Audio Video Conferencing

Geoffrey Fox, Gurhan Gunduz and Ahmet Uyar
Florida State University
Department of Computer Science and
CSIT (School of Computational Science and Information Technology)
400 Dirac Science Library
Tallahassee
Florida 32306-4120
fox@csit.fsu.edu

1 Introduction

This report describes a focused effort studying the Access Grid and HearMe audio video systems in detail. The original Tango Interactive system had its own audio video conferencing system Buena Vista built in. We considered this important for two reasons. Firstly it allowed a single invocation and user registration process for all parts of the collaboration; secondly the alternatives were not very satisfactory when we developed Buena Vista some 4 years ago. We have made a different choice with Garnet [1] as it is now longer realistic or even useful to develop our own audio video support; rather we will use best of practice solutions from other commercial or research developers. In some cases we can use an available API to link invocation and registration; in others we view audio video and shared document support as separate stand-alone systems. We can classify audio video support into three classes:

1. **Low-end:** Illustrated by HearMe Audio and CUSeeMe Desktop Video
2. **Medium:** Illustrated by PictureTel and Polycomm
3. **High-end:** Illustrated by Access Grid and a system Admire from BUAA University in Beijing, China.

In this report we discuss the Access Grid and HearMe systems and in the rest of the introduction describe our general plans in this regard. The last two sections describe these two systems in technical detail.

The Access Grid (AG) is developed originally by Argonne [3] but extended as part of the NCSA Alliance. There are now over 50 of these high-end audio video conferencing systems installed worldwide. 2 Systems are being installed at FSU – one fixed and one transportable model aimed for "teachers" in a distance education scenario or for use in a small conference room. The Access Grid is described at http://www.mcs.anl.gov/fl/accessgrid/ and training sessions are available. NCSA recently offered AG training to ERDC personnel at the Access Center in Washington and we expect that we can arrange this for all the major PET and MSRC sites that wish to install AG nodes. We intend that a "train the trainers" model so that once a cadre of experts exists at HPCMO/PET sites they can help other sites in this community. This will build a critical mass of AG systems to enable electronic collaboration to be effective in HPCMO/PET. Currently ERDC JSU and ARL intend AG installations.

Although the AG is an impressive system there are some issues to address. We recommend using different shared document systems when the simple shared PowerPoint of the AG is insufficient. This motivates our system to use AG for community conferencing but Garnet or commercial collaboration systems like Centra or Webex for document sharing. Further we think the AG community should look at H.323 or SIP (the audio video interoperability standards) compliance for this technology. This would allow one to support hybrid sessions involving simultaneously systems such as AG, PictureTel, Polycomm, CUSeeMe and HearMe. We discuss the H.323 and SIP standards in a detail review we produced [4]. Finally we need the AG to support partitioning of clients so that multiple communities can be separately administered. An important (for HPCMO) recent enhancement to the AG supports encrypted media streams using the new AES standard [5].

HearMe http://www.hearme.com [2] is a low-end audio conferencing system supporting general mix of phones and Internet clients with participant control. The phone option is helpful as it allows audio communication with better quality of service than can be guaranteed on the Internet. Note that the phone and Internet options are integrated as both are converted to the same codecs and recorded for later replay through the web. We are adding SMIL (the W3C standard for multi-stream multi-media files) based replay of session by converting G.711 or G.723 codecs digitized on the HearMe server to RealAudio. We have a HearMe server installed at FSU with a license for 20 simultaneous users.

The Access Grid produces a "designed space" aimed at supporting groups interacting with groups – PictureTel or desktop systems are more optimized for individual interactions. The AG features hands free high quality audio, multiple (4) video and audio streams, and lifesize displays. 4 PC's control it and AG equipment includes a echo canceling box, multiple camera and projector or frame buffer displays (at least 3). HearMe provides two types of conferences: standard and moderated. In the standard conference, every client has the same privileges and can talk at any time. In a moderated conference, there are three types of users: moderator, panelist and participant. The Moderator is the creator of the conference and has full control over the session. Panelists are those given the right to talk while participants can listen but need the permission of the moderator to talk.

2 HearMe voice over IP system

URL: http://www.hearme.com/

2.1 Introduction

HearMe [2] is a voice over IP application to do voice conferencing. It provides full-duplex voice communication among participants. **It has no video capability**

Today, there are three solutions for teleconferencing: Firstly one can use the Internet as a medium; people attend conferences by using PCs. Secondly one can use phone lines with a conference typically arranged by telephone companies. Thirdly one can use both Internet and phone lines as a medium. In this case people can attend conferences either by using PCs or phones. The HearMe system is based on the third solution.

