Multimedia Conferencing: from prototype to national pilot

Mark J. Handley
Prof. Steve. R. Wilbur

Department of Computer Science
University College London
Gower Street
London WC1E 6BT, UK
M.Handley@cs.ucl.ac.uk
S.Wilbur@cs.ucl.ac.uk

Abstract

At University College we have developed a multimedia conferencing system as part of the CAR project funded by the European Commission RACE Program. This conferencing system is aimed at assisting distributed collaborative design in the automotive industry. It runs over basic rate ISDN lines, with video and audio provided by integrated video switching and ISDN video codecs. This paper describes some of our experiences with wide area multimedia conferencing, both from the workstation based approach and from the point of view of interconnecting studios and large conferences using digital networks. We are currently involved in the planning of the SuperJanet broadband network which will enable routine use of multimedia services between research and academic institutions in the United Kingdom. We describe some of the services that it will be possible to provide, both for workstation and studio based multimedia services, and the necessary local infrastructure that will be required.

Introduction

With the advent of ISDN, practical wide area multimedia conferencing has become possible at reasonable costs. One such prototype system is the CAR multimedia conferencing developed at UCL under the RACE programme, and aimed at collaborative design in the automotive industry. We will describe aspects of this system, and some of the implications of the design decisions taken.

ISDN however does not provide a flexible or efficient communications medium for large scale conferencing, and high bandwidth is still expensive. The JANET research and academic network in the UK is soon to be superseded by a new high bandwidth network dubbed SuperJanet. When this becomes available, applications such as CAR conferencing will have a new and much more powerful network, and we will be able to be provide multimedia services rather than the experiments that today's research network can support. As SuperJanet will support ATM technology, and high speed multicast protocols, more efficient ways to implement these services can be found. We will describe some of the possible ways to implement large scale desktop conferences on a network such as SuperJanet, and the implications these have for the video compression technology, shared window systems and the multicast protocols themselves.

The CAR Project

At University College London we have developed a multimedia conferencing system as part of the CAR project (CAD/CAM for the Automotive industry under RACE (Research for Advanced Communications in Europe)). The CAR system is a multimedia conference management and support system. It is aimed at assisting automotive designers at different sites to interactively discuss design decisions in order to reduce the development time of new cars.

The conferencing system performs the following services:

- A directory of authorised users, multimedia stations, video and audio peripherals.
- A conference server, which keeps track of active conferences and their participants.
- A switching service, controlling local analogue video and audio switching and mixing.
- A conference control application, which provides a user interface to transparently access the above services, and provides floor control and application control services.
- The integration of a shared window system which allows existing applications to be shared in a conference.

- Support for new conference aware applications.

Applications

Applications which can be included in a CAR conference can be divided into two categories - collaboration aware and collaboration transparent.

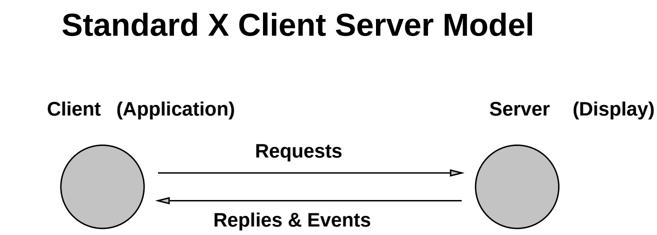
Collaboration aware applications are those which have been written specifically for the conferencing system. An example is a distributed sketchpad developed for CAR by FhG-IGD at Darmstadt, Germany. Such an application receives asynchronous notifications (similar to callbacks in the X window system) informing it of events such as change of floor holder. It can obtain information about the conference through calls to the CAR library, which uses ANSA RPC to contact the conference server.

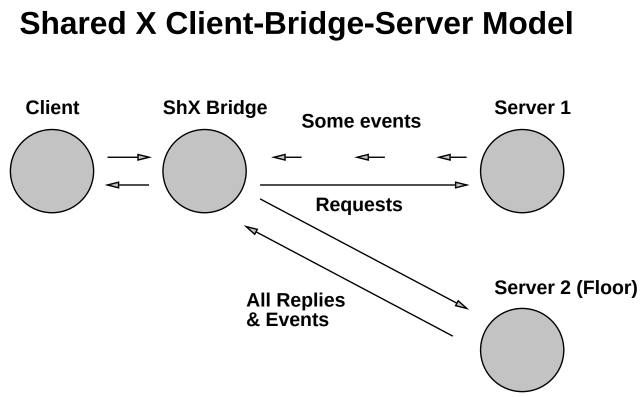
Collaboration transparent applications are existing X window system applications which can be shared in a conference using a shared window system (based on Shared-X, developed under the CEC Nestor project at DEC Karlsruhe). This allows us to include important applications such as CAD packages, which would require much work to rewrite them specifically for the conference environment. The shared window system is integrated into the conferencing system, and collaboration transparent applications are subject to the same conference mechanisms as collaboration aware applications.

Shared-X

The standard client-server distributed window system model of computing implemented in the X Window System consists of a server process which controls a workstation screen, and client (application) processes which require their output to be displayed. The application sends requests to the server such as “Draw me a line”, and the server sends back a reply to the application to confirm what happened. To make the model interactive, the server can send events to the application such as “The user just clicked the mouse button at position ...”. Applications can choose which types of events they are interested in, and so don’t get informed about irrelevant events. These requests, replies and events are all performed or controlled by calls to functions in Xlib, the X library. The application process and the X server can be on the same machine, but they don’t have to be - the Xlib functions just need to be told which display to use on which machine.

Shared-X was intended to aid remote teaching by allowing the replication of the display and control of an X application on more than one display. To this end, a version of Xlib was developed which keeps a copy of all the state information pertaining to an application that is normally stored inaccessible in the X server, and permits its duplication to other X servers. Programs can be recompiled using the Shared X library instead of the standard Xlib,





but more typically the duplication is performed using a Shared X bridge, which converts standard Xlib requests into Xlib requests for each server. This bridge process sits between the client application and the server(s). It looks like a server to the client, and like a client to all the servers.

When the Shared-X bridge receives a request from the client, it keeps track of any resources that would be created in the server, and the passes the request on to the servers. Replies and events coming from the servers to the client are filtered, as the application only expects to hear from one server, and hence a floor control regime is imposed which only allows user input events from one server through to the application at any time.

One problem with Shared-X is that it does not provide any form of distributed window management. Thus it is possible to have different sized windows on different displays, and for one person to be working in a part of a window that is not visible to other users. A form of distributed window manager would be a desirable extension.

Another major drawback is Shared-X’s centralised design, which greatly affects performance when replicating to a large group of displays. The communications implications of this centralised design are discussed later.

Video

At UCL we have a digitally switched analogue video network connecting several offices and laboratories. The video switch is controlled by software running on a UNIX workstation, and is transparently integrated into CAR multimedia conferences. The conference video channel

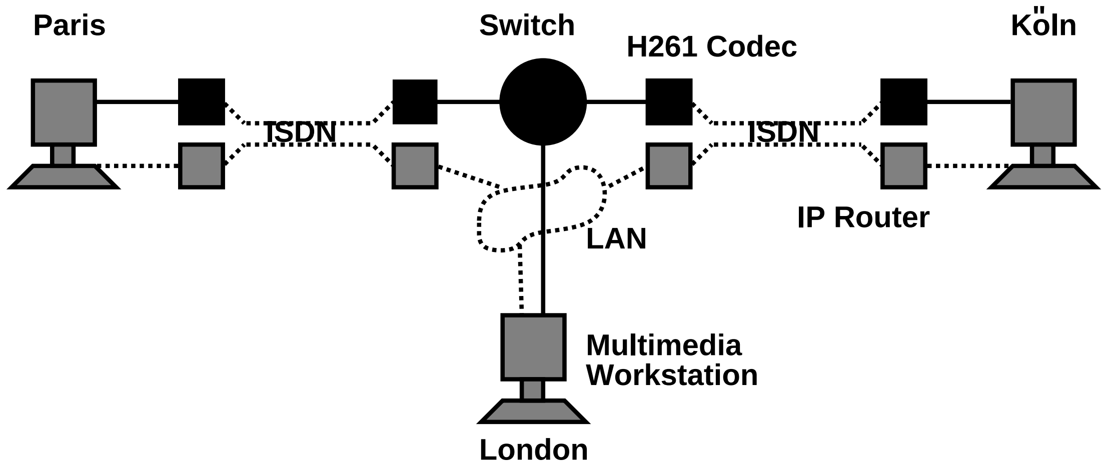


Figure 1: Hub and spoke arrangement for CAR ISDN multimedia conference trials

is controlled by the floor holder through the conferencing software, and switches automatically when there is a change of floor. This degree of integration ensures that users view the video and applications as being one conference, rather than separate. The floor holder can also select quad mixing, which allows four conference sites to be visible to everybody, which is especially useful for small interactive conferences.

The audio channel for a conference is open-channel, and so anyone can speak at any time. To prevent feedback, we employ an n-1 audio mixer, which returns mixed audio to a site from all the other sites, but not to the originator.

CAR user trials show that in a collaborative design conference, much of the user's attention is taken up by the design applications. There tend to be quite long periods when a designer is concentrating on a design task, and often he does not explain verbally what he is trying to do. At these times, a video link is invaluable to find out what the other participant(s) are doing, and whether they expect you to be doing something. For this purpose high quality video is really not necessary. However for larger lecture style conferences, if video is required, greater bandwidth would be desirable.

Our findings indicate that many users (especially designers) do not wish to sacrifice screen space for a video window, however small. For the purposes of interactive design conferences, we also find that a frontal head and shoulders view does not include enough contextual information. Much better appears to be a view from one side, which allows the remote site to see the user from the waist upwards. Ideally it should include the workstation keyboard, so the remote site can clearly see when the user is busy. To create an illusion of co-presence, the camera and video display must be situated close together. This enables a user to talk to the *image* of the remote site, and feel he is being talked to. The implication of this is that many users will not require the video to be displayed on the workstation screen, but rather on a separate monitor situated away to one side of the desk.

Using analogue video for local distribution provides us with maximum flexibility, and very high quality at relatively low cost. However it is clearly not cost effective to use analogue video for wide area distribution, and the problems of scaling switch control are certainly not trivial. For wide area distribution we employ both proprietary Picturetel and H261 video codecs. As video codecs are currently expensive it makes sense to use them as a shared resource, and hence connecting them to our analogue video switch rather than to an individual workstation is a cost-effective solution.

We are currently preparing for the second phase of trials of the CAR system with designers from Ford, Peugeot SA, and IAD Ltd. This will involve sites in Germany, France, and the UK. To connect these sites we employ ISDN connections, and prototype British Telecom H261 video codes. ISDN will provide 64Kb/s for the applications data, and 64Kb/s for the combined video and audio data. For these trials UCL will feature as a hub site to allow us to perform video switching and quad mixing. Whilst we appreciate that 48Kb/s video is of rather low quality, the results of the first phase of CAR user trials suggest that due to the uses the video is put to, this will not be a problem. However the delay of passing through two pairs of codecs, each of which adds a 400ms delay means that we abandon any attempts to get synchronisation between the audio which is very important, and the video which is less time critical.

Although car designers value the real estate of the workstation screen, it may prove much more cost effective to display video there. We estimate that on workstations with greater than 100 MIPS processing power, H261 video compression will be possible in software without expensive hardware assistance. This is supported by findings at INRIA (Huitema, 1992) on current top end DEC workstations. Suitable high power workstations should be available within 18 months. As a result, digital multimedia will become much more widely available, which will have far reaching effects for tomorrow's networks.

Digital Conferences and Teaching

In March 1992 as an experiment, audio from the IETF conference in San Diego was successfully multicast over the internet to many sites in the US, and others worldwide, including the UK. This was done using ISI's *vt* (voice teleconferencing tool) and Lawrence Berkeley Laboratories *vat* (visual audio tool).

At UCL we have been experimenting with multicasting both audio (using vat) and slow frame rate video using our own software. To date we have multicast several conferences from the London University (Livenet) video network to sites in the UK including Cambridge and Edinburgh over JIPS (JANET Internet Protocol Service). As we cannot assume any special hardware at the receiving sites, and as JIPS is heavily used and of restricted bandwidth, we have been restricting video to one frame per second with no special compression, but this proves to be quite adequate to create a feeling of cohesion and, most importantly, to disseminate any slides. Given more network bandwidth, there appear to be few problems increasing this rate to about 8 fps, which is the maximum rate we can obtain from the Sun VideoPix digitiser board we are using.

An alternative approach might be to make the slides available before the seminar for file transfer, which being lower priority, has less effect on network traffic, and to provide synchronisation data to display each slide at the correct time. Unfortunately this eliminates the potential benefits of multicast to reduce total network traffic, and loses the immediacy that even low rate video can provide.

Although these experiments have shown the potential for conference dissemination, there is currently not enough bandwidth available in the UK for this multimedia traffic to be considered as a regular service. It does however provide an important stepping stone towards integrated multimedia services. These experiments also spread awareness of the potential uses of this technology, build a potential user group, and thus steer the development of the support technology and of the networks themselves.

Towards SuperJanet

Plans are being laid for a new UK academic network to provide a basis for high bandwidth and novel applications including multimedia teaching and conferencing. The original JANET X25 network was one of the first European national academic and research networks. Over the years it has grown in size and performance to the point where some 200 UK sites are connected, with 25% of sites connected at 2 Mb/s.

With the growth of workstation based systems, it was decided to provide an IP overlay on the JANET network, ie. IP packets were tunnelled over JANET. Approximately one third of the 2 Mb/s backbone is used for this JIPS (Janet IP Service) traffic. Since it became available in October 1991, JIPS traffic has seen explosive growth.

Projecting the existing demands forward, data services alone will probably need 34 Mb/s trunks before the end of the decade. In addition there will be potentially exponential growth in demand for multimedia services as increased processing power drastically reduces the specialist hardware required. Clearly a new network is required if this potential is to be exploited.

To this end, the SuperJanet Project was founded. Its aims are:

- from an early stage, to provide higher capacity conventional data communications on a service basis allowing existing use to grow naturally;
- to provide a testbed for users to develop new applications which only really become feasible with significant bandwidth;
- to allow the communications and systems research community to use part of the network for experiments in high performance communication;
- to allow collaborating companies from European industry access to a large-scale testbed for novel communications and applications.

The current thinking is to use SDH (Synchronous Data Hierarchy) technology, with perhaps three rings covering roughly the south, middle and north of the UK. Each ring might operate at 622 Mb/s initially with cross connects between the rings. Connections to user sites would probably be at 155 Mb/s initially. Although in many ways ATM (Asynchronous Transfer Mode) standards would offer an ideally flexible way of sharing bandwidth, it seems likely that products and standards will be too immature for the initial network. However, an ATM pilot service is planned for early in 1993 with multimedia applications.

Multimedia Services and SuperJanet

Eventually isochronous multimedia services such as voice and video will be provided over ATM, but initially it is likely that 2 Mb/s circuits will be set up between H261 codecs, or packetised and sent over the packet networks.

The first uses of these services will be to provide cost effective connections between the existing local video networks and studios such as the University of London Livenet, and similar networks at the Universities of Ulster and Wales (C5C). These sites already have video teaching rooms in place. The cost of suitable codecs to connect these sites is currently about $50000, which is not expensive when compared with the cost of equipping a studio.

We have been running a pilot video conferencing service between the studio at UCL and various sites in the US over the 384 kb/s UK-US IP link. Currently we are using 128 kb/s for video and 64 kb/s for audio, which provides excellent quality sound, and acceptable picture quality. However this link is obviously limited to one conference at a time. With the advent of SuperJanet, it will be possible for this sort of conference with higher quality video to become relatively commonplace within the UK.

Connecting studios in this way presents few technical problems as the video is not computer mediated, and it can be safely assumed that all sites are likely to have high bandwidth network connections. It greatly enhances the uses of these local video networks. It also serves to build a user base for multimedia services, which will them drive the development of smaller scale systems, including workstation based conferencing.

Towards desktop conferencing

With the advent of SuperJanet in the UK, and particularly with an ATM service, we will have a network that can provide the connectivity, flexibility and bandwidth to support multimedia services employing this technology. Connecting two homogeneous multimedia workstations together for video and audio is not technically difficult, and many such products will rapidly become available. However, conferences with more than two participants cause more problems. Even on high bandwidth networks we wish to avoid having to send separately to all participants, and to this end multicast appears to provide a desirable solution, as it avoids traffic duplication where possible. However, any real large scale conference user group will contain a variety of equipment and variation of bandwidth of connections to the network backbone. This leads us to a dilemma: we wish to send the same data stream to everyone (multicast), but different users have different bandwidth and facilities for decompression. We really do not want to have to take the lowest common denominator, and send low quality to everyone, just because a few sites have poor connections.

One approach is to code the audio and video data using a hierarchical encoding scheme, in which higher levels of the hierarchy provide additional quality. Thus the most significant part of the signal is sent to all receiving sites, and the higher levels of the hierarchy are only forwarded to sites which can cope with them. This also has the advantage that the low levels of encoding are the more important for intelligibility, and thus we have a priority scheme to ensure that the most important packets get through if a network becomes congested. Wakeman & Crowcroft (1992) suggest several ways these objectives can be achieved.

Wide Area Conferencing

As we move from local conferences and simple ISDN hub and spoke topologies towards true wide area high speed networks such as SuperJanet and ATM with large numbers of large scale conferences, some of the limitations under which current conferencing systems such as CAR were developed no longer apply. Facilities such as multicast can make more efficient use of these networks and reduce response times. Many lessons can be learnt from the existing systems which should affect the way we design systems for tomorrow's networks, and determine the services these networks should provide.

Conference Support

Directory service

Conference directory services are required to find equipment, resource and user details, so that these may be integrated into the conference control software. Clearly this directory information should be distributed so that each site holds information about its own resources. To satisfy these requirements, we intend to replace the existing CAR directory service with one based on the OSI X500 directory standard. However if we intend to use the same A/V equipment with other software, a standard for specifying the actual information would be desirable.

Conference server

Conference servers provide a means to discover details about conferences currently in progress. A centralised architecture in not sufficiently fault tolerant, and clearly does not scale. However this information may be too timely and transient to allow us to use X500 or similar "heavyweight" standards. A lightweight protocol is required, which does not need to guarantee consistency. As the same information will be frequently required by many sites, this protocol should make use of multicast protocols with caching, which should allow for elegant scaling.

Video and Audio

As the CAR project has been based around the control of an analogue video switch with ISDN as a communication medium, we opted for a centralised hub and spoke arrangement for the distribution and switching of video (see figure 1). However as we progress to fully digital video and broadband packet/cell networks, other topologies have benefits:

- Hub and spoke
    - The simplest solution for connection based networks such as ISDN

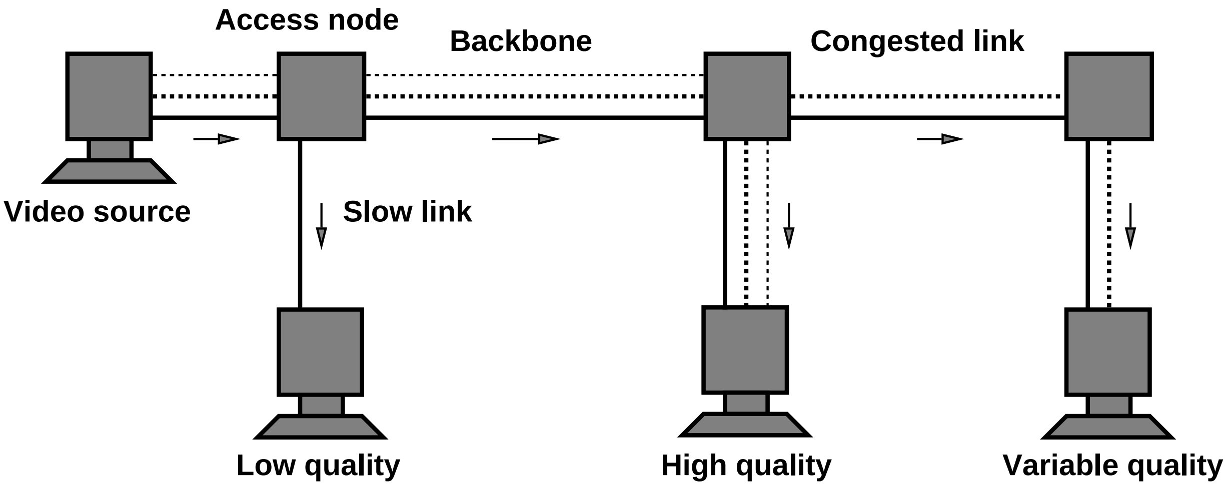


Figure 2: Hierarchical encoding of video and audio for wide area multicast

- Less good over ATM: we can multicast from the hub, but unless the hub is on the network backbone, multicasting from source is likely to be more efficient.
- Good for single channel conferences (the floor holder controls who sees what). We only need to send a source to the hub when the channel is switched to that source
- Less good for multi-channel conferences (the receiver controls what they see), as the number of sources sending to the hub simultaneously may be large.

• Peer group with multicast

- Good for single and multi-channel conferences.
- Needs multicast to minimise scaling problems.
- Don't want everyone sending to everyone else continuously, so we need a fast way to select/deselect forwarding. As we already know the members of the group, group leave and join would be inefficient, and restrict scalability.

Hybrid topologies are also possible, but the same basic issues apply.

The different topologies and technologies each have implications for video switching, and the compression algorithms used:

• Switched analogue

- Simple for local systems
- Switch control does not scale
- Expensive over wide areas

• Switched analogue hub with digital spokes

- Simple
- Allows standards conversion - can put several codecs on one switch
- Only really feasible for single channel conferences. Multiple channels require one codec per spoke.
- The video must go through two pairs of codecs. At low data rates each pair may add a delay of around 400ms, which makes communication between spoke sites difficult.

• Digital hub with single channel video

- Simple
- May be a delay whilst the video stream from the source to the hub sets up.
- It is important for the video to switch immediately so the floor holder gets confirmation of the switch, but the other users must be informed that the break in signal is intended.
- The video source codec must generate a synchronisation point immediately the call is set up. With some codecs this may be a problem.

• Digital hub with multiple channel video

- This is not a good solution - there is a risk that the hub's network connections will be swamped. A peer group system does not have this bottleneck, but is a more complex control problem.
- Each time we add a receiver to a source, we must tell the source to generate a resynchronisation point immediately.

• Peer Group multicast with single channel video

- This is more network efficient than equivalent hub systems.

  - Due to unusual routings, it may be difficult to ensure that the old source ceases multicast before the new source begins. On limited bandwidth links to some receivers this may cause problems.
  - The new source must send a resynchronisation point, but only after we are sure that this will not be lost due to the end of the stream from the previous source.
  - We need a fast efficient way to select/deselect multicast (without actually leaving the multicast group), but this is relatively simple, as it only needs to be an on/off selection at source.

- Peer Group multicast with multiple channel video
  - This is the most complex, and the most flexible system.
  - It requires a fast efficient way to select/deselect multicast, but for this topology we need the routers to respond to the select so we can add/remove a receiver on a multicast route.
  - The source codecs must be capable of producing a resynchronisation point every time that source is reselected by a receiver.

IP Multicast

The Internet Protocol multicast routing system distributes traffic across a minimum spanning tree of all the networks and routers where there are members of a multicast group. To join a multicast group, a process issues a *group join*, which is propagated globally.

This scheme is scalable, and works well for such traffic as audio conferences (cf. *vat*). It makes an assumption that the group is symmetric in that no member of the group is more important than the others, and thus any member of the group can multicast to the group. For audio this is a reasonable assumption, as silence suppression decides which group member sends. If more than one member sends, mixing or filtering can always be performed at the receiver.

However for video we do not want more than one site sending simultaneously to a multicast group, and when there are many potential sources, it is the *receiver* that decides which one to receive from. What is required is for members of a conference to join a group at the start of the conference, but to be able to select from time to time whether they wish to be forwarded traffic from that group. We do not wish to use group join for this purpose as it is propagated globally, which is not necessary as we already know the group members.

What is needed is some traffic filter in the multicast routers and mechanism to send an activation message to them. This can be done in two ways:

- multicast an "activate group" control message to the group. As this is multicast it will only propagate to the existing group routers.
- send a message to the source requesting activation of the group. The source then propagates the "activate group" message along with the existing stream of video packets to the routers downstream.

As the source must also send a synchronisation point in the video for the new receiver, the latter method ensures that this arrives after the filtering of the multicast stream has been removed.

This filtering scheme could be part of a more general scheme for hierarchically encoded video as described earlier.

Applications

When we examine the communication of conference applications, it becomes clear that many collaboration aware applications lend themselves to the use of multicast protocols, and thus to efficient network usage in large conferences. However Shared-X which allows us to run collaboration transparent applications does not, and cannot be modified to do so.

Shared-X works by mapping requests and resource ids from a single X application onto requests and resource ids for multiple X servers. The problem with this model is that as the mapping is performed at the client, this results in a different request stream being sent to each X server, even though these request streams are supposed to perform the same task, (albeit on different resources such as the colourmap).

The X protocol makes the complete distribution of this resource mapping difficult, as clearly only one consistent reply and event stream must reach the client. However if the request and resource mapping were performed in a pseudo server associated with each real X server, then the requests from the client can be multicast to the pseudo servers. To do this, the pseudo servers must implement a virtual X space, similar in concept to virtual memory. For instance such a pseudo server would map a virtual colourmap (which is the same for the client and all pseudo servers) onto the logical colourmap at its X server.

X heterogeneity

Heterogeneity within the X window system can cause shared window systems a problem at two levels:

- X servers within a conference may have different display types, sizes, or depths of colourmap and so on.
- even homogeneous hardware may have different resources available due to other applications already running.

For collaboration transparency this can cause resource mapping problems. For example if our conference includes several 24 and 8 bit plane displays and one 1 bit (black & white) display, what display type do we tell the application it is connected to? Assuming the lowest common denominator causes the least mapping problems, but may prove too restrictive for some applications. Worse still, what do we do if the user of the 1 bit display joins the conference part way through? The only good course of action is for the user who starts the application to decide depending on the application and the expected user (and hence hardware) group. In some cases it may be necessary to specify minimal resources, and a user whose system does not meet these requirements receives either an incomplete display or no display at all.

It is clear that the X window system was not designed to allow display replication, and as a result attempting to design extensions or additions to do so elegantly is not trivial. However allowing collaboration transparent applications is extremely desirable, as we cannot expect many useful applications to be rewritten in a collaboration aware manner.

Conclusions

Multimedia applications are now starting to become available, initially with local area networks but increasingly with ISDN and other connection based networks in mind. Although conference systems designed for these networks will run over the next generation of networks such as SuperJanet and other ATM networks, for the most part they do not take advantage of the increased flexibility, and potential for multicast that these networks can provide. However to make full use of such networks requires extensions of existing protocols. In particular, the current implementation of IP multicast does not elegantly support video switching, as it makes assumptions about the nature of multicast traffic which are not entirely justified for video traffic.

The X window system is now sufficiently well established and supported that it is a natural base on which to build conferencing systems. However, for shared applications it does not elegantly support collaboration transparent replication of displays. Shared window extensions such as Shared-X do permit display replication, but at the cost of a centralised architecture, which prevents the use of multicast or very large conference groups. It should be possible to design a decentralised architecture for a shared X window system, though ideally this should be included in the X server itself.

Wide area multimedia conferencing is still in its infancy, and although their are standards such as H261 for video compression, they do not map well onto heterogeneous receivers on different bandwidth networks. Such standards clearly need to be developed. There are still many aspects of multimedia conferencing for which no standards exist. However for some of these it is clearly still too early to be setting such standards, and much more experimentation and experience is required.

Acknowledgements

We wish to thank Jon Crowcroft and Ian Wakeman for some of the ideas presented in this paper. We would also like to thank all our colleagues in the CAR project who are involved in the development of the conferencing system.

This work has been funded by the CAR project of the CEC RACE Programme and Department of Computer Science, UCL

This paper is dedicated to the ruins of Llanthony Abbey in the Black Mountains of South Wales.

References

[1] Deering, S. & Cheriton, D. E. "A Multicast Extension to the Internet Protocol", RFC-966, Dec 1985

[2] Huitema, C. "Software codecs and work-station video conferences", Proceeding on INET '92, June 1992

[3] Shacham, N. "Multipoint Communication by Hierarchically Encoded Data", Proceedings of INFOCOM92, May 1992

[4] Wakeman, I. & Crowcroft, J. "Note on the use of hierarchical encoding in ameliorating the problems of dissemination of audio visual data over catenets of varying bandwidth", UCL Internal Note IN/92/3

[5] Wilbur, S. "The UK National Research Network: From Janet to SuperJanet", Proceedings of IEEE Eurocon '92, May 1992