

AN INTEGRATED CIRCUIT/PACKET SWITCHED VIDEOCONFERENCING SYSTEM^{1 2}

H. A. KIPPENHAN, JR., W. P. LIDINSKY, G. A. ROEDIGER
HEP Network Resource Center
Fermi National Accelerator Laboratory, Batavia, IL 60510, USA

T. A. WATTS
Department of Physics and Astronomy
Busch Campus, Rutgers University, Piscataway, NJ 08854, USA

The HEP Network Resource Center (HEPNRC) at Fermilab and the Collider Detector Facility (CDF) collaboration have evolved a flexible, cost-effective, widely accessible videoconferencing system for use by high energy physics collaborations and others wishing to use videoconferencing. No current systems seemed to fully meet the needs of high energy physics collaborations. However, two classes of videoconferencing technology: circuit-switched and packet-switched, if integrated, might encompass most of HEP's needs. It was also realized that, even with this integration, some additional functions were needed and some of the existing functions were not always wanted. HEPNRC with the help of members of the CDF collaboration set out to develop such an integrated system using as many existing subsystems and components as possible. This system is called VUPAC (Videoconferencing Using PAcKets and Circuits). This paper begins with brief descriptions of the circuit-switched and packet-switched videoconferencing systems. Following this, issues and limitations of these systems are considered. Next the VUPAC system is described. Integration is accomplished primarily by a circuit/packet videoconferencing interface. Augmentation is centered in another subsystem called MSB (Multiport multiSession Bridge). Finally, there is a discussion of the future work needed in the evolution of this system.

1. INTRODUCTION

The HEP Network Resource Center (HEPNRC) at Fermilab and the Collider Detector Facility (CDF) collaboration have been working to evolve a flexible, cost-effective, widely accessible videoconferencing system for use by high energy physics collaborations and others wishing to use video conferencing. No current systems seemed to fully meet the needs of high energy physics collaborations. Two classes of videoconferencing technology: circuit-switched and packet-switched, if integrated, might encompass most of HEP's needs. But even with this integration, some additional functions were needed and some of the existing functions were not always wanted. HEPNRC in collaboration with CDF set out to evolve such an integrated system using as many existing subsystems and components as possible. We call this system VUPAC (Videoconferencing Using PAcKets and Circuits). A key component of VUPAC is the MSB (Multi-Session Bridge).

Since DOE/ER has a room-based circuit-switched videoconferencing system (ERVN--the Energy Research Video Network) and CDF was already using ERVN for some of its conferences, ERVN was the natural and necessary choice for the circuit-switched system. With respect to packet-switching, work within the Internet community (also partially supported by DOE/ER) has let to a set of experimental tools and protocols

¹ Supported in part by the US Department of Energy under contract number DE-AC02-76CH03000.

² Extensive involvement and consultation were made by the following CDF collaborators: T. Phillips, Duke U.; L. Kirsch, Brandeis U.; T. Thomas, U. of New Mexico, and E. Wicklund, Fermilab.

that can be used together as a videoconferencing system. This system, which is in the public domain, has become widely deployed and used. While it is sometimes referred to as MBONE (Multicast backBONE), this backbone is only part of the system; other parts include workstation software tools that packetize, compress, send and receive video and audio, and present the video and audio on a workstation. HEPNRC and CDF felt that while the tools were necessary, the use of the multicast backbone was undesirable in some circumstances. In this paper, MBONE will refer only to the multicast backbone. When referring to entire packet-switched videoconferencing systems, a more encompassing term PSVID is used.

2. BACKGROUND

2.1 Energy Research Videoconferencing Network - ERVN

ERVN was conceived of in about 1989. It was first embodied as a pilot with conference room sites at three U.S. Department of Energy laboratories: LBL, SSCL, and FNAL. It used proprietary equipment and protocols and dedicated bit rate from Energy Science Network (Esnet). As a pilot, the number of sites expanded to eight, began to be used by researchers, and was slowly converted to the H.320/H.261 standard protocols. In 1994 it was converted to a production system run primarily by ESnet.

Today ERVN uses standard circuit switched protocols, employs equipment and software from multiple vendors, uses Basic Rate ISDN for communication, has over 35 sites, and is used extensively. ERVN is a circuit-switched videoconferencing system that, until recently, was primarily room-based. Recently desktop-oriented systems have become available.

2.2 Packet-Switched Videoconferencing (PSVID)

There is much interest in using personal computers and workstations (hard to tell the difference these days) for communication of video including voice. There are many proprietary and semi-proprietary systems. Many of these use circuit switched communications; some use packets. One of the packet efforts consists of public domain software developed and used on the Internet. It consists of (1) router software designed to multicast conferences on irregular mesh packet switched networks, and (2) software tools that packetize, compress, and send and receive video and audio on a workstation. The virtual IP multicast network that is overlaid on the Internet is known as MBONE; it processes and routes IP multicast packets over the normal Internet IP. Initially the mrouter that comprised MBONE were implemented in user workstations. This, however, put excess traffic on local area networks attached to these mrouter workstations. Now, most of the commercial router vendors either have or are putting multicast routing code into their products.

There are presently four workstation-resident tools that are most widely used. Network Video (nv) accepts video from a camera, compresses it, and transmits it to a specific IP multicast address. It also accepts video from the network or other communication link for a specific IP multicast address, decompresses it, and displays it in a window on the workstation monitor. The Video Audio Tool (vat) does much the same for audio. A tool called Whiteboard (wb) allows several networked workstations to view a shared window and in basic ways edit what is on the screen. Finally, the Session Director (sd) provides a place to set up and advertise sessions and also a mechanism for attaching to a session, bringing up both nv and vat in the process. These tools have been

ported to several different UNIX-based workstations and are presently being ported to Microsoft Windows.

There are a number of other tools that are less widely used but which, in some cases, are gaining in popularity. These include vic, imm, nevot, and mmcc. These tools are being considered for future incorporation into VUPAC. This is discussed more in a later section of this paper.

A video capture board is needed to interface with a camera. This means that to transmit video, such a board is needed. However, almost any workstation running the software can attach to a conference, receive video, and both receive and transmit audio (if it has or is provided with audio capability). Vat, nv, and wb can be used on a point-to-point basis without MBONE.

This public domain packet-switched system has been extensively used on the Internet for the broadcast of public meetings and conferences, private meetings and conferences, speeches, etc.

3. NEEDS AND LIMITATIONS

3.1 Cost

ERVN site systems have a cost per site of \$20K - \$50K for a room-based system and from \$5K - \$15K for a desktop system. ERVN also requires a room in which to install the equipment and to use as the videoconference room. This combination often represent an insurmountable obstacle for would-be sites. Conversely, almost anyone with a UNIX workstation and Internet connectivity can listen, view, and talk to an Internet audio/video session. Since UNIX workstations are frequently available to researchers, passively participating in conferences often costs nothing other than the use of Internet bit rate. Active participation requires a video and perhaps audio board. Upgrades for active participation presently cost about \$1K - \$2K.

Also, an ERVN Basic Rate ISDN (BRI) call costs about \$30 per hour. Each site must place a call to the videoconference bridge. Out-of-pocket costs for site participation in an ERVN conference might be several hundred dollars per conference. Using Internet audio and video, the out-of-pocket communication costs are zero. The costs of using the Internet of course are not zero and indeed may be greater than that of ISDN. However, this is either a sunk cost or one that is not seen in individual budgets.

3.2 Privacy

With ERVN it is generally possible to achieve an adequate level of privacy since the conference sites are connected using circuits. Sites other than those explicitly connected cannot easily eavesdrop. Also, private conferences are not publicly announced. In contrast the Internet packet-switched system, when using sd and MBONE, is public, advertised, and multicast. Even if sd is not used, privacy is by obscurity--i.e., the multicast IP address is not known.

Using the present tools, encryption can be applied to audio but not video. (While nv does not have encryption, more recent tools such as vic do.) Encryption can help, but the encrypted packet stream is still available, can be copied, and decrypted off line.

A way to achieve a degree of privacy within Internet packet audio/video that is comparable to that of ERVN is to avoid the use of multicast (e.g., MBONE) and sd, and to provide additional packet switched functionality similar to that of an audio or video bridge. This in turn begins to lead toward the eventual use of virtual circuits.

3.3 Connectivity

Basic Rate ISDN (BRI) is a long way from achieving ubiquity. Many sites that might otherwise be an ERVN site cannot be so because these needed communication services are not available. BRI is becoming available in many metropolitan areas, but has a long way to go, especially in more rural localities.