Although using only Internet for teleconferencing is cheap, the quality of voice is often not satisfactory. On the other hand, the quality of voice usually reasonable when phone lines are used. However using phones is unaffordable and inconvenient for many people. The third solution combines the quality of phone lines and low cost of Internet. The idea is that the speaker will talk on the phone providing better voice quality and listeners can either use phones or PCs. In addition to its cost benefits this solution is also more convenient than the other two solutions. A phone-to-PC gateway is used to connect phone lines to Internet.

2.2 Services

HearMe provides two types of conferences, standard and moderated. In a standard conference, everyone has the same privileges. Anyone can talk at any time. On the other hand, in a moderated conference there are three types of users, moderator, panelist and participant. Moderator is the one who has full control over the conference. He or she gives the permission to talk and has the right to eject a participant from the conference etc. a panelist has right to talk by default. Participants need permission to talk.

HearMe provides a recording mechanism for live sessions. But unfortunately right now they do not provide any tool to replay recorded conferences. Recorded conferences are in HearMe proprietary format and one needs to write its own decoder to replay it. FSU is currently implementing replay using Internet standards: RealPlayer or Microsoft technology.

2.3 Architecture

There are three servers, talkserver, MCU, and bridgeserver. The talkserver is used to manage the conferences such as creating a conference, destroying a conference, getting information about a conference etc. The talkserver is basically used by administrators. MCU(Multi-point control unit) is the one who does the real job, getting voice packages from different people and transmitting them to appropriate recipients. In addition MCU can record the conferences. Users directly connect to the MCU. Bridge server and an IP gateway is used to include phone connections into conferences. Gateway converts analog voice signals to digital form and vice versa. Bridge server is used as a bridge between the gateway and the MCU.

