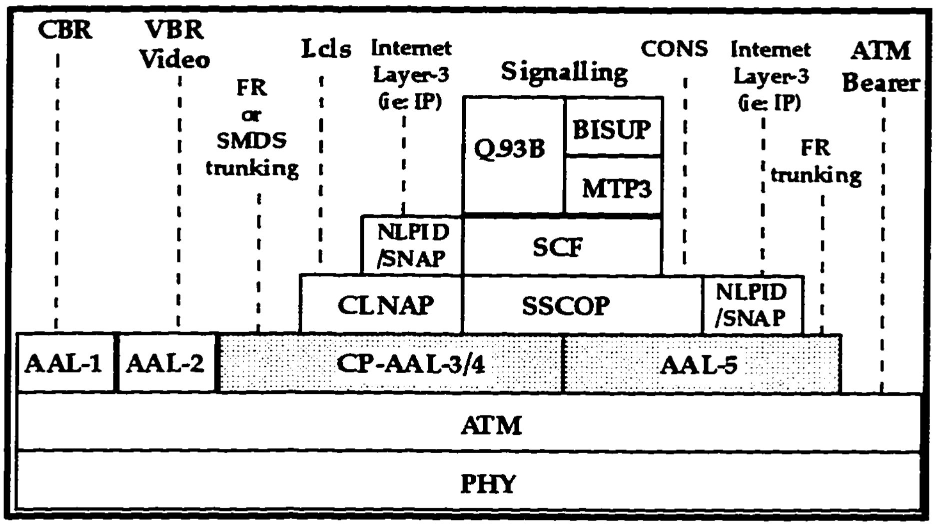
AAL-5 eliminates the need to process and keep state information for MIDs, sequence numbers and BE Tags.

9.4. Background

9.4.1. Services Context

Support for Frame Relay Services and SMDS over B-ISDN are used herein as example services that may use these AAL. B-ISDN signalling will also utilize one of the AAL. The AAL comparisons apply to frame based services in general.

The figure below provides some examples of where these AAL may be used within a B-ISDN protocol stack to support a variety of services. This figure is meant only to provide some context and should not be used as an information reference for B-ISDN protocols. Some of the protocols shown are preliminary proposals (i.e.: SSCOP and SCF) yet to be defined in any detail. Some others are outside of CCITT or ANSI standards (i.e.: IETF/Internet NLPID/SNAP). Also, a given horizontal line should not be interpreted as implying an equivalent level of service. For example, AAL-1 and AAL-2 provide drastically different services than AAL 3/4 or 5.



Example Placement of VBR AALs in B-ISDN Protocols

CLNAP	Connectionless Network Access Protocol (I.cls equivalent of SIP-3)
SSCOP	Service Specific Connection Oriented Protocol (provides assured service & possibly functions such as fragmentation and streaming support)
SCF	Services Coordination Function
MTP3	Message Transfer Part/Protocol Layer 3
NLPID	Network Layer Protocol ID
SNAP	Subnetwork Access Protocol
CONS	Connection Oriented Network Service

9.4.2. AAL-3/4 Description

Figures showing key features of AAL-3/4 and AAL-5 are provided on the following pages for quick reference. For details see references [8] & [9].

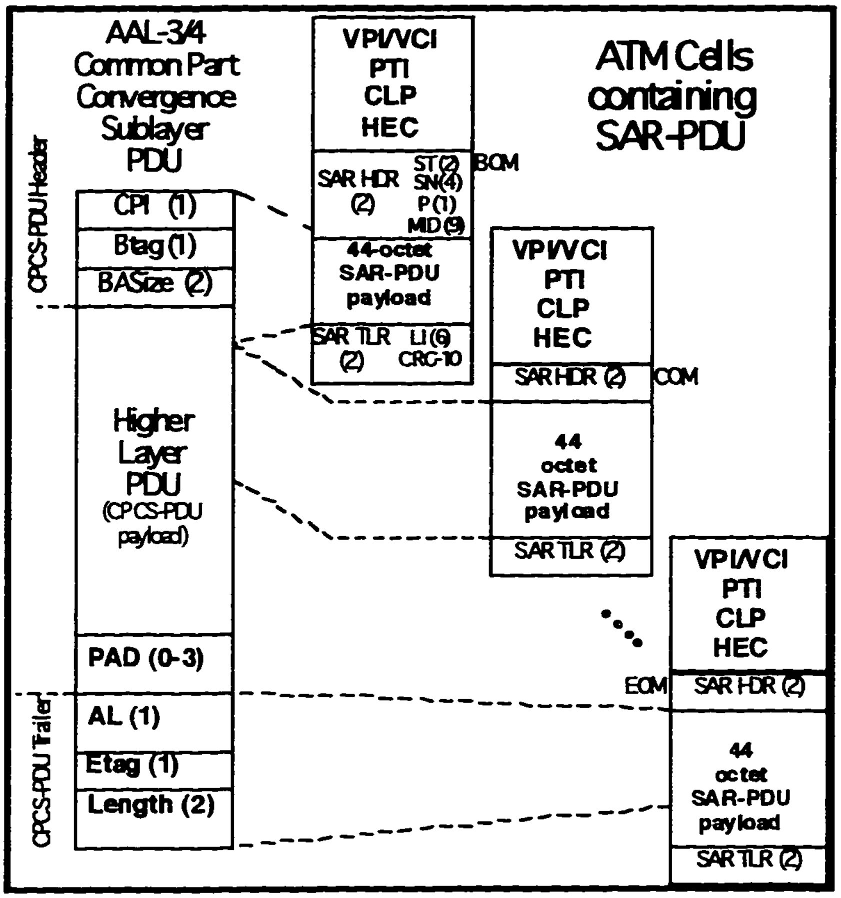


Figure 1: Common Part AAL-3/4

CPCS-PDU Fields	
CPI	Common Part Indicator
Btag	Begin Tag
BASize	Buffer Allocation Size ≥ Payload Length
AL	Alignment (unused)
Etag	End Tag
Length	Payload Length in octets

