Global Serverless Video Conferencing over IP[†]

Thomas C. Schmidt, Matthias Wählisch
{schmidt,mw}@fhtw-berlin.de

Computer Centre, Fachhochschule für Technik und Wirtschaft Berlin
Treskowallee 8, 10318 Berlin, Germany.

Hans L. Cycon, Mark Palkow
{hcycon,mpalkow}@fhtw-berlin.de

FB Ingenieurwissenschaften I, Fachhochschule für Technik und Wirtschaft Berlin
Allee der Kosmonauten 20-22, 10315 Berlin, Germany
.

Keywords: *Peer-to-Peer Video Conferencing, User Locating, User Address Resolution, Multicast Videoconferencing, Wavelet Transform*

Abstract

In recent years the capabilities of the common Internet infrastructure have increased to an extent where data intensive communication services may mature to become popular, reliable applications. Videoconferencing over IP can be seen as such a highly prominent candidate. However, heavy infrastructure and uneasy call handling hinder acceptance of standard solutions.

This paper presents a more lightweight framework - both communication scheme and conferencing software - to overcome these deficiencies. A simple, ready-to-use global location scheme for conference users is proposed. First practical experiences are reported.

1. Introduction

In recent times video conference solutions communicating via the Internet Protocol have become more and more available and mature. Establishing feasible audio-visual sessions between Internet-connected desktop computers no longer remain an ambitious task, provided all partners access compatible tools and know of each others location.

In adopting the Internet protocol standards as the underlying, commonly available communication infrastructure Video Conferencing over IP (VCoIP) can be expected to be broadly around, soon. When heading towards VCoIP as a standard Internet service important steps for usability on a global scale have to be taken. Any demand

[†] This work was supported in part by the EFRE program of the European Commission.

of specific hardware or dedicated networking infrastructure is likely to hinder VCoIP roll-out and should therefore be avoided.

Up until now the employment of video conference applications has been dominated by ISDN systems. This traditional technology offers a person to person, respectively meeting oriented, private service as telephony in general does. The communication paradigm consists of a point to point connection between dedicated devices under specific user attendance.

VCoIP is in contrast embeded into general Internet-connected working devices and today oriented towards more or less public conference groups. As employment of VCoIP grows more mature, though, the need for meeting oriented, private sessions urgently has to be met. Since it addresses people instead of devices it should adapt to the common internetworking communication paradigm of mobile users accessing services, not equipment.

In the present paper we address the issue of global, decentralised VCoIP communication infrastructure. We present a simple, ready-to-use approach to user look-up without modification of the current Internet information infrastructure as well as a serverless, highly efficient VCoIP software implementing our information strategy. The aim of our solution has to be seen in rigorous ease and functionality on the price of loss in generality.

This paper is organised as follows. In section 2 we discuss communication strategies, introduce our basic ideas and examples of related work. Section 3 presents the daViCo video conference software and its core technologies. Finally, section 4 is dedicated to conclusions and an outlook on practical experiences of the solution.

2. A Distributed Global Communication Framework

2.1. VCoIP Architectures and Related Works

Video conferencing over IP still waits to be established as a regular communication service. To progress its dissemination throughout the Internet community the most simple application scenario should be kept in mind: Any Internet user may call any online partner by just starting an appropriate software tool and addressing a common name.

Video conference communication is a person-oriented service. As the Internet in general accounts for location independent access of walking users a look-up strategy is needed to transparently find any desired partner. Implementation then has to take care of the appropriate user/device mapping. Internet electronic mail presently is organised in an according fashion with the significant distinction, however, of mail being a light-weight, asynchronous process.

The traditional, ISDN compatible architecture of VCoIP systems has been defined in the ITU standard H.323 [1]. Central parts of this model are derived from a client-server principle with a Multipoint Control Unit (MCU) serving videostreams in multipoint conferences and a Gatekeeper providing connection control and address translation. One advantage of the MCU facility design lies in its ability to transform data streams between different video/audio codecs. The major disadvantages of course are drawn from the request for heavy infrastructural changes and significant latency additives [2].