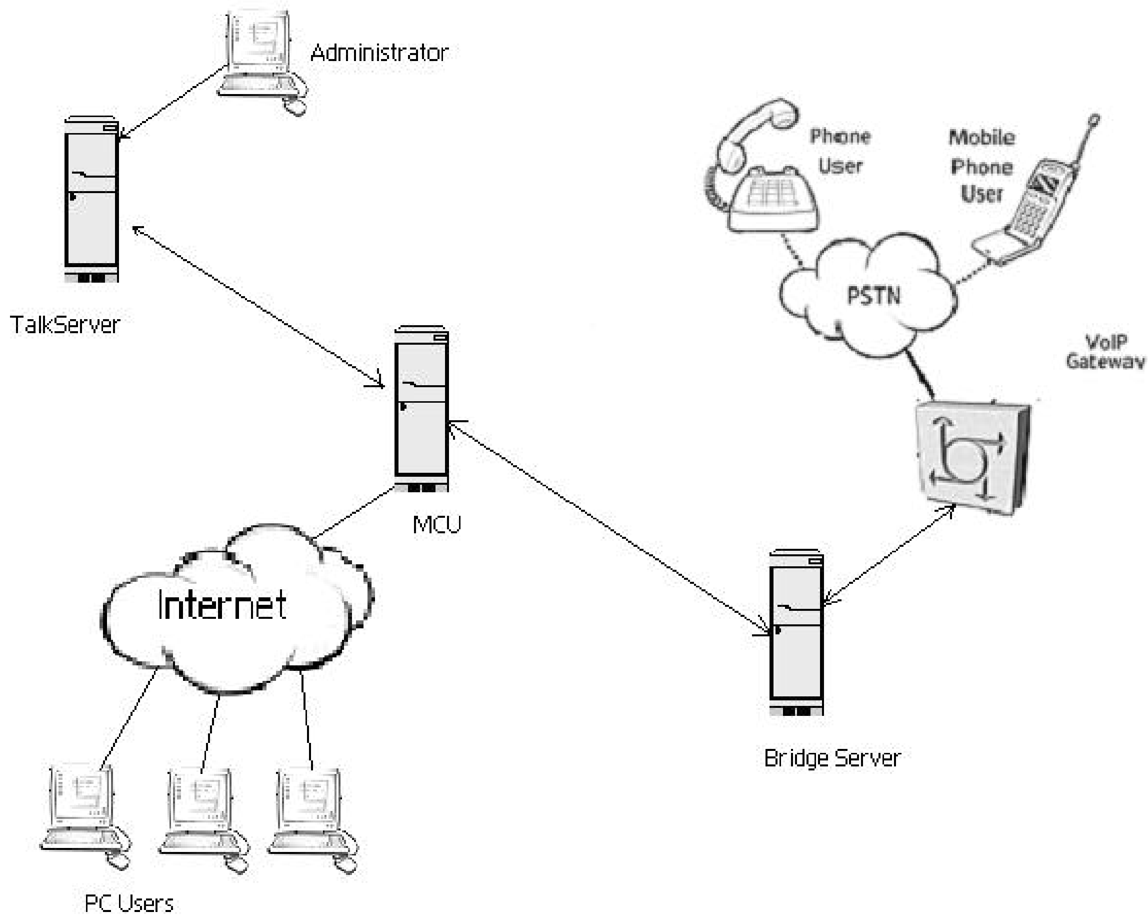


Figure 1: the architecture of HearMe voice over IP system.

2.4 Protocols

HearMe uses industry standards in their voice over IP system. Their system architecture is based on the H.323 standard described in ref. [4] that is a recommendation from International Telecommunication Union (ITU). It sets standards for multimedia communications over Networks that do not provide quality of service. It sets standards for voice, video and data. HearMe currently uses G.723.1 for voice compression. G.723.1 is also a recommendation of ITU and widely used for Internet telephony and web conferencing. They are also using ITU G.711 for voice compression, which provides better voice quality and requires higher bandwidth, but it is currently not fully functional. In addition HearMe uses session initiation protocol (SIP) to initiate sessions.

2.5 Bandwidth requirements

Each client needs 28.8 Kbps or greater Internet connection.

2.6 Client side System requirements

The minimum system requirements for each client is

- Pentium 166MHz

- 32Mb of RAM
- Sound Blaster compatible 16-bit sound card
- Headset or speakers and microphone
- Windows 95, 98, or NT
- Internet Explorer 4.0 or later/Netscape 4.5 or later

2.7 Server side System requirements

TalkServer:

- Pentium III @ 500MHz
- 256 MB RAM
- 10 GB disk
- 100 Mbit/sec network interface card
- RedHat Linux 6.1
- Oracle 8i

MCU:

- Pentium III @ 500MHz
- 256 MB RAM
- 10 GB disk
- 100 Mbit/sec network interface card
- RedHat Linux 6.1

BridgeServer:

- Pentium III @ 500MHz
- 256 MB RAM
- 10 GB disk
- 100 Mbit/sec network interface card
- RedHat Linux 6.1
- H.323 VoIP Gateway (ref.:Cisco AS5300)

2.8 Cost

The cost of a HearMe Voice Developer's Kit is $10,000. It includes:

- Server software for TalkServer, MCU and BridgeServer.
- License files to allow service for up to 16 concurrent customers. One can add more at additional cost.
- HearMe Voice SDKs

2.9 Conclusion

HearMe provides a solution for the voice conferencing over the Internet and it also allows telephone users to attend these conferences. It is relatively cheap and high quality compared to other solutions existed on the market today. Although they lack some features like replaying recorded conferences, they are on the right track and they will add those features in future releases.