SAR-PDU Fields	
ST	Segment Type
SN	Sequence Number
P	Priority
MID	Multiplexing ID
LI	Length Indication (≤44)

Cell Header Fields	
VPI/VCI	Virtual Path/Channel Identification
PT	Payload Type
CLP	Cell Loss Priority
HEC	Header Error Control

9.4.3. AAL-5 Description

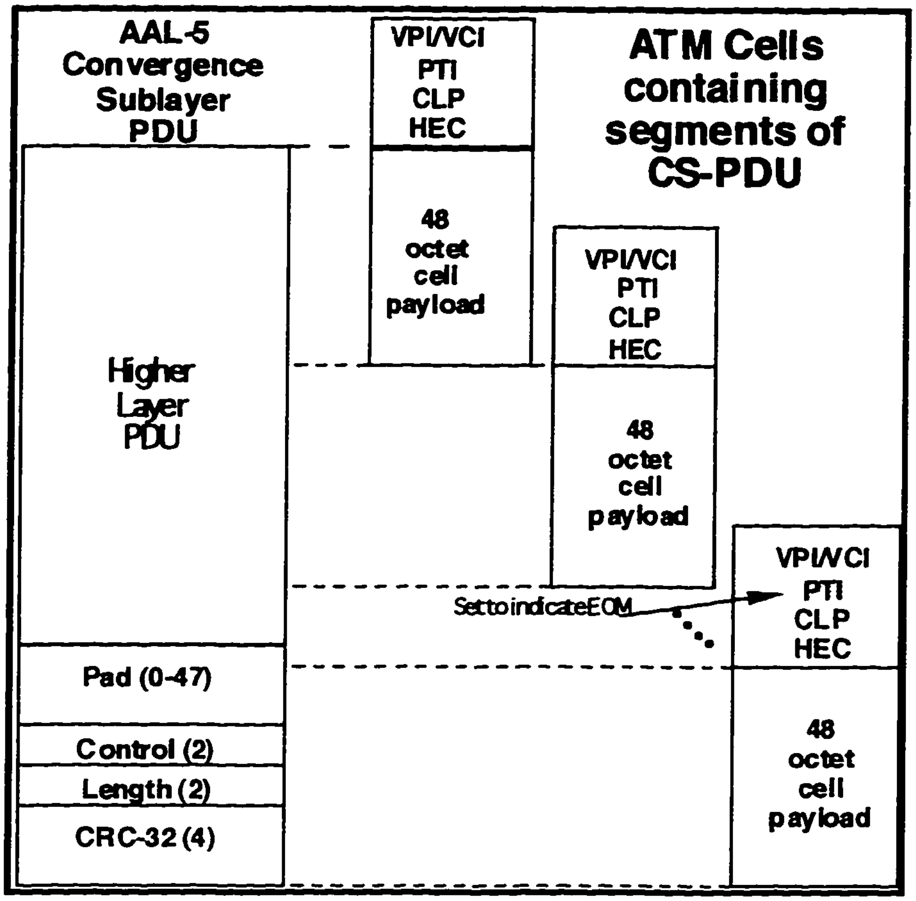


Figure 2: NOKIA Version of AAL-5

Note - Neither the ***Length*** field nor the ***Control*** field were in the original Sun Microsystems SEAL contribution to T1S1. SUNs SEAL included only ***Pad***, ***CRC*** and End-Of-Message indication in ***PTI***. The ***Length*** and ***Control*** fields were proposed in a NOKIA (Finland) proposal to ETSI on FR/ATM interworking and in the IETF draft RFC on Multiprotocol Interconnect over ATM. Precisely what fields will exist in AAL-5 once standardized is yet to be determined.

9.5. Comparisons

9.5.1. MID Multiplexing & Address Space

Two conclusions are reached in the following section. First, simple calculations (dividing transmission system bandwidth by number of connection identifiers) are used to argue that MID submultiplexing on VCI is not necessary. Secondly, if there is a possible connection granularity problem it is due to limited VPI space, and MID submultiplexing does not alleviate this problem.

Some have argued that the connection address space offered by VPI/VCI alone may prove to be insufficient, especially on very high speed NNI. The 9 or 10-bit Multiplexing ID ([1]MID) in AAL-3/4 can increase connection multiplexing address space by a factor of 512 or 1,024.

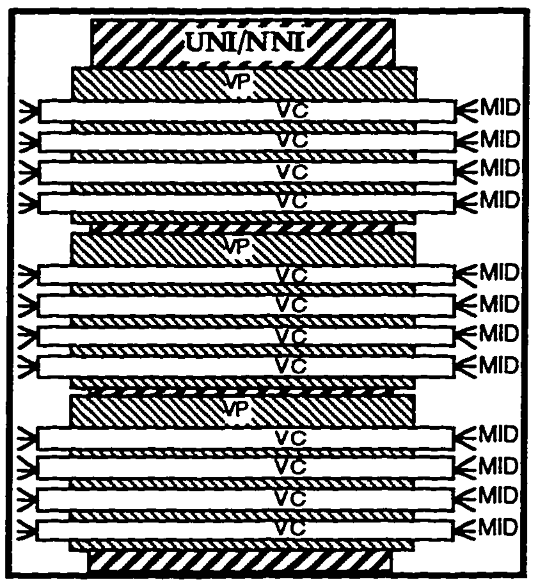


Figure 3: MID Submultiplexing on VCCs

It seems likely, however, that bandwidth will become a limited resource before connection address space does. Simply dividing transmission system bandwidth by the number of connections which may be supported via VPI/VCI yields a bandwidth granularity which is small enough to support expected applications.

It can also be argued that by the time transmission systems reach much higher speeds (thus eliminating this restriction of limited average bandwidth per VPI/VCI) switched

[1] MIDs were originally conceived as Message IDentifiers in DQDB for interleaving frames. Frames from a variety of sources to a variety of destinations could re-use the same MID at different times. It is equally possible, however, to use the MID field to identify a connection which has a fixed end-to-end source-destination relationship over time. MID is defined as a Multiplexing ID by CCITT.

services will reduce the need for large numbers of pre-established connections (PVC) which tie up address space.