The H.323 architecture must be considered as local in the sense that all participants need to agree on the common MCU and Gatekeeper servers which, obviously at least for the MCU, suffer from severe scaling deficiencies. No global naming is defined except for telephone numbers handled by ISDN gateways and the Q.931-compatible signalling protocol H.225.0. H.323 concepts center around the ideas of telephone-based wide area connectivity and are obsoleted by the simple observation that use of video conferencing via telephony does not grow. Consequently attempts are made to overcome local restricitions in addressing by interconnecting Gatekeepers via meta-directory servers as done in the Video Development Initiative [3].

H.323 terminals may be used independent of servers for bilateral conferences. In this way MS Netmeeting a.o. operate. The serverless extension to multipoint abilities in the IP world are most efficiently done via multicast transport, where any client in the conference simultaneously takes the role of multicast source and destination. Multicasting is employed at the price of communicating in more or less full public. Multicast features do not conform to H.323 and have been implemented e.g. by the Mbone Tools [4], Vcon [5], Ivisit [6].

User location services of the available conference tools remain rudimentary. Beside direct addressing of manually discovered devices and static listings some terminals can connect to a directory server and dynamically update user locations. This can be done for example with Netmeeting and the MS Internet Locator Server [6]. In this way a conference attendee may select partners from people currently registered at his previously selected directory server. The SDR Mbone tool – though attained through advertising multicasts – exhibits a similar behaviour. The communicative aspect of these services, however, remains far from a steering at will and comparable to chat groups.

The problem yet to be solved concerns strategies of locating appropriate services and to contact a communication partner at will on a global scale. Thereby, in order to ensure a short-term success, any solution should not involve changes to the present Internet information structure. A fairly general attempt has been made along with the Session Initialisation Protocol (SIP) [8]. Roughly speaking a global addressing scheme of the form <user>@<SIP-server>, where the SIP-server contains a name mapping directory learned from client registrations or proactively driven by unspecified server inquiries, is proposed. The SIP concept proposes either a significant roll-out of SIP selflearning infrastructure or just the presence of single, isolated information servers. In the latter case strategies on locating these information servers remain vague. In the following section we will introduce a mechanism similar to second type in SIP that precisely specifies location strategies and operates without inventing new addresses.

2.2. A Global User Location Scheme

Video conferencing is a heavily-weight, synchronous form of communication requesting online presence of the participants. To retrieve the information on how to direct data flows to the appropriate user’s device a dynamic user session recording has proven advantageous. In our system introduced here we denote this by User Session Locator (USL) and store appropriate session information within an LDAP directory server.

The video conference clients update information about ongoing sessions regularly so that outdated session records can be identified by their timestamps. The USL server can be arranged within a local infrastructure not only to enhance scalability by distribution, but also to adopt local knowledge of the identity of users as well as a method for authentication. Note the importance of authentification procedures for user session registration: Private communication channels are directed by advertising user session data. Whereas a local search of the USL server can be performed in a straight forward fashion (see Fig. 1), the global user look-up problem reduces to deciding on a unique user addressing and discovering the appropriate directory server for a given address.

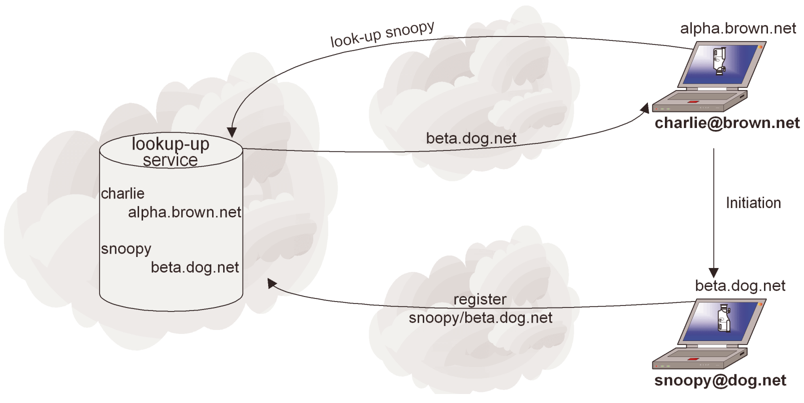

Figure 1: Centralised user look-up

Currently the only uniformly available user addressing scheme throughout the Internet is given by mail. Mail addresses are not only globally unique but also device independent, commonly known or easily retrieved. The uniqueness and popularity of mail naming has been noticed by several vendors so that calling a video conference user by his mail name has gained some popularity. With respect to its feasibility and ease of use our system restricts user addressing to mail addresses.