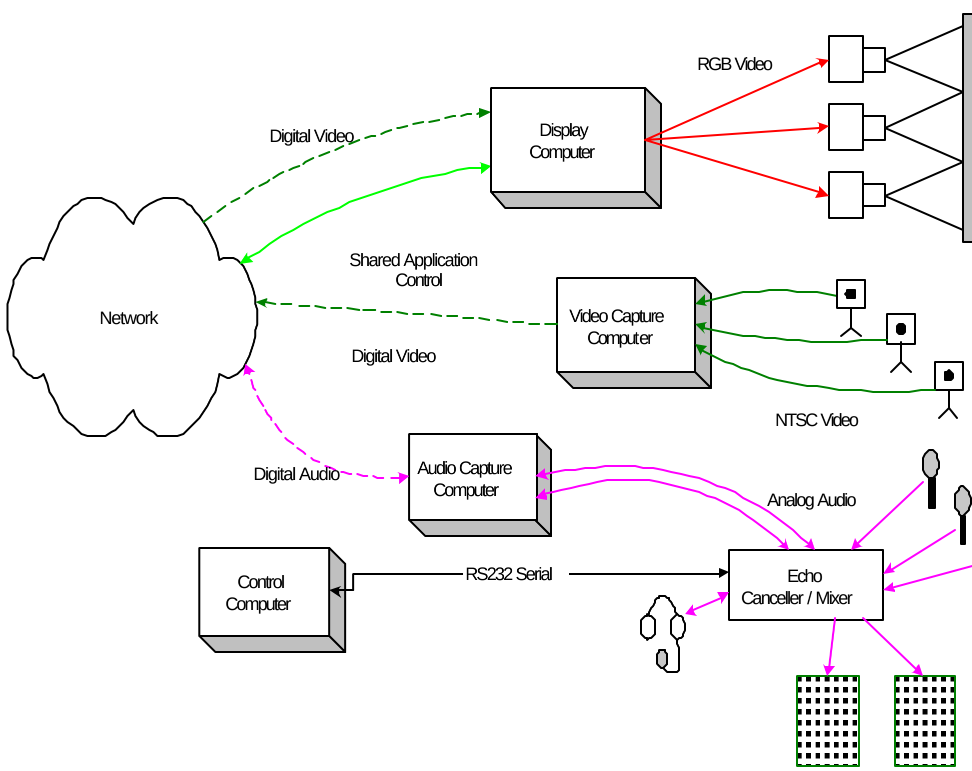
3 Access Grid

URL: http://www-fp.mcs.anl.gov/fl/accessgrid/default.htm

3.1 Introduction

The Access Grid [3], designed by Argonne National Laboratory, is a system that enables group-to-group collaboration across Internet by providing multiple video and audio streams among groups. The Access Grid consists of many AG nodes around the country. AG node is a special room designed to participate in AG meetings. It consists of video cameras, projectors, audio equipment, computing equipment and high-speed Internet connection. There are currently around 50 AG nodes in US.

The Access Grid project focus is to enable *groups* of people to interact with grid resources and to use the grid technology to support group-to-group collaboration at a distance. This is the main difference between desktop-based collaboration tools and the AG. The AG is designed in a way to give sense of presence to remote participants. AG nodes have large displays, multiple video and audio streams. Audio system is designed in a way that every participant can talk hands free.



3.2 Video

Each AG node has four video cameras. It is important to be able to see every participant in a remote site. One of them is used to get the video stream of presenter. Second one is for display screen shot (it is important for remote sites to see what we are seeing). The last two are for audience shot. Video cameras should be placed in a way to facilitate the feeling of eye contact

3.3 Audio

The most important thing in audio configuration is to make very participants be able to talk hands free. Therefore there should be adequate number of microphones placed around the room properly. There must be also an echo canceller device in each AG node. Two speakers are used to project good quality of audio into the space.

3.4 Projectors

Large display screens are used in each AG node, because it is important to get real life-size images of participants at remote sites. This is accomplished by using three high-resolution projectors. Each node gets 4 video streams from every participating nodes, so there are a lot of video streams coming to one node. Therefore, it is important to have three projectors.

3.5 Computers

There are four computers, display computer, video capture computer, audio capture computer and control computer, in each AG node.

Display computer is used to get video streams from other sites and display them on screens. It has a special software running on it to manage the video streams on screens. It runs Windows 2000 operating system and has a multi-headed video card.

Video capture computer is used to get the video streams from the cameras in the room. It has fours video capture cards on it and runs Linux operating system.

Audio capture computer gets audio streams from the microphones in the room and encodes and broadcasts them to other nodes. It also gets audio streams from remote nodes and decodes them. It runs Linux operating system.

Control computer is used to run control software for the audio gear(echo canceller). It runs Windows 98 operating system.

3.6 Software

Access Grid partners have developed several pieces of software. One of them is a multicast beacon that is used to monitor the network status of nodes. Another one is distributed PowerPoint tool that is used to share PowerPoint slides in a session. Persistence and scope are provided by using the Virtual Venue software developed at Argonne. It has components that run on the Display, Video, and Audio machines, as well as a central server. VIC is another software that is used to manage displays. RAT software is used to manage audio.

3.7 Network

The access grid uses network multicast among AG nodes. A full AG session can deliver many dozens of video streams to a node. The bandwidth required for each stream can vary from 128 Kb/s to 512Kb/s depending on the settings. Inadequate bandwidth results in unintelligible audio and jerky-motion video.

3.8 Protocols

The Access Grid uses Robust Audio Tool (RAT), an open source software, for handling audio. It is an audio conferencing and streaming application that allows users to participate in audio conferences over Internet. RAT is based on IETF standards and uses RTP above UDP/IP as its transport protocol. RAT features a range of different rate and quality codecs, G.711(64kb/s), Wide-Band ADPCM(64kb/s), G.726 ADPCM (16-40kb/s), DVI ADPCM (32kb/s), Variate Rate DVI ADPCM (~32kb/s), Full Rate GSM (13kb/s), LPC (5.6kb/s). It also features encryption so you can keep your conversations private.

The Access Grid uses Video Conferencing Tool (VIC) for handling video. VIC is a real-time, multimedia application for video conferencing over the Internet. It is developed by Network Research Group at the Lawrence Berkeley National Laboratory in collaboration with the University of California, Berkeley. VIC is based on Real Time Transport Protocol (RTP) developed by IETF. To be able to use conferencing capabilities of VIC, your system should support IP multicast. VIC uses H.261 protocol to encode and decode video streams. H.261 is the protocol that defines the video portion of H.323.

3.9 Recording/Playback

Argonne has built a recording and playback engine, Voyager Multimedia Multistream, that can record and playback live sessions. It saves multiple video and audio streams to disks without loss. It also synchronizes in time the multiple audio and video streams when playing back.

3.10 Required Equipments

An Access Grid node consists of several hardware equipments. These are basically;

- 4 PCs
  - Display computer
  - Video capture computer
  - Audio capture computer
  - Control computer
- 4 cameras
- Several microphones
- Echo canceller device
- Three projectors or displays

3.11 Cost

Computing equipment	$12,455
Network equipment	$750
Other computing equipment (monitors, KVM switch)	$1,800

audio configuration	$10,564
Video cameras (4 Sony EVI-D30)	$5,196
Projectors (3 Epson 710c)	$15,900
Total (January 2001)	***$46,665***

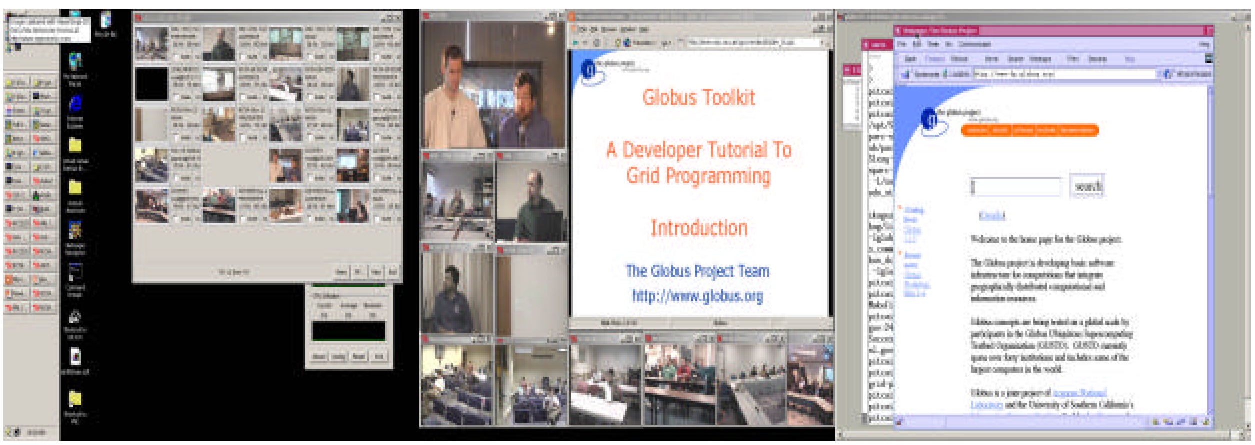
These prices and equipment may vary depending on the configuration of the AG node.

Access Grid software is free and will be available on a CD.

3.12 Conclusion

Today the group-to-group collaboration is a need in many areas and it is not easy to gather everyone to the same place. Access Grid is trying to make this happen in remote locations by providing real life size images and hands free audio. They are quite successful on this and the number of institutions that are installing the Access Grid is increasing rapidly.

We conclude with some pictures from an Access Grid Session



4 References

1) Geoffrey C. Fox, "Architecture and Implementation of a Collaborative Computing and Education Portal", ERDC Technical report May 2001.
2) HearMe http://www.hearme.com/
3) Access Grid http://www-fp.mcs.anl.gov/fl/accessgrid/default.htm
4) FSU Review of Collaboration Tools http://aspen.csit.fsu.edu/collabtools/CollabReviewmay09-01.doc
5) AES Encryption standard http://csrc.nist.gov/encryption/aes/