For example, to support FR trunking or interworking between FRS and B-ISDN CONS, each 28-bit VPI/VCI on an NNI would be equivalent to a DLCI. If this address space is flat and may be flexibly allocated to connections it would appear that you will run out of bandwidth before you run out of connection identifiers.

Interface	VPI/VCI field size	Average Bandwidth Available Per Connection
OC3 NNI	28-bit	0.5 bit/s
OC3 UNI	24-bit	9 bit/s
OC12 UNI	24-bit	36 bit/s
ANSI DS3 UNI	20-bit	43 bit/s
ANSI OC3 UNI	20-bit	143 bit/s

The address space is not flat however, but hierarchical as shown below.

CCITT

- UNI - VPI/VCI size is negotiable up to 24 bits, divided 8/16 bits.
- NNI - VPI/VCI size is negotiable up to 28 bits, divided 12/16 bits.

ANSI

- UNI - VPI/VCI size is currently being debated. The CCITT values above are likely to be adopted. Previously ANSI has been considering restricting UNI VPI/VCI to 20 bits and possibly one of: 4/16 bits VPI/VCI or8/12 bits VPI/VCI
- NNI - VPI/VCI size is 28 bits.

Interestingly, this introduces an address allocation inflexibility problem analogous to TDMs bandwidth allocation inflexibility. 4,096 VPI exist on an NNI. If each VPC has some end-to-end significance across the network (i.e.: supporting a FRS virtual NNI/trunk) then the VCI within that VP are only available to that end-to-end VP user. Unused VCI on one VP are not available to other users of the NNI trunk (analogous to TDMs unused bandwidth on one channel not being available to other channels).

However, at OC12 rates with 4,096 VPs, approximately 146 Kbit/s average is available per VP if all VPs were used. Intuitively this does not seem a very high bandwidth for supporting, for example, a FRS VP-based virtual NNI. Again, bandwidth seems more a limit than address space.

A 10-bit MID facilitates sub-multiplexing up to 1,024 connections (or 1,024 cell interleaved datagrams) on a single Virtual Channel Connection. If the MID is 9-bits with 1-bit priority, there can be up to 512 connections or simultaneous cell-interleaved frames on a VCC. While this does provide more total connection space, the application requiring this space is yet to be identified.

Furthermore, the threat of VPs becoming the dominant scarce resource remains. This latter problem has resulted in proposals in [1] which proposes the use of a single VP across the UNI to support a Broadband Virtual Private Network rather than N-1 VP to form part of a fully-connected mesh.

9.5.2. MID Multiplexing & Multipoint Services

The AAL will be used to support multipoint data services (see Appendix B). ATM-level multipoint connections will support the AAL in this respect. One particular ATM connection type, multipoint-to-point or *Merge* poses unique problems for frame reassembly which make these connections:

- complicated to support with AAL-3/4 and
- impossible to support with AAL-5 without higher level session control (which is impractical).

Multipoint services may be provided utilizing a variety of underlying unidirectional connection types including: Point-to-Point (Unicast), Point-to-Multipoint (Multicast) and Multipoint-to-Point (Merge). These connection types are shown schematically below.

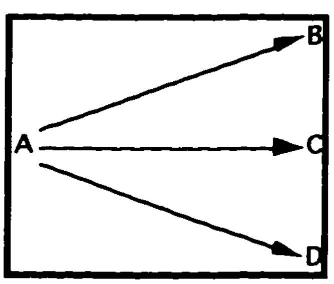


Figure 4: Three Unicast Connections

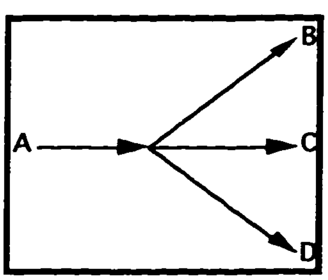


Figure 5: One Multicast Connection

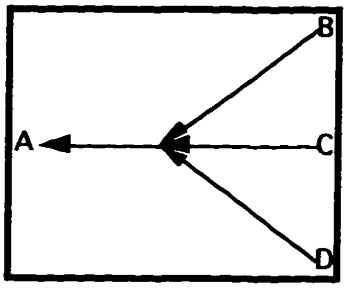


Figure 6: One Merge Connection

Arrows in the above diagrams represent virtual connections (i.e.: VC or VP). The topology of the arrows is not intended to imply a certain implementation. For example, the replication for the multicast may occur at a single point (switch) as implied or it may occur via a hierarchy of copies in a spanning tree. The key characteristics of the connection types are unchanged from the user (A, B, C & D) perspective.

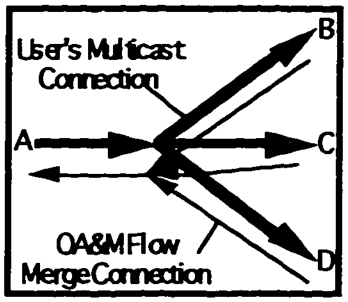
In the Unicast case, A sends on a unique connection identifier to reach a specific destination, and those destinations can uniquely identify the source A. With Multicast, A can not identify a single destination at the connection level but the destinations may be certain of the source. With Merge, the sources do uniquely identify the destination A, but A has no way of knowing the unique source.

The first two connection types (Unicast and Multicast) may be supported equally well by either AAL.

The third connection type (Merge) is the source of addressing/multiplexing difficulties which make it complex to support with AAL-3/4 and impractical to support with AAL-5.

If sources B, C & D are independently segmenting and sending variable length frames to A, there will be times when frames from multiple sources are cell interleaved when they arrive at A. Some form of per-cell sub-multiplexing on the connection is required for A to demultiplex these frames for reassembly.

OA&M Flows

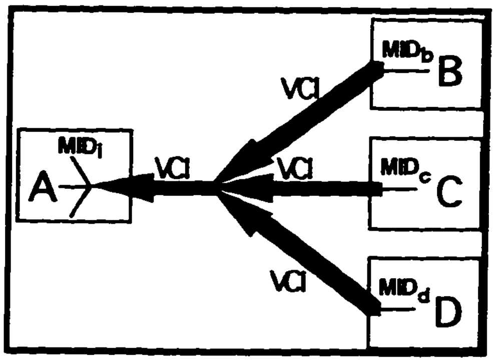


One possible requirement to support multipoint-to-point Merge VCs is for OA&M cell flows on the return path of a point-to-multipoint Multicast VC. This may not be a demultiplexing problem for OA&M messages which are a single cell in length (although interpreting echoed PINGs or FERF messages from multiple leaves at the root of a multicast tree may prove complex). There will be a demultiplexing requirement for longer OA&M messages (i.e.: AAL frame Ping which can occur end-to-end) which demands a per-cell multiplexing field.

AAL-3/4 can utilize its MID for this purpose as discussed below in i) and ii).

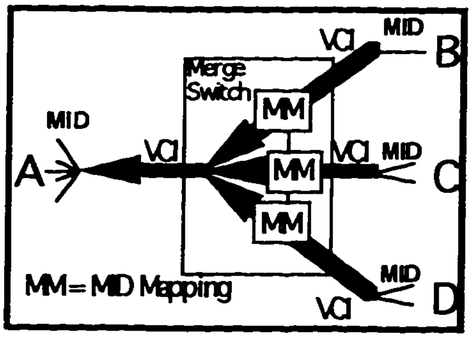
AAL-5 does not have a MID-like (per-cell) sub-multiplexing ID and hence Merge VCC can not be utilized. Two alternative solutions to this problem, equally applicable to either AAL, have been identified and are described in iii) and iv). Solutions involving some form of session control (such as token passing) to ensure that only one source station is transmitting at a given time are considered impractical and not discussed.

i) AAL-3/4 & MID Allocation



With AAL-3/4, unique MIDs can be allocated to each source terminal on the Merge Virtual Channel Connection. This will enable the receiving terminal to correctly identify and demultiplex cells from different frames for reassembly. This is precisely what is done for multi-CPE DQDB access to SMDS since all traffic is transmitted on a single VCI. DQDB has a simple, distributed MID allocation procedure operating at the PLCP-level to ensure MID uniqueness for each frame. MID allocation for ATM merge connections would require development of some other layer management procedure for MID allocation. The CPI field in AAL-3/4 has been proposed to indicate MID allocation frames.

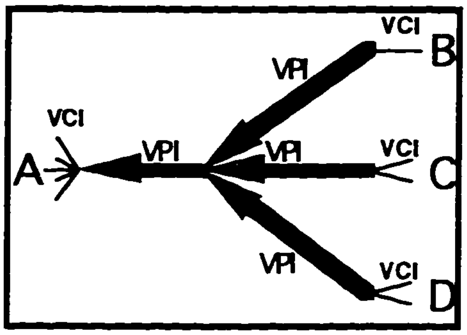
ii) AAL-3/4 & MID Mapping



A conceivable alternative to the MID allocation requirement above is to map MIDs (to new MID values) within the network at any merge point. This leaves the source free to independently select MIDs. In addition to performing a VCI translation at the switch where the merging occurs, MID uniqueness on the egress VC can be ensured via MID mapping. This would require that processing be done at the AAL SAR-PDU level and involve recalculation of the CRC-10. It would also require that MIDs be interpreted as

Message IDs that be randomly mapped to any single value for a given frame, not as end-to-end connection identifiers.

iii) Merge VP with Either AAL



A VP could be utilized to provide a multipoint-to-point Merge connection. The ATM switch(es) performing the merge need only process the VPI field. Unique VCI can be allocated to each source terminal on the Merge VPC to facilitate demultiplexing at the receive terminal. The largest problem with this proposal is that VPI are a scarce resource, especially at the UNI. If one was dedicated to the merge connection, address space would soon be exhausted.

iv) No Merge Type Connections with Either AAL

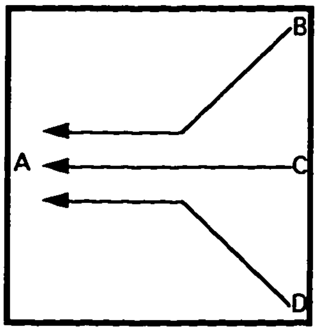
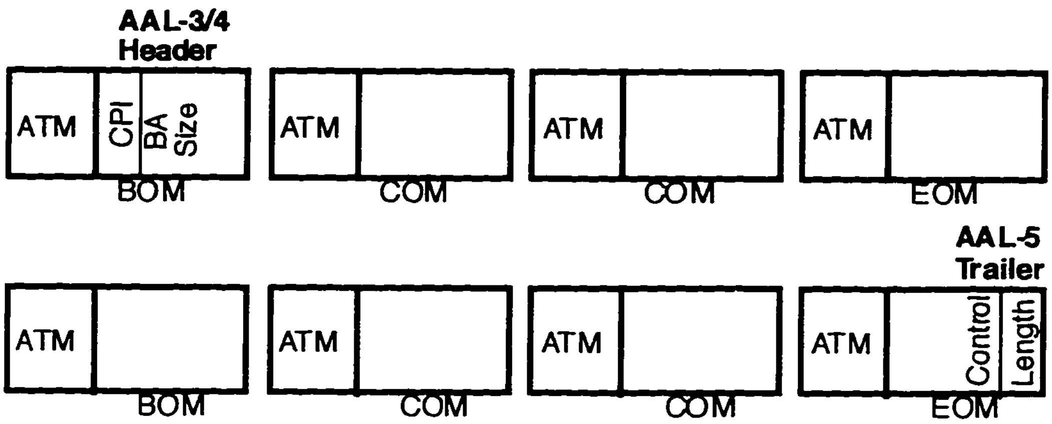


Figure 7: Multipoint-to-Point (No Merge)

The suggestion above to use unique VC for each source in the merge connection is really equivalent to eliminating the idea of a merge connection (at least at the VC level). Traffic will, of course, need to merge in the physical transmission system at the egress UNI but need not merge onto the same connection identifiers. Merge connections can be replaced by "N-1" Unicast connections as shown below. The primary negative implication of this choice is that the destination terminal needs to manage more egress VCI.

9.5.3. Location of Header/Trailer in BOM or EOM



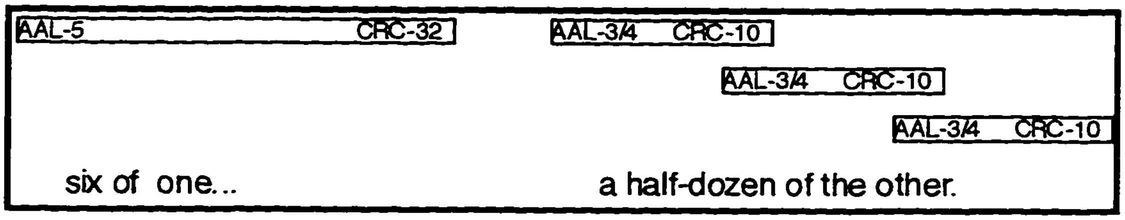
One clear difference between AAL-3/4 and AAL-5 is the absence of a header containing Length and AAL Control information in AAL-5.

The absence of a length indication header field in AAL-5 eliminates the possibility of terminals being able to dynamically allocate reassembly buffers unless they process some upper layer protocol encapsulated within the AAL frame that provides length information in its header.

AAL-3/4 management frames may be indicated via code points in the CPI field of the AAL-3/4 header. This supports functions such as a frame-level PING for continuity and performance testing and MID allocation. The ATM-level Payload Type Indicator would treat segments of these frames as normal user data.

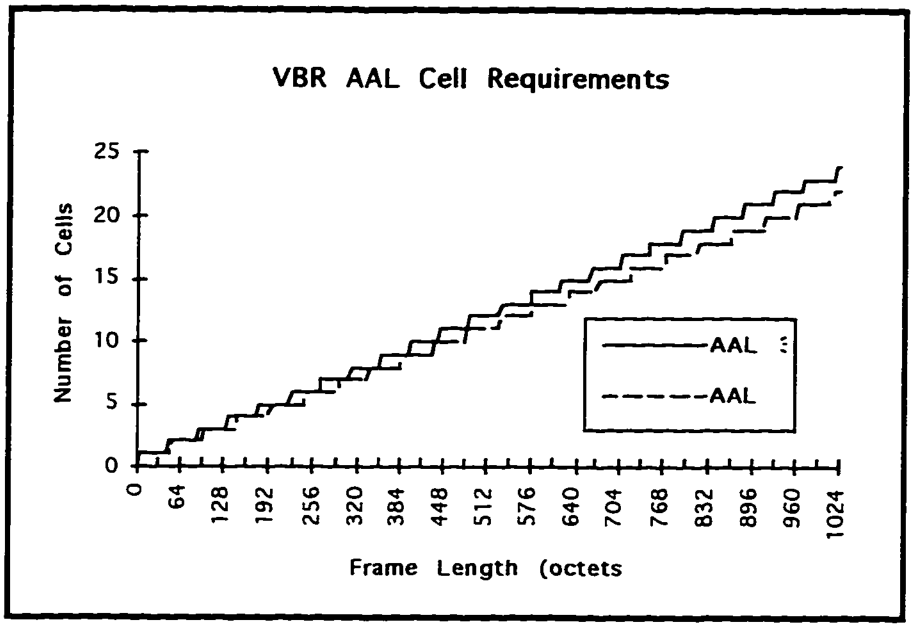
AAL-5 as currently defined does not include frame-level management functions such as PING. Presuming these functions were adopted for AAL-5 and the CPI-equivalent code points were contained in the AAL-5 trailer Control field, these frames could not be identified as being different from user data until the AAL-5 End-Of-Message cell was received. This is not an issue if OA&M frames of this type are Single-Segment-Messages.

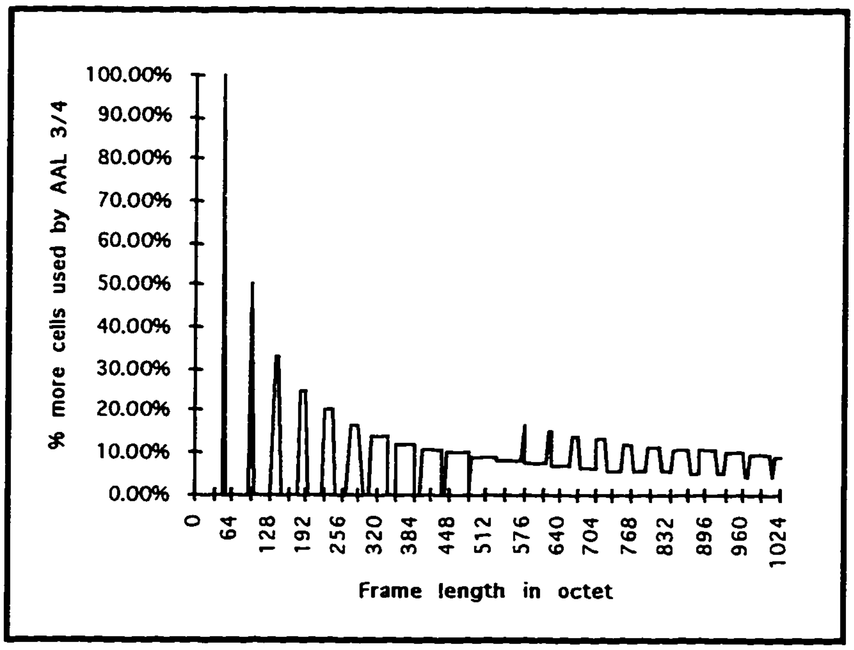
9.5.4. Error Robustness



Much of the discussion to date on the VBR AAL debate has focussed on error performance. However, depending on the error conditions chosen, either AAL-3/4 or SEAL can be shown to provide better error detection capabilities. For fiber there is no reason to choose one AAL over the other on the basis of robustness against errors.

9.5.5. Efficiency





The advantages of AAL-5 in this respect are clear at the cell level. AAL-3/4 utilizes 4 octets per 48 octet cell payload for the SAR-PDU header & trailer. AAL-3/4 thus consumes 8.3% of the cell payload capacity for SAR control information.

Differences at the frame level between the overhead of the AAL-3/4 header & trailer versus the AAL-5 trailer are, relatively speaking, insignificant. Given this as an assumption (the exact size of the AAL-5 trailer has yet to be agreed upon) then:

For frames shorter than 485 Bytes (max 484 octets = 11 cells), AAL-5 will require less cells than, or an equal number of cells to, AAL-3/4. For frames of length 485 octets or greater, AAL-5 will always require less total cells. In the limit for very long frames, AAL-5 would require 8.3% less cells.

9.5.6. Simplicity

It's argued in the SUN proposal [6] that AAL-5 can be implemented keeping only 4 octets of additional state information per VC for CRC-32 above what is required for either AAL to keep track of current buffer address, etc. SEAL eliminates processing for MID fields, Sequence Numbers and BE Tags. MIDs in particular are criticized for greatly increasing the amount of interface logic and complexity in MID translation and allocation. Comments from development groups on this aspect of comparison is requested.

10. APPENDIX D - Routers : ARP, Inverse ARP, Reverse ARP

A known data application which is made more efficient via network multicast support in Address Resolution Protocols. The description below assume IP as the layer-3 protocol but presumably many protocols utilize similar functions.

Three types of ARPs are used on catanets of multiple LANs today, namely:

- ARP - "Where are you?"... requests a physical (MAC) address for a device when you know the IP address.
- Inverse ARP - "Who are you?"... requests a network (IP) address given a knowledge of a specific physical (MAC) address.
- Reverse ARP - "Who am I?"... requests assignment of an IP address to ones self

When routers are interconnected not by LANs but by a connection oriented ATM network, globally unique MAC addresses no longer exist. Instead, only Virtual Path/Channel Identifiers (VPI/VCI) exist as "physical" addresses. The VPI/VCI value is only locally unique and locally significant across a given interface.

Interconnected routers still need ARP, Inverse ARP and Reverse ARP. This may accomplished by:

ARP (Where are you?) is multicasted to other routers in a Closed User Group (Virtual Private Network). The Router recognizing the IP address as its own can give a single unicast response. The VPI/VCI on which this response is received is the physical address (i.e.: it can not be provided explicitly by the responding router).

Inverse ARP (Who Are You?) if multicasted to the group on a multicast VPI/VCI will solicit many responses (IP addresses back from all distant routers). If Inverse ARP is unicast selectively on point-to-point VPI/VCI each will yield a single response.

Reverse ARP (Who am I?) could be multicast at initialization time and only the distant router providing name service will respond.

10.1. Standard Service Definitions

Multipoint services are in the process of being defined in the ITU (formerly) CCITT (both I.361 ATM cell and X.6 packet), the Frame Relay Forum and in Bellcore FA-NWT-001110. These proposals include Unicast (pt-pt), Multicast (pt-mpt) and Merge (mpt-pt) connections in a multipoint call.

The three Frame Relay Forum Multicast service definitions are shown below. (Note that these definitions would perhaps be better named *Multipoint* services, and the term *Multicast* be reserved for certain connection types within a multipoint call.)

In these figures, arrows at the UNI represent unique connection identifiers (Virtual Path & Channel Identifiers in the case of ATM). The topology of connections shown within the network cloud (circle) does not reflect the physical point in the network where the multicast replication or merge multiplexing occurs. For example, for "One-Way Multicast", the replication of frames sent by A on its One-Way Multicast DLCI (OMdlci) needn't occur at the first network switch. Multicast replication could occur via a spanning tree topology distributed over several switches.

Typical networks providing Frame Relay Service today are HDLC based and process contiguous variable length frames, each identified by a DLCI unique to a physical transmission link. However, once the network evolves to ATM, the frames will be segmented into multiple ATM cells. Only the first cell of the frame, the Beginning of Message or BOM cell, will contain the DLCI.

Merge connections provided by ATM Virtual Channels will cell interleave frames. In order to associate these cells with the frame to which they belong, an identifier is required in every cell. A per-cell sub-multiplexing field is required. This is true even in the case where the same terminal is the source for multiple merging connections (as in One-Way Multicast) since the Unicast and Multicast connections which merge may experience differences in delay and ordering between them can not be ensured.

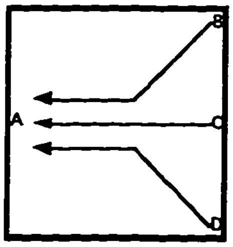
With AAL-3/4, unique MIDs can be allocated to each source terminal on the Merge connection. This will enable the receiving terminal to correctly identify and demultiplex cells from different frames for reassembly. AAL-3/4 is required for multi-CPE DQDB access to SMDS since all SMDS traffic is transmitted on a single VCI. If SMDS over B-ISDN supports multi-CPE via multipoint-to-point VC, AAL-3/4 & MID allocation are still required.

With AAL-5, this sub-multiplexing does not exist and Merge VCs can not be utilized. Three alternative solutions to this problem are described below. A third solution involves some form of session control (such as token passing) to ensure that only one source station is transmitting at a given time.

1) **Merge VP** : First of all, a VP could be utilized to provide a multipoint-to-point "merge" connection. The ATM switch(es) performing the merge need only process the VPI field. Unique VCI can be allocated to each source terminal on the "merge" VPC to facilitate demultiplexing at the receive terminal. This requires an algorithm and/or protocol to ensure that terminals are allocated unique VCIs. The largest problem with this proposal is that VPI are a relatively scarce resource, especially at the UNI where there are only 256 8-bit VPI. The number of simultaneous services that utilize VP will be constrained by these 8-bits in addition to the bandwidth of 34, 45, 150 or 600 Mbps.

2) **No Merge Type Connections** : The suggestion above to use unique point-to-point VC for each source in the merge connection is really equivalent to eliminating the idea of a merge connection. Traffic will, of course, need to merge in the physical transmission system at the egress UNI but need not merge onto the same connection identifiers. New multipoint service definitions are shown below which utilize only unidirectional point-to-multipoint (multicast) and point-to-point (unicast) connections within the multipoint call types.

3) **Session Control** : A third solution to cell interleaving on merge connections involves some form of session control (such as token passing) to ensure that only one source station is transmitting at a given time and hence frames will not be cell interleaved. This is equivalent to Media Access Control (MAC) on shared medium Local Area Networks (LAN).



Multipoint-to-Point Service Utilizing only Unicast Connections (No Merge)

Each of these new definitions maintains the connectivity of the previous three definitions. However, the new definitions make it explicit that at any UNI, unique connection IDs (VPI/VCI) will identify the source of any egress traffic.

The main implication of this is a higher demand on the VPI/VCI address space in the egress direction at the UNI. For example, for N-Way multipoint call, N-1 egress VCI are required rather than 1. If, in general, each member of a multicast group requires both:

a) to be able to reach the group via One-Way Multicast and

b) to be able to selectively unicast to each member

The connection topology would then consist of N Unidirectional Multicast connections (see N-Way Multicast) plus Nx(N-1)/2 Bidirectional Unicast connections.

Thus, each terminal must support the following unidirectional connections:

- 2x(N-1) receive (egress) connections of which (N-1) are unicast and (N-1) are multicast and all uniquely identify the source
- and N transmit (ingress) connections, one of which is a Multicast connection and (N-1) of which are.

This level of interconnection would be appropriate for peer Routers within a routing domain.

To support applications such as ARP, a router would need to associate multiple (2) receive connection IDs (one multicast and one unicast VPI/VCI) with a single physical address (unicast transmit VPI/VCI).

11. Glossary of Abbreviations Used

AAL5	ATM Adaptation Layer type 5, used for carrying variable length data frames over ATM's fixed length cells.
ATM	Asynchronous Transfer Mode: The fast packet switching technology that is the basis of B-ISDN, a growing number of Intranets and the core of the Internet.
AV	Audio Visual or Audio and Video
B-ISDN	Broadband ISDN. The true ISDN based on ATM.
BOM	Beginning of Message: A cell identified as containing the start of a higher layer data frame.
CIF	Common Interchange Format: the highest resolution image used in H.320 conferencing.
CODEC	Audio or Video COder/DECoder responsible for digitization and compression of temporal (time continuous) media
COM	Continuation of Message: A cell identified as containing a intermidiary portion of a higher layer data frame.
DCT	Discrete Cosine Transform: The signal processing technique that is the basis of JPEG and MPEG I-frame compression.
EOM	End of Message: A cell identified as containing the end of a higher layer data frame.
fps	Video frames per second.
H.261	The video coding standard used in Px64 video systems.
H.3xx	A suite of video conferencing standards from the ITU-T for operation over various network technologies.
I-MUX	Inverse Multiplexer: Equipment used to aggregate multiple 56/64 kbps channels into a single wideband channel. Eliminates differential delay between subchannels by adding delay to some. Enables "dial-up" video conferencing when connected to a switched digital service such as ISDN. Permits only point to point operation unless connected to an MCU.
IETF	Internet Engineering Task Force: Technical body responsible for generating specifications to evolve the Internet.
ISDN	Integrated Services Digital Network
JPEG	Joint Photographic Expert Group of the ISO/IEC responsible for the popular DCT based image compression standard.
MBONE	Multicast Backbone on the Internet: The subset of the Internet which interconnects Multicast Routers via IP tunnels to achieve multicast applications such as the broadcasting of IETF sessions.

MCU	Multipoint Control Unit: centralized server for Multipoint Video Conferencing responsible for selecting, switching, mixing and multicasting incoming audio, video, data and control streams from/to participants in a multipoint conference.
MPEG	Motion Picture Experts Group of the ISO/IEC responsible for the MPEG 1 and 2 digital video compression standards.
MTU	Maximum Transmission Unit: The maximum packet size that can be handled by all intermediate systems between source and sink. For ATM and unlike IP, only the capabilities of the terminals determines MTU size.
OA&M	Operations, Administration and Maintenance (aka Network Management) systems or operations.
OC12	A fiber based transmission system operating at roughly 600 Mbps.
PCM	Pulse Code Modulation.
POTS	Plain Old Telephone Service
Px64	Video conferencing systems based on H.320 standards that aggregate "P" 64 kbps channels to get sufficient bandwidth for audio, video, data and control.
QCIF	Quarter CIF.
QoS	Quality of Service: Performance characteristics of a network transfer service. Can be measure at various levels in a system. For example, transfer delay and delay variation, loss probability for the "reserved" bandwidth are important measures at the network level. Fidelity (frame rate, resolution and audio quality) and delay are useful measures of QoS at the video conferencing application level.
RSVP	Resource Reservation Setup Protocol: planned signalling for the new Internet, namely, the Integrated Services Packet Network.
SEAL	The Simple Efficient Adaptation Layer was an early rhetorical name for what became AAL-5.
TDM	Time Division Multiplexed: adjective referring to networks providing digital channels in increments of 56 or 64 kbps. The networking technology used to support most voice telephony today (e.g. POTS and ISDN).
VC	An ATM Virtual Channel connection is identifed in every cell flowing across an ATM network and allows the network to route the cell to its destination.
VCI	VC Identifier.

VoD	Video on Demand: Network Service providing users access to a menu of available videos to select and initiate playing.
VP	An ATM Virtual Path. A VP connection contains a bundle of 64K Vcs.
VPI	VP Indentifier.

12. References

[1] Beyond Teleconferencing Toward Enhanced Collaboration, SR-598, Institute for the Future, April 1996.

[2] The North American Video Equipment Market, YankeeWatch, Vol.9, No.10, September 1994.

[3] Future of Videoconferencing and Dataconferencing, Forward Concepts Electronics Market Research, December 1995.

[4] U.S. Videoconferencing Systems and Services Markets, Frost and Sullivan, 1996.

[5] ATM RendezView: Enabling Real-Time Multimedia to-the-desktop, Master's Thesis, Keith Smith, University of Ottawa, 1997.

[6] A Secure Audio Teleconference System, David Steer, Leo Strawczynski, Whitfield Diffie & Michael Wiener, BNR, Lecture Notes in Computer Science - Advances in Cryptology, Crypto '88, Springer-Verlang

[7] An Architectural Comparison of ST-II and RSVP, Danny Mitzel: USC/Hughes, Deborah Estrin: USC, Scott Shenker and Lixia Zhang: Xerox PARC,10 pages, February 1996

[8] An ISDN Multimedia Conference Bridge, D.H.Horn, V.M.Shaw, IEEE TENCON'90, p. 853-6 vol. 2, IEEE Region 10 Conference on Computer and Communication Systems, Hong Kong, September 1990.

[9] ATM Forum, UNI v4, October 1995.

[10] ATM Forum: Basline Text for Multiprotocol over ATM (MPOA), ATMF/95-0824r5, January 1996

[11] ATM Forum: LAN Emulation over ATM Specification - Version 1.0, ATM Forum/94-0035r9, January 1996

[12] ATM Forum: Traffic Management Specification, Version 4.0, 61 pages, February 1996

[13] Bellcore Multipoint Multimedia Conferencing Control Unit, GR-1337-CORE, Bellcore, 1994.

[14] Bellcore FA-NWT-001110, Key Topics Proposed for B-ISDN Requirements, Section 4: Multipoint Network Connectivity.

[15] Can ATM Really Support Video?, L.Lang, Telephony, January 1995.

[16] Data and Computer Communications, Fourth Edition, W.Stallings, Macmillan Publishing Company, 1994.

[17] Desktop Videoconferencing: Not Ready for Prime Time, K.Taylor, K.Tolly, Data Communications [TE37], April 1995.

[18] Expanded Lineup of ISDN Multipoint Audio Conferencing System, K.Murata, K.Shukushima, NTT Review, Vol.7 No.2, March 1995.

[19] Flat Networks Position Paper, Peter Ashton, Eric Livermore, Pierre Goyer and Scott Smith, Nortel Technology, June 1996

[20] FORE Systems Inc., SPANS UNI: Simple Protocol for ATM Network Signalling, Release 3.0, ForeThought Partners Adapter Documentation, MANU0031 - Rev. B, Mar. 1995

[21] Frame Relay Forum FRF 92.07, Frame Relay Multicast Service Description Issues.

[22] IETF: A Framework for Supporting RSVP Flows over ATM Networks, Onvural and Srinivasan, Draft RFC, 1996.

[23] IETF: Experimental Internet Stream Protocol, Version 2 (ST-II), RFC 1190, October 1990

[24] IETF: Integrated Services in the Internet Architecture: an Overview, IETF RFC 1633, June 1994

[25] IETF: Inter-Domain Multicast Routing (IDMR), Core Based Trees (CBT) Multicast Architecture, Internet Draft, February 9th, 1996

[26] IETF: Internet Protocol, Version 6 (Ipv6) Specification, RFC 1883, December 1995

[27] IETF: IP over ATM: A Framework Document, Internet Draft, April 1995

[28] IETF: Multicast Transport Protocol, RFC 1301, February 1992

[29] IETF: Native ATM Support in the Internet, RFC XXXX, Steve Jackowski: NetManage, Inc.

[30] IETF: Requirements for Multicast Protocols, RFC 1458, May 1993

[31] IETF: Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification, Internet Draft, 101 pages, February 21, 1996

[32] ITU Draft Recommendation H.24i, Interworking of H.series Multimedia Terminals December 19, 1996,

[33] ITU Draft Recommendation X.6, Packet Multicast Service Definition.

[34] ITU Recommendation I.361, B-ISDN ATM Layer Specification, March 1993.

[35] ITU Recommendation I.363, B-ISDN Adaptation Layer (AAL) Specification, April 1994.

[36] ITU Recommendation G.722: 7KHz Audio-Coding within 64Kbit/s, AP IX-142-E.

[37] ITU Recommendation Q.2931, B-ISDN User to Network Signalling, October 1994

[38] ITU Recommendations H.261, H.221, H231, H.243 & H.200, Working Party XV/1 - Audiovisual, Study Group XV , March 1993

[39] Multimedia Conferencing Systems for the Research Community, P.T.Kirstein, Information Networks and Data Communication (C-23), 1994.

[40] Multimedia Telecommunications Markets, Frost & Sullivan, June 1994.

[41] Multimedia: Computing, Communications & Applications, R.Steinmetz, K.Nahrstedt, Prentice-Hall PTR, 1995.

[42] Multipoint Video: Where's the Service, E.M.Gold, Data Communications, May 1994.

[43] Networking Constraints in Multimedia Conferencing and the Role of ATM Networks, R.R.Roy, AT&T Technical Journal, July/August 1994.

[44] New Directions in Videoconferencing, R.Shuchat, Link Resources Corporation, Filing Information, Vol.1, May 1995.

[45] Service Description of Communication Systems Supporting Multimedia Multi-user Applications, G.j.Heijenk, X.Hou, I.G.Niemegeers, IEEE Network special issue on Distributed Systems for Telecommunications, June 1993.

[46] Sharing the Cost of Multicast Trees: An Axiomatic Analysis, S. Herzog: USC, S. Shenker: Xerox PARC, D. Estrin: USC, ACM SIGCOMM '95 Conference, 15 pages, Augsut 1995

[47] T1S1.5/91-292, Simple Efficient Adaptation Layer (SEAL) Contribution, Sun Microsystems.

[48] The U.S. Data Communications Market, Data Communications, p.73-75, December 1994.

[49] Video Compression Techniques for Multimedia Communications, Toffman, Müller and Vogt, Electrical Communication, 1993.

[50] Video Development Environment Reference Guide, Parallax Graphics Inc., Part No. 917-026526 Rev. c, 1994

[51] X and Motif Quick Reference Guide, X Window System Version 11, Release 5, OSF/Motif 1.2, Randi J. Rost, Digital Press, 1993