Session X: Communication III

Chair: Duane Northcutt, SUN Microsystems Laboratories

The issues related to system support for digital audio and video can only be properly explored within the context of actual systems; it is rarely possible to effectively study a single aspect of the system support problem (e.g., process scheduling, buffer management, network control, etc.) in isolation. This is because of the interactions between the individual resource management activities that must be performed in the end-to-end process of communicating digital audio and video. Therefore, it is necessary to construct complete systems in order to validate proposed solutions and achieve the needed level of understanding of the relationships among various system-level trade-offs. All of the papers in this session describe work that has gained benefits from such empirical systems engineering efforts. These papers address various communications-related aspects of the general problem of transporting digital audio and video across computer communications networks.

The focus of the first paper is on the problems faced in delivering interactive compound documents (which include a rich variety of information types, including continuous media) to end-users via a digital communications network. The paper defines an interesting and technically challenging application area, and describes various difficulties encountered in the authors' attempt at emulating the computing environment needed to support such applications. The described work is intended to serve as a stepping stone for more ambitious experiments with continuous-media distribution and presentation. The second paper describes a CCITT standard protocol designed to meet the needs of videophone teleconferencing. This protocol is aimed at limited-bandwidth transport services and has provisions for interoperability among videophone units with differing capabilities. The author also indicates that his research institute has implemented working prototypes of videophones which use this protocol. In the third paper, key aspects of the design and implementation of a professional-grade digital audio system is described. This all-digital audio system is based on FDDI network technology and makes use of techniques for the synchronization of multiple independent streams of digital audio. The fourth and final paper describes a valuable technique for making effective use of the synchronous mode of transfer in FDDI systems for the transport of digital video information. The proposed network management technique is aimed at minimizing the total amount of buffering required by the system, while ensuring that the time constraints of digital video are met. Analytical techniques are used to prove that the proposed technique provides the desired properties, and work is reportedly underway to validate this work through empirical means.

In his presentation of a paper he wrote together with Gil Cruz and Thomas Judd, "Presenting Multimedia Documents Over a Digital Network," Jonathan Rosenberg of Bellcore described the general problem of distributing interactive compound documents to a large community of users over limited bandwidth connections. He defined the type of information to be distributed to be composed of a wide variety of different information types (e.g., text, graphics, audio, still images, video, synthetic imagery, etc.), that is stored at some remote server site, and accessed interactively by a (potentially large) set of (potentially widely geographically distributed) independent users. The transport service for the delivery of this information to a single user was assumed to be limited in terms of its available bandwidth, in that it would provide much less than is required to transmit a single stream of uncompressed video. Specifically, the assertion was made that the "last mile" problem would persist for the next few decades, and therefore the authors assume a data rate on the order of 1 Mbps.

Jonathan also described the three main areas that are being addressed in this work, and the approach they are taking on each of them. These key areas are:

- Media compression – a direct implication of the limited bandwidth assumption, and a means of effectively expanding the amount of transport bandwidth that is available. Their plan is to make use of existing hardware and software technology for reducing the bandwidth demands of media transport.
- Structural transmission – a means of making more effective use of the limited transport bandwidth by taking advantage of the structure inherent in compound documents and segmenting them into their (independent) constituent components and transmitting them separately (according to their individual requirements). The techniques being explored involve new partitionings of applications between source and destination.
- Time shifting – another technique for the efficient utilization of the limited transport bandwidth, that also uses the structure of the presentation material to predict access patterns and distribute the transport of desired information over a longer period of time (thereby reducing instantaneous bandwidth demands for the transfer of specific information items). In effect, this amounts to the pre-fetching of information.

Jonathan went on to describe the experimental (PC- and Ethernet-based) system that has been developed as part of project DEMON (Delivery of Electronic Multimedia Over the Network) at Bellcore. He presented a system architecture for an initial prototype that embodied three simplifying assumptions: single source/destination document delivery, circuit-switched network connections, and that the machine at the destination is dedicated to document presentation. Jonathan presented a taped demonstration of some of the initial results of this effort, which took the form of an interactive travel guide to Seattle. This demonstration showed a large (fairly static) image as a background, with a number of text, graphic, and image windows being dynamically created, deleted, and moved on the screen. The overall screen resolution was reported to be 900×675 pixels, with 16 bits/pixel, and a video rate of eight frames/second was shown.

Jonathan concluded by saying that this exercise had proved to his group's satisfaction that a practical system that can support their desired class of applications can be built. He went on to say that this experience has served to reinforce the notion that some form of hardware assistance will be needed to properly support such applications. In particular, he noted that the delivered video frame rate was almost half that predicted based on the theoretical hardware throughput figures.

An interesting discussion followed Jonathan's presentation, which centered on the question of regulatory restrictions on the ability of common telecommunications carriers to provide the type of services described in this presentation. Jonathan noted that while some restrictions have recently been lifted, there are still a large number of regulations and regulatory agencies to be dealt with, so the question is far from being resolved.

In response to questions regarding the assumptions of limited bandwidth to a wide client base, Jonathan noted that field trials are underway with digital service at rates of 1.5 Mbps into homes and 16 Kbps back out. However, this is being done over copper and fiber is a long way off. He noted that New Jersey Bell has estimated that it will take 30 years to completely replace their 56 Million miles of copper with fiber.

The paper "The CCITT Communication Protocol for Videophone Teleconferencing Equipment" was presented by Ralf Hinz of Daimler-Benz AG Ulm, Germany. It describes the recently defined protocol for videophone conferencing developed by CCITT Study Group XV. This protocol supports from one to six channels of synchronized audio, video, and data, all on a single 64 Kbps (ISDN-B) channel. While the protocol is designed to be network-independent, the intention is that this protocol should work with existing, low bandwidth (i.e., from 64 Kbps to 2 Mbps) digital communications paths (and not ATM or gigabit networks).

Ralf provided an overview of the CCITT recommendations that are applicable to this protocol (which included H.261 as a possible means of compression), and described the key features of the protocol. These include: in-band signaling, support for (de)multiplexing of streams, support for both point-to-point and multi-point connection topologies, and allowances for both centralized and distributed network synchronization mechanisms. Furthermore, the protocol encodes eight sub-channels in each transmitted octet; on basic rate ISDN channels, this provides 8 Kbps per subchannel. In addition, a connection can be composed of up to six ISDN-B channels, so channel numbering has been added to allow for the synchronization of multiple sub-channels from different ISDN channels.

The protocol Ralf described includes a means for endpoint device capability exchange during the setup phase among potential communicating partners. This exchange permits the negotiation of the type of service to be used for the new connection. As part of the negotiation process, each station in the desired connection indicates its capabilities – e.g., number of frames per second of video it can source/sink, audio/video data formats supported, type of compression, type of codec, etc. In addition, the communications mode for a connection can be renegotiated during the course of a session. This capability permits a single communication stream to be used alternatively to send video and audio, FAX, or data, without having to acquire additional capacity or re-establish the connection.

Ralf went on to indicate that, in order to gain experience in the area of videophone conferencing, his research institute has produced working prototypes of videophones that use this protocol. These videophones incorporate the necessary codec's and line interface units, and include mechanisms to balance out differences in delays between different paths (e.g., the audio and the video in one channel of a teleconference).

It was noted that by considering data rates of up to 2 Mbps, it would be possible to use something other than the CCITT standard compression scheme (i.e., H.261) – for example, MPEG.

In addition, questions were raised about the error recovery capabilities of the protocol, especially with respect to the transfer of compressed video. Ralf pointed out that the protocol has a number of features for error recovery, including synchronization markers for video frames that will allow recovery following data loss.

Udo Zölzer presented in "FDDI-Based Digital Audio Interfaces" joint work with N. Kalff at the Technical University of Hamburg-Harburg, Germany. He gave a description of a rather large and sophisticated digital audio system that is based on FDDI technology at the data link layer. This system provided the capability to acquire, process, and present multiple, concurrent streams of digital audio, as well as distribute them among a distributed suite of audio sources, sinks, and processing units. This system manages the distribution of multiple digital audio channels in a variety of different formats – i.e., one of three different sampling frequencies (48 kHz, 44.1 kHz, and 32 kHz), and two different transmission formats (AES/EBU and MADI).

As this is a specialized system, it consists of dedicated-function units that use a TDM bus to connect individual audio channels to the network, and use TAXI parts to interface to the fiber. Each functional unit's fiber connection leads to a distribution switch, and all functional units (as well as the switches) in the suite make use of a common reference clock (i.e., the "house clock"). These data paths contain pure audio data and all of the control functionality exists in separate controllers, connected by traditional LANs. There is no general purpose computing done within the functional units, and no control traffic coexists with the digital audio samples.

Because of the dedicated-function nature of the system and the existence of a common, global clock, inter-sample and intra-sample synchronization is quite straightforward. However, when asynchronous inputs are introduced into the system (i.e., audio streams that are not sampled with the house clock – e.g., from remote studios, vans, etc.), a more difficult form of synchronization is required. Asynchronous inputs must undergo a process of sample rate conversion upon introduction into the system. In addition, when multiple streams of audio are to be synchronized, there are two types of issues to be addressed – one, for the case where the audio streams were acquired at the same nominal sample rate, but have different phasing, and the other for the case when the audio streams were acquired at different sample rates. All of these synchronization issues are addressed in this system, by way of sophisticated DSP techniques for re-sampling of the data streams.

Given the approach to re-sampling used in this system involves a form of interpolation, a question was raised as to what amount of delay is induced in the process. Udo indicated that only a small amount (on the order of one sample time) of delay is introduced.

Another discussion centered on the question of the practicality of house clocks in a more general computer networking environment, and helped draw distinctions between the type of dedicated-function system described here and more general distributed computing environments.

A further discussion was held that attempted to establish the potential viability of implementing traditional DSP functions (such as re-sampling) with high-performance (i.e., high clock-speed, high instructions-per-clock), general-purpose processors, as opposed to DSPs. From the discussion, it became clear that while these algorithms are costly, increases in general-purpose processor performance should make it practical to implement such functions for small numbers of streams, but (as in this system) support for large numbers of streams will require more cost-effective dedicated-function solutions.

The final presentation of this session was given by Bernard Cousin of ENSERB-LaBRI, Bordeaux, France on "Digital Video Transmission and the FDDI Token Ring Protocol." He described a technique for effectively distributing digital video using the synchronous mode of transmission defined in the FDDI specification. The aim of this work is to simultaneously allow the time constraints of the video to be met and arrive at an effective balance between the amount of buffer space needed at sources and destinations and the overhead induced by the passing of the network access control token. In the scheme presented here, this is achieved by choosing the size of transmission units and send/receive buffers in concert with the bandwidth needs of the streams and the network's chosen token rotation time.

By controlling the amount of data that is sent at each opportunity (i.e., token rotation), it is possible to ensure the effective use of a minimal amount of buffering at both the source and destination. In his presentation, Bernard indicated that he has analytical proofs that the proposed transport control scheme can meet the temporal constraints of video streams with a given (minimum) delay bound.

Bernard also discussed the importance of properly choosing the token rotation time for the network – in effect, it should be fast enough to provide good interactivity (i.e., granularity of access) and minimize worst-case buffer size requirements, but not so fast as to induce significant network overhead (and thereby lower the effective bandwidth of the interconnect).

In conclusion, Bernard noted that his group is in the process of constructing a prototype network that uses this control method to transport video over FDDI.

Initial discussions revolved around the question of whether it would be possible to achieve similar delay bounds with FDDI's asynchronous mode. In the course of this discussion, it became clear that the proposed scheme relies quite heavily on the hardware-provided periodicity of token rotation in the synchronous mode of operation.

It was also noted that, while the FDDI specification includes the synchronous mode, most of the currently available implementations only support the asynchronous mode of operation. Bernard replied that this is exactly why they were compelled to construct their own FDDI network interface units.

Another discussion followed that centered on the notion that the proposed scheme involves a three-way trade-off between buffer sizes, token rotation time, and transmission unit sizes. Which of these variables are changed and which are given can differ based on the available degrees of freedom in a given system. For example, given a specific negotiated token rotation time and a given transmission unit size, this scheme can be used to determine the amount of buffering needed to meet the temporal demands of a given video stream.

Presenting Multimedia Documents Over a Digital Network

Jonathan Rosenberg
Gil Cruz
Thomas Judd

Bellcore
445 South Street
Morristown, NJ 07962-1910
USA

October 25, 1991

Abstract

This paper discusses an experimental prototype system for presenting integrated multimedia documents over a digital network. This prototype is the first in a series investigating the requirements placed on the network in support of applications presenting multimedia information. The information consists of multiple media in digital form, including multi-font text, geometric graphics, photographic images, audio and motion video.

The paper describes the motivation for this line of research and the initial focus and goals of our first prototype. We describe the hardware comprising this prototype and the current status of our efforts. This is followed by a discussion of some early results we have obtained in dealing with integrated digital media, including motion video, with off-the-shelf components. Finally, we draw some conclusions about required support for digital media.

1. Introduction

The Information Networks Research group at Bellcore is investigating the network requirements for supporting near-term (five to ten years from now) multimedia information applications. Many researchers believe that multimedia information applications will be important in business, educational and residential environments. Applications might include interactive training, remote classroom lessons and home information services.

We are initially investigating the network requirements of residential electronic multimedia information applications. The home is appealing because the public switched network provides an individually addressable, electronic connection to virtually every residence in the United States. On the other hand, the home is a challenge because the local access loop does not presently support the bandwidth required by multimedia applications, especially those utilizing photographic images or motion video.

Although there are many open research questions, our primary focus is on techniques to enable the presentation of high-quality multimedia documents over networks with *limited bandwidth.* We consider limited bandwidth to be any rate at least an order of magnitude less than required to present uncompressed full-motion video. As we will discuss later, we are initially focusing on a bit rate of 1.5 Mbits/sec.

We are investigating these network presentation techniques in a project known as Delivery of Electronic Multimedia Over the Network (DEMON). As part of DEMON, we will be building prototype applications that present interactive, multimedia documents over a network. The construction of complete end-to-end experimental prototypes will allow us to fully explore the requirements that this class of applications places on the network and display devices.

The documents we are interested in include multi-font text, graphics, photographic images, audio and motion video. We are concerned with the presentation of these media when represented digitally (which we will call *digital media*) over a single network providing a fixed bandwidth connection. We are not considering adjunct analog networks. This is appropriate because the local access loop of the public switched network is expected to evolve to provide digital transport.

Our early work on an experimental prototype application has shown that it is possible to present high-quality multimedia documents, including embedded motion video, over a network at 1.5 Mbits/sec. The relative ease with which we were able to accomplish this is due to two factors. First, we have identified practical techniques for presenting multimedia documents over limited bit-rate connections. Second, by constraining the bandwidth to 1.5 Mbits/sec we have largely avoided pushing against the limits of off-the-shelf hardware and software, including networks, CPUs, operating systems, compression hardware and graphics systems. Despite this, we appear to have hit the wall in dealing with low-rate digital video on current workstation architectures.

This paper is primarily a discussion of the experimental prototype system we are constructing and what we have learned about the network delivery and presentation of digital media. The next section provides a short introduction to our view of multimedia document presentation over networks: why such presentation is important, why we are initially focusing on 1.5 Mbits/sec, why we believe this work will have wide applicability and the techniques we have identified. This is followed by sections briefly describing the DEMON project, the prototype[1] and some early results and conclusions.

2. Document Presentation Over a Network

One might ask why we should worry about presenting multimedia documents over a network. After all, there are standalone platforms available that can present multimedia documents delivered from local storage media, such as CD-ROM, laser disc or even magnetic disc. For some applications, of course, these platforms will be perfectly suitable.

However, network document presentation will be essential when the timeliness of information or the cost of acquiring and storing information locally make the use of a standalone system infeasible. For example, access from homes to a large collection of multimedia encyclopedias may only be economically practical if the documents are stored remotely and information presented over the network on demand. Other applications include daily in-house multimedia

---

[1]It is the policy of Bellcore to avoid any statements of comparative analysis or evaluation of products or vendors. Any mention of products or vendors in this paper is done where necessary for the sake of scientific accuracy and precision, or for background information to a point of technology analysis, or to provide an example of a technology for illustrative purposes, and should not be construed as either positive or negative commentary on that product or that vendor. Neither the inclusion of a product or a vendor in this paper, nor the omission of a product or a vendor, should be interpreted as indicating a position or opinion of that product or vendor on the part of the authors or of Bellcore.

newsletters for geographically distributed businesses and personalized information packages delivered to homes. In these examples, the cost and time required for traditional publication could be prohibitive. Network presentation may be an attractive alternative.

A major roadblock in the way of network presentation of multimedia documents is the bandwidth required by some media. The presentation of uncompressed, NTSC-quality digital video, for example, requires in excess of 80 Mbits/sec.[2] There is much work in the development of high-speed transmission for the public switched network that promises to alleviate this bandwidth bottleneck. Unfortunately, networking capabilities with sufficient bandwidth to easily support presentation of multimedia documents are unlikely to be available to any significant number of residences before the year 2010 [9].

There have, however, been recent advances into attaining increased bit rates over the existing copper loop facilities, including Asymmetric Digital Subscriber Line (ADSL) technology [6]. This technology allows a 1.5 Mbits/sec channel to a customer with a simultaneous low bit-rate return channel (suitable for, at least, a voice connection), over most of the existing copper-loop plant.

The 1.5 Mbits/sec rate is also interesting because of recent standards activity in full-motion video compression. Two emerging schemes, MPEG [4] and H.261 [5], can compress full-motion video for transmission at 1.5 Mbits/sec. Of the two standards, we are considering only MPEG because it is designed to support the the manipulation of digtal video. Recent studies have indicated that MPEG-encoded entertainment video at 1.5 Mbits/sec may be nearly equivalent in quality to VHS video [2].

Despite our current focus on residential applications, we believe that our techniques will also be applicable in non-residential applications. In particular, some techniques for supporting multimedia applications over packet-switching networks operate by reserving resources to guarantee a minimum bit rate connection to an application [1]. In these systems, an application needs to reserve the peak bandwidth required by the document, thus wasting much of the bandwidth (and resources). Our techniques will allow an application to reserve a lower bandwidth, of which little will be wasted.

In addition, even if a workstation is attached to a high-speed network, only a small fraction of the total bandwidth will, in general, be available. For example, on an FDDI network providing total throughput of 100 Mbits/sec, an application might obtain a channel guaranteeing 14 Mbits/sec. This is still nowhere near the bandwidth required for uncompressed full-motion video. Furthermore, although network speeds will increase, we can expect users to never have enough bandwidth, much in the same way that users run out of disk space and RAM, no matter how much you give them.

We have argued that our techniques will be useful not just in residential document presentation, but also in presenting documents over packet-switching networks providing support for multimedia applications. In the remainder of this section, we will describe three techniques we will use in presenting interactive multimedia documents over networks with fixed moderate bit rates: *media compression*, *structural transmission* and *time shifting*.

[2]This figure assumes an uncompressed 480×640 pixel image with 8 bits of color per pixel [3].

Media compression makes use of medium-specific and human cognitive properties to compress bandwidth-intensive media like photographic images and full-motion video. JPEG [8], MPEG and H.261 are examples of media compression techniques. For our purposes, we view media compression as available technology.

Structural transmission makes use of the structure of a document to save transmission bandwidth. Consider a presentation that has a fixed pixmap background upon which parts of the presentation will appear and disappear. Using structural transmission, the background is sent only once. The display machine understands the semantics of a background and will repaint parts of the pixmap as necessary, without the need for network communication. Another kind of structural transmission makes use of media semantics. For example, rather than using several thousand bytes to transmit a blue rectangle as a pixmap, it is possible to send several bytes representing a request to draw that rectangle and color it blue.

Time shifting makes use of periods of low channel utilization to deliver some media ahead of their presentation times. An example of this is the playing of CD-quality audio during a video presentation over a 1.5 Mbits/sec network channel. Using time shifting, the audio would be delivered before it was required during a part of the presentation that left spare channel capacity.

The use of structural transmission breaks a document into its constituent media and logical components. This tends to create peaks and valleys in the bandwidth requirement of the document as its presentation progresses. Media compression is used to squash down some of the peaks. Time shifting chops off peaks that are still too high and places them in earlier valleys. The result of these techniques is to reduce and smooth the bandwidth requirements of a document to make better use of a fixed bandwidth allocation.

3. The DEMON Project

The primary goal of the DEMON project is to study the network requirements of applications for presenting interactive multimedia information. Our initial focus is on the presentation of personalized, interactive information packages to residences. The packages will consist of articles chosen according to a subscriber's interests. The articles will contain little text, consisting primarily of photographic images, full-motion video, audio and geometric graphics. Like television, the articles in DEMON will be highly dependent on temporal media, such as audio and motion video.

Because we work in the context of end-to-end applications, our research will also involve areas besides network document presentation. We are also investigating interfaces for residential users and authoring tools to ease the task of creating multimedia articles. More details on DEMON — including motivation for its user model, the system architecture and examples of articles — can be found elsewhere [7].

4. The First DEMON Prototype

We are planning on building a series of experimental prototype systems exploring and demonstrating issues related to network presentation of multimedia documents. The first prototype, described in this section, is constrained to simplify many aspects of the problem. Future prototypes will explore increasingly complex models of network document presentation, including those required by non-residential applications.

For the first prototype, we are restricting our attention to the presentation of documents from an Information Provider[3] machine to a single residence machine over a circuit-switched connection with ADSL characteristics: 1.5 Mbits/sec available to the customer and a low-bandwidth return channel. This architecture is illustrated in Figure 1.

This architecture makes several simplifications to the general problem of network document presentation:

- *Documents are delivered from a single source machine to a single destination machine.* An economically realistic scenario might require that machines support multiple connections.
- *The network connection is circuit switched.* In many environments, a packet-switching network is the only network available.
- *The destination machine is dedicated to the task of document presentation.* In many environments (for example, the office), a workstation must perform many tasks simultaneously.

Note that these restrictions simplify the presentation of temporal multimedia documents by eliminating much of the resource contention typical in computer systems. We have effectively eliminated contention for network, operating system and some workstation resources, such as screen space. This was done in the interests of making some progress on this problem. We expected network document presentation even under these conditions to be rather difficult and we were determined to make our first step manageable.

In addition, because we are exploring new territory, the first prototype is being built using off-the-shelf hardware. The hardware comprising this prototype is illustrated in Figure 2. Although we are most comfortable with Unix[4] workstations, we have chosen a 486, 33 MHz PC running MS-DOS[5] as the receiving machine. We made this decision for several reasons:

- MS-DOS is a uniprocess operating system, and we could carefully control the scheduling of time-critical activities.
- There are a large variety of graphics systems available for PC platforms (in contrast to the meager choices available for Unix workstations). Our prototype is currently outfitted with a Matrox Image Series 1280[6], which provides powerful graphics and image processing capabilities and a flexibly configurable 8 MB of screen-mapped on-board RAM.
- Hardware support for compression invariably appears for PC platforms well before it is available on Unix workstations. The prototype incorporates hardware JPEG support in the form of a Squeeze-AT board[7] from Rapid Technology.

---

[3]*Information Provider* is our generic term for the entity responsible for the end-to-end delivery of documents.

[4]Unix is a registered trademark of AT&T.

[5]MS-DOS is a trademark of Microsoft.

[6]Matrox Image Series 1280 is a trademark of Matrox.

[7]Squeeze-AT is a trademark of Rapid Technology.

The PC, augmented with compression and graphics hardware, matches our functional view of the TV or information/entertainment appliance of the future. We expect this device to be like a TV, but with RAM and significant media formatting and networking capabilities.

We are running standard operating systems on both machines: SunOS 4.1 on the Sun workstation and MS-DOS 5.0 on the PC. At the moment, we are using TCP/IP as our transport protocol. The software for delivering documents and for displaying them was written by us.

The prototype is simulating a circuit-switched connection by making use of a dedicated Ethernet network. We have used Ethernet technology because it permits us to experiment with a range of bandwidths, expanding our research in both directions from 1.5 Mbits/sec. We will do this by installing a *throttle* (see Figure 2) in the network driver code on the Sun SPARCstation 2.[8] Applications will specify a maximum bit rate, which will be enforced by the throttle.

As discussed in the Introduction, one of our goals for DEMON is to investigate the integration of digital media. To us, this means treating all media, including temporal media, equivalently. This implies that all media (as far as feasible) utilize the same storage devices, network facilities and processing paths. We are, therefore, avoiding such typical tactics as utilizing an adjunct analog network for video transmission or installing hardware to allow video to bypass the operating system and be displayed directly on the monitor.

The first DEMON prototype is in an early stage of construction. We are, however, able to present some documents consisting of multi-font text, geometric graphics, photographic images and motion video (all in 16-bit color) over the Ethernet at approximately 1.5 Mbits/sec.[9]

Currently, the documents presented by the prototype have to be preprocessed by hand for effective network presentation. The initial software was written to demonstrate the general validity of our ideas and to gain some experience into our hardware infrastructure. The next implementation phase will generalize and automate the document presentation process.

Although we are early in the development of the first experimental prototype, we have learned some things about integrating digital media using conventional hardware and software. The next section provides some details about the media characteristics we are dealing with and discusses our experience so far.

5. Early Results from the DEMON Prototype

The first thing we considered for multimedia document presentation at 1.5 Mbits/sec was whether it would be satisfactory to treat an entire presentation as video and use MPEG to compress it. We did this to a short segment of one of our documents.[10] The MPEG process

---

[8]SPARCstation 2 is a trademark of Sun Microsystems.

[9]The throttle is not yet implemented, so we are measuring the attained bit rate, which is currently averaging near 1.5 Mbits/sec.

[10]We took the display output as video, digitized it and used a software system at Bellcore that simulates MPEG encoding and decoding.

produced what we characterize as a"fuzzy" version of the document segment. The motion video looked fine, but much of the text was blurry and difficult to read and the sharp lines comprising the geometric graphics were lost. We consider this quality unacceptable as it severely limits the amount and kinds of normation that can be usefully included in a document.

The documents presented in the prototype were preprocessed by hand to use media compression and structural transmission. Our use of media compression was limited to JPEG (discussed further below), but we found that these techniques were easy to use and produced the desired effect: we were able to present multimedia documents over a network at 1.5 Mbits/sec. The documents were equivalent in visual quality to those we can produce on a standalone workstation.

As mentioned previously, the documents we have been presenting contain multi-font text, geometric graphics, photographic images and motion video. As might be expected, we experienced several levels of difficulty in working with these media. We discuss our experiences with each class of media in the remainder of this section.

The simplest class of media to work with included multi-font text and geometric graphics. This comes as no great surprise since these media typically have low bandwidth requirements. Furthermore, our use of structural compression (sending a command to draw and color an object instead of sending the pixels, for example) lowered the requirements even further. In addition, the powerful capabilities of our graphics system meant that little processing was required by the PC upon receipt of the information.

Photographic images required more effort, although they were still relatively simple to transmit effectively. We have a JPEG board that provided us with flexible input and output. The board allows us to specify arbitrary routines to direct both (compressed) input and (uncompressed) output, so that we are able to feed the board from anywhere and send the output to anywhere. This flexibility is important because it allowed us complete freedom in the graphics systems and peripherals we can use. There were faster boards available but they attain this speed at the cost of being tied (via hardware) to a particular graphics system, and none of these graphics systems provided the power and resolution we needed. Of course, the flexibility we obtained was bought at a cost, as we shall discuss below.

Our strategy for photographic images is to compress each image only as much as necessary to enable the image to be transmitted, decompressed and displayed according to the document presentation schedule. This strategy caused us a problem since it required us to be able to attain a specified compression ratio. Unfortunately, the JPEG board only allows the specification of a *Q factor* during the compression of a image. The Q factor, which ranges from 1 to 255, is a rough idea of how much compression should be applied.

In practice, the compression ratio obtained for a specific image at a given Q factor is highly dependent on the image itself. This meant that we had to try several Q factors for each image until we obtained a compression ratio near the one desired. This is an inherent property of the JPEG compression algorithm. It is interesting to note that we typically were able to use compression ratios of 25:1 without producing any objectionable degradation of an image.

It will be no surprise to hear that transmitting and displaying motion video gave us the most trouble. We are counting on the availability of hardware support for MPEG to allow us to achieve good-quality video at 1.5 Mbits/sec. Unfortunately, no MPEG-decoder boards are currently available for purchase.

We are, therefore, currently making do with JPEG. We decided to do this, rather than waiting for hardware MPEG support, in order to gain some experience in working with digital video. The idea is to compress individual frames off-line, then transmit, uncompress and display them as rapidly as possible. Given the uncertain availability of MPEG, this technique is beginning to appeal to many researchers, developers and manufacturers. Using this technique, we have been able to achieve reasonable-quality motion video with a displayed size of 900×675 pixels and a rate of approximately 6.5 frames/sec over Ethernet at 1.5 Mbits/sec. This has required some tricks, however, as we discuss below.[11]

First, some figures on the performance of the JPEG board and our PC. We can decompress a 300×225,[12] 16-bit image (compressed at approximately 17:1 to about 8 Kbytes) from RAM and deliver it to the graphics system in about 115 ms. This allows us to display about 8 frames/sec from RAM.

Experimentation has shown that we can decompress those same images in only 85 ms if we throw the resultant bits away instead of shipping them to the graphics system. The raw speed of the board should, therefore, allow us to display about 11 frames/sec, but we have not been able to attain that rate. Our unsubstantiated belief is that we are suffering the effects of bus contention. Consider that, for each frame we are transferring bits from memory to the JPEG board, from the board back to memory and then from memory to the graphics system.

If we analyze this process, we see that for each frame we are transferring 8 Kbytes (the size of a compressed frame) from memory to the JPEG board, 135 Kbytes (the size of an uncompressed frame) from the board to memory and 135 Kbytes from memory to the graphics system. This is close to 300 Kbytes for each frame. For a rate of 11 frames/sec, this would be about 3 Mbytes/sec. The graphics system will accept data at the rate of 4 Mbytes/sec, although 11 frames/sec would only require about 1.5 Mbytes/sec. We believe, therefore, that this bit rate is greater than the bus can handle due to contention. It is for exactly this reason that some JPEG boards bypass the system bus to achieve higher effective rates.

Although 8 frames/sec is fast enough to achieve reasonable motion, the frame size of 300×225 pixels is disturbingly small in our application.[13] We would really like something two to three times that size in each dimension. Unfortunately, if we use source images of these sizes, the frame rate slows to the point of unacceptability.

It also turns out that compressing the images at greater than 17:1, thus reducing their size in RAM, does not allow a higher frame rate with larger images. The greater compression clearly

---

[11]There are a series of tricks we use, but do not discuss here. These involve the way input video frames are converted to digital frames. For example, for a video source converted from film, we take every other field and double each field to form a frame. This avoids the blurry frames caused by the 3:2 pulldown during conversion from film.

[12]The 300×225 size is somewhat arbitrary, being a convenient size for the capture system we are using. In addition, it is in the 3:4 aspect ratio used for movies.

[13]We are displaying the documents on a large screen display (60" diagonal) with 1280×1000 pixel resolution.

must reduce the time required to transfer the images to the JPEG board. Unfortunately, the time required to decompress an image is essentially independent of the compression ratio, depending only on the resultant size of the image. Therefore, we reduce somewhat the time to transfer the compressed image to the board, but the decompression time and the time to transfer from the board to the graphics system is effectively unchanged.

The trick we have come up with for increasing the effective size of video is to make use of the hardware zoom capability of the graphics system. This allows us to transmit the same number of bits to the graphics system and let the explosion in bits happen on the board. Of course, the video suffers some from blockiness caused by the expansion (via pixel replication) of the image. We believe that for motion video, a larger frame size is more important than smooth spatial resolution.

We have experimented with sending these images over the Ethernet, then decompressing them and displaying them. With this setup we were able to achieve about 6.5 frames/sec. This still gives reasonable-quality motion video, but calculations tell us that we should be achieving a greater frame rate. Since we are obtaining at least 1.5 Mbits/sec over the Ethernet, we should be able to transmit about 20 frames/sec (each frame is about 8 Kbytes in compressed form). We know that currently we can not do better than 8 frames/sec due to limitations on the PC, but the rate we obtain is consistently less than this. The compressed frames are stored in Unix files and we are not being clever in opening and reading these files. We suspect, therefore, that the time is being eaten up in file access overhead.

We are continuing to push our current configuration as far as possible to support motion video. We will continuing to investigate network delivery and the behavior of the system bus. In addition, we will be looking for methods of reducing the bus load.

An ongoing task for us is the attempt to draw general conclusions from the specific work we are performing. We expect our continued work to yield insights about networks, operating systems and machine architectures for digital media.

6. Conclusions

Even the small amount of work we have done is enough to validate our premise that high-quality multimedia documents can be presented over a 1.5 Mbits/sec network. These documents can contain multi-font text, geometric graphics and photographic images. Although we have not yet worked with digital audio, we do not anticipate any major problems in using it in documents.[14] While we would not claim to have achieved success with motion video, we are quite optimistic given our early experience.

It appears possible to press JPEG into supporting motion video with a minimum of effort, although full-screen, full-motion video is probably infeasible with any reasonable image quality. This is important since it appears that MPEG boards will not be available for some time.

---

[14]We do, however, expect that the inclusion of audio will force us to face the issues involved in close media synchronization.

The ease with which we have obtained our first results is partially due to the fact that we are working with a data stream traveling at only 1.5 Mbits/sec. This low data rate means that we do not push much against the limits of the network, storage media, operating systems and CPU. Presumably, this would not be true were we dealing with a higher bandwidth network. If our data were moving at 10 Mbits/sec we would undoubtedly be experiencing bottlenecks at more places, including the operating system and processor.

The problem we have been studying is a constrained version of the general problem of dealing with networked, integrated digital multimedia. Even with limited bit rates and the elimination of most resource contention, we have begun to hit some of the limits of conventional workstation architectures and operating systems.

The most obvious example of this is the system bus, which appears likely to become the bottleneck for our system (and others) as we process digital audio and video. This confirms our belief that current workstation architectures are inappropriate for supporting multimedia applications that deal with digital media. Workstations are designed to be computation intensive, which is fine for traditional data processing applications. Multimedia applications require architectures that are communication intensive and allow the fast establishment of high bandwidth connections among arbitrary subsystems.

Many researchers and manufacturers attempt to work around this problem by building hardware that bypasses the standard data paths. For example, video is usually delivered in analog form. The video may be displayed on the screen by digitizing it in real time and feeding the bits directly to the frame buffer. Another technique keys the video to the analog signal driving the display. Both techniques avoid the system bus and the operating system.

Unfortunately, by bypassing the standard system data paths, these solutions tend to have limited functionality and it becomes difficult to integrate them with other hardware and software components. This can lead to unpleasant side effects like the disappearance of the screen cursor when it enters a video window or the inability to overlay video with graphics.

Finally, we believe that the time is ripe for the investigation we have begun into network document presentation. This is due largely to the recent advances in compression technology and processor speed and the plummeting costs of memory, storage and graphics engines.

We make use of these advances by trading local processing power for network bandwidth using a variety of techniques. In addition, recent research on improving the bandwidth of cooper loop facilities and in providing multimedia support on packet-switching networks has made it desirable to pursue the network presentation of multimedia documents at moderate bit rates.

Acknowledgements

Abel Weinrib gave me useful comments on an earlier version of this paper and Rong Chang and Bob Kraut provided criticisms of this paper. Paul Crumley was helpful in discussing the PC architecture.

References

[1] Domenico Ferrari.
Real-Time Communication in Packet-Switching Wide-Area Networks.
Technical Report TR-89-022, University of California and ICSI, 1989.

[2] Robert S. Fish and Thomas H. Judd.
A Subjective Visual Quality Comparison of NTSC, VHS, & Compressed DS1-Compatible Video.
Proceedings of the Society for Information Display , May, 1991.

[3] Howard P. Katseff, Robert D. Gaglianello, Thomas B. London, Bethany S. Robinson, Donald B. Swicker.
Experiences with the Liaison Network Multimedia Workstation.
In *Proceedings of Distributed & Multiprocessor Systems (SEDMS II).* 1990.

[4] Didier Le Gall.
MPEG: A Video Compression Standard for Multimedia Applications.
Communications of the ACM 34(4), April, 1991.

[5] Ming Liou.
Overview of the px64 kbit/s Video Coding Standard.
Communications of the ACM 34(4), April, 1991.

[6] Earl E. Manchester.
New uses for residential copper.
Telephony , June, 1991.

[7] Jonathan Rosenberg, Robert Kraut, Louis Gomez and C. Alan Buzzard.
Multimedia Communications: a User-centered Viewpoint.
IEEE Communications (Special Issue on Multimedia Communications) , May, 1992.

[8] Gregory K. Wallace.
The JPEG Still Picture Compression Standard.
Communications of the ACM 34(4), April, 1991.

[9] R. S. Wolff.
What's Ahead for Copper?
Telephone Engineer and Management , October, 1988.

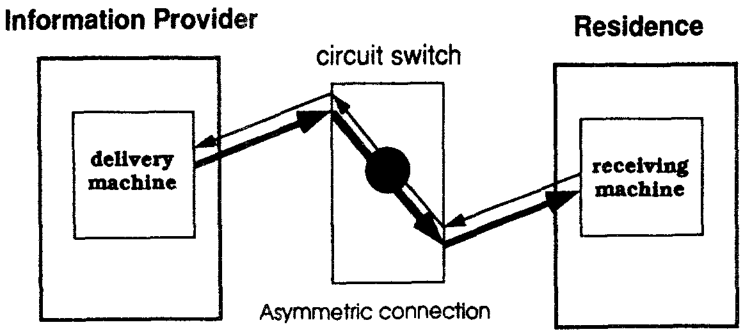

Figure 1: Architecture of First DEMON Prototype

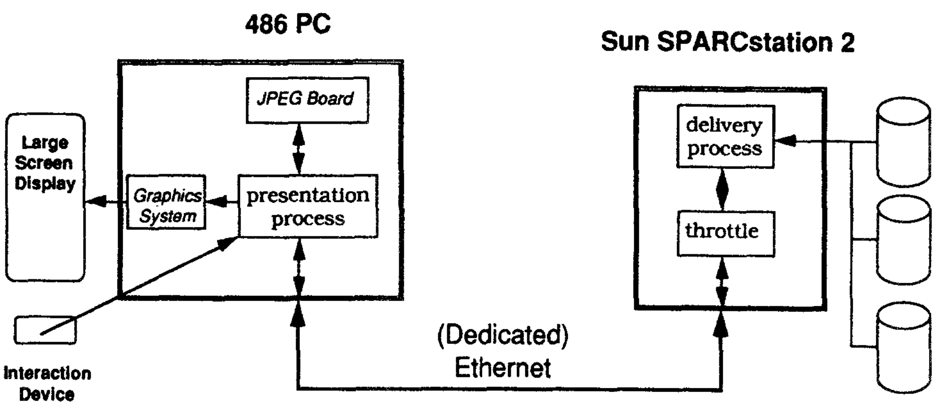

Figure 2: Configuration of First DEMON Prototype