The Internet, while connected to nearly every site that might be interested in participating in a audio/video session, sometimes has poor performance characteristics for isochronous traffic such as packet audio or video. Hops from source to destination may very large, resulting in long delays and large delay variance making conversation unintelligible. A link or router along the path may have a great deal of traffic to handle. This will also result in large delays and delay variance. Also, the bit rate available is sometimes not enough even if fully dedicated to a video conference.

There are four ways to aid time-critical packet delivery in a packet network:

- (1) give the time-critical or isochronous packets priority,
- (2) reserve bit rate and router capacity,
- (3) engineer sufficient capacity so that no packets exceed delivery time limits, or
- (4) bypass the packet network in favor of other connectivity.

The merits of all of these approaches are known. (1) and (2) are being considered by the Internet Society. (1) is inherently unfair to asynchronous traffic and presents problems for competing isochronous traffic. Right now there is no effective way of implementing approach (2). The third approach increases the capacity of the installed base of links and routers and installs additional links and routers to known "hot" sites. This is very costly. The last scheme can be selectively implemented and requires no changes to the packet network. Costs are reasonable.

The Internet in its present form was designed for asynchronous data; not for isochronous voice and video. Handling isochronous traffic is inherently difficult.

3.4 Scalability and Congestion

The number of ERVN sites participating is limited in large measure to the number of ports on an ERVN conference bridge. Also more than eight to ten sites per conference becomes either unwieldy or impossible to use. If several conferences of this size occur simultaneously, the capacity of present ERVN bridges is exceeded. Furthermore, present MCUs are expensive, the cost per MCU port being about \$8K - \$10K.

The Internet system can theoretically support hundreds of participants per multicast session and still have multiple sessions. This is possible because of the distributed nature of irregular mesh networks. On the other hand, "hot spots" of high packet traffic can and do occur, resulting in poor video and unintelligible audio. The needed Internet bit rates are given in Table 1 with the values for NTSC video for comparison.

Table 1: Needed Internet Bit Rates

Image Resolution (pixels)	Frame Rate (frames/sec.)	Compressed		
		Video Bit Rate (Kbps)	Audio Bit Rate PCM2 (Kbps)	Total Bit Rate Needed (Kbps)
640 x 480	2	130	70	200
352 x 288 (full CIF)	4	130	70	200
~352 x 288 (NTSC)	30	---	---	~40,000

At 200 Kbps, if several sessions traverse the same line or a session is carried on a line heavily used by data traffic, congestion will occur resulting in delayed packets. This will impinge on data users and also increase both the delay and the delay variance of the audio and video. Delay variance of audio is the real killer. As delay exceeds the maximum allowable delay for voice, packets are discarded resulting in voice dropout. Audio rapidly becomes unintelligible. The video, while unsatisfactory under these conditions, can usually still be used with effort. If there is a large amount of multicast traffic, improved multicast pruning should help the delay and delay variance problems for the current level of Internet audio/video traffic.

3.5 Quality

The best video resolution obtainable from packet video at this time (640 x 480) is better than either circuit switched video such as ERVN or NTSC home television (352 x 288 pixels interlaced). However the frame rates are much slower. Packet video is somewhat better for non-moving images e.g., viewgraph displays.

With respect to audio, the quality of packet audio is definitely worse than that of ERVN. This is largely due to the inherently connectionless nature of the Internet transport used by the audio tool (i.e., UDP). Packets are lost. Others are delivered out of order and are discarded since discarding them results in a less intelligibility than playing them. Also since lossy compression is used, unplayed packets have a greater effect than would be the case if there was no compression.

4. VUPAC - AN INTEGRATED SYSTEM

To solve or circumvent some of the above limitations while still using existing systems, PSVID and ERVN have been integrated and augmented in the VUPAC system.³ Figure 1 shows the overall VUPAC system. The Interface Workstation is a UNIX workstation that runs the PSVID tool suite (i.e., nv, vat, wb, ...). Similarly, the Interface codec system is an ERVN codec system that consists of a codec, cameras, monitors, and microphones. It employs the ITU-CCITT H.320 suite of protocols.

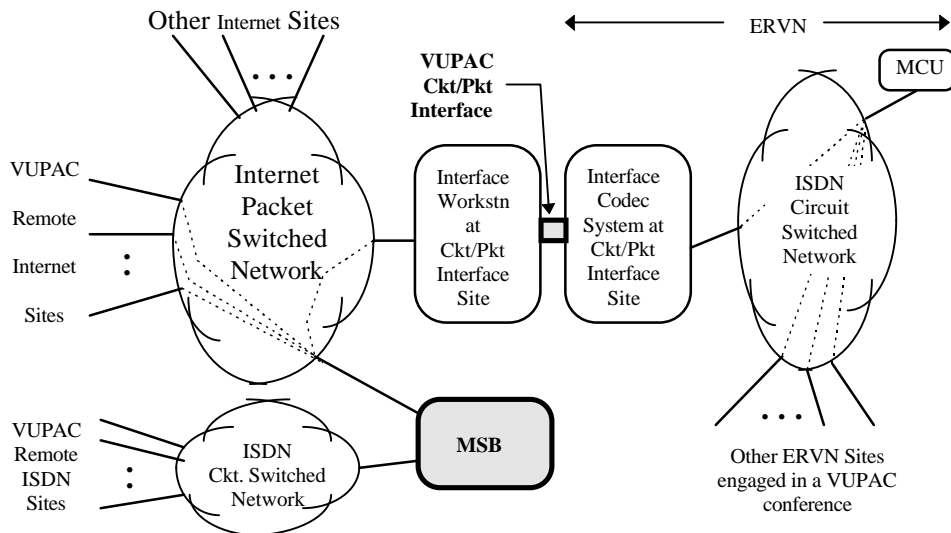


Figure 1: VUPAC Video Conferencing System

³ While VUPAC is currently experimental as is the Internet packet video/audio suite of protocols and tools, VUPAC is being used by the CDF high energy physics collaboration on a regular basis.

The ISDN circuit switched network is the commercial network offering. Although one network, it is shown twice at the lower left and at the right in Figure 1 to simplify the drawing. Codecs in the ERVN system connect via ISDN BRI service to the MCU (Multiport Control Units) located at the ESnet site at LLNL. The MCU, a videoconference bridge, connects together the codec systems designated for a specific conference.

The Internet "cloud" represents the entire Internet. The VUPAC Remote Internet Sites are workstations that are participating in a VUPAC conference using Internet packet switching connectivity to carry IP packets with video and audio information. The VUPAC *Remote ISDN Sites* are workstations that are participating in a VUPAC conference using ISDN circuit switching connectivity to carry IP packets with the same video and audio information. The *Other Internet Sites* are included in Figure 1 for completeness. They represent all other computers connected to the Internet that are not involved in a VUPAC conference.

The MSB code was written to provide a bridge between combinations of the conventional multicast video conferences and circuit connected sites. In practice it has also been used to connect other types of sites where control over conference data streams was required. Packet video conferences have both audio and video streams. Addresses associated with a session (one or more sessions make up a video conference) can be point-to-point or multicast in scope. The MSB code runs on a machine that sees all the sessions to be bridged. All the unicast sites actually unicast to the MSB machine. The multicast sessions must be in the ttl (IP time to live) scope of the session. The MSB takes the streams from each session and in a simplex fashion and decides which sessions will receive a copy of the stream. Thus each remote site can be configured to receive any audio and or video stream. For ISDN circuit connections the bandwidth is limited by a number of B channels aggregated together to form the switched circuit. A site connected with 3 B channels has 192kbps of bandwidth. The 192kbps circuit can handle a 70kkbs PCM audio stream and 100 Kbps video stream. Other sites in the MSB-bridged conference may have more bandwidth and thus could be configured to see more than 1 video stream.

Remote sites that are participating in the multisession video conference will run a small piece of code, *vtconnect*, that connects to the conference. The form of the command for this code is:

```
vtconnect <msb_machine_IP_name> <password>
```

The MSB uses the remote machines address and the password to decide what that site has been configured for in this conference. The MSB tells the vtconnect program what programs to start with the appropriate parameters. The user will see the video and audio tools automatically started. When the MSB is notified that the video and audio tools at the remote site are started, it will begin to forward the bridged audio and video streams. When the remote site on a unicast connection stops the video and audio tools the MSB is notified by vtconnect and will then stop sending the streams.

5. FUTURE HEPNRC WORK

There is a built-in tracroute capability in the MSB. This will be used in the future to find overlapping routes and congestion. With this information a decision can be made to set up an ISDN circuit (long distance toll charges applied) or use the Internet if bandwidth is available.

The switched video based on who is talking will also be added. This has been requested by people using the Pacific links. In this way a static configuration decision need not be made about video but rather by keying the microphone one can determine who uses the expensive Pacific link bandwidth for video.