The Internet mail system, though, provides a mechanism for resolving user location due to its interaction with the Domain Name System via the MX record type for referencing a mail exchanger. Following this example the appropriate proposition for session based services would call for a new DNS service record pointing at the USL directory server for a given domain name. The extension of the DNS by SRV records has been proposed in RFC 2782 [9] and referred to in [8]. It requests for a change in Internet information structure, though, and remains at present stage proposed. Similarly, but with less significant changes in Internet naming, the DNS TXT record could be employed to store the location of a USL look-up server as proposed in RFC 1464 [10].

Since these two approaches beside their straightness imply global modifications on DNS content structure which cannot be easily achieved we chose a much simpler strategy. DNS data provided today are ready to cope with it: As the mail exchange record indicates a domain where any requested user is identifiable along with a method of authentication it is the appropriate location for a USL server. Within this domain

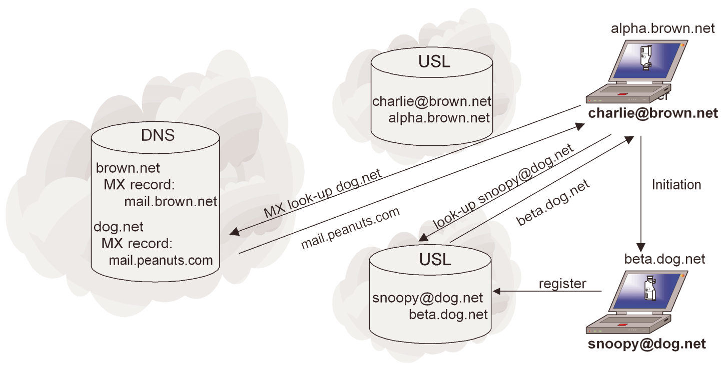

Figure 2: Distributed User Location Scheme

the look-up server can be identified by the common approach of a naming convention, i.e. *usl.<mailexchanger-domain>.* Consequently, a global user look-up proceeds in two steps. Firstly, the MX record for the target user is inquired, secondly the directory server hostname formed from along the above naming convention has to be resolved (see fig. 2).

Though simple, this user session information architecture neither relies on infrastructural changes nor requires dedicated user knowledge on the application side. It is easily integrated into existing local infrastructure and may push video conferencing into establishing a serious, regular Internet communication service.

2.3. A Directory Schema in LDAP

The definition and implementation of an appropriate directory schema for conferencing services [15] bears essentially four topics:

1. Integration into global naming structures to provide world wide user tracking.
2. Integration into local directory structures.
3. Scalability.
4. Definition of actual conferencing session data.

In following the lookup strategy defined in the previous section we omit problems one and three. Our user lookup scheme does not require a global directory schema and is thus left with local directories of limited size and complexity. A data definition for the description of conferencing sessions suitable in our case looks as follows:

```
DN:
dn: mail=charly@brown.net,dc=application

Attributes:
objectclass (< OID > NAME ‘VCoIP’ SUP top AUXILIARY
      DESC ‘Video Conferencing over IP Session Information’
      MUST ( VCoIPipHostNumber $ VCoIPipServicePort $ VCoIPServiceProtocol
         $ VCoIPTimeStamp $ mail $ cn
      )
      MAY ( VCoIPMcastGroup $ VCoIPAppID $ VCoIPAppVer $ VCoIPAppProtocol
         $ VCoIPMimeType $ VCoIPPrivateipHostNumber $
         VCoIPPrivateipServicePort $ VCoIPStatusFlag
      )
)
```

Integration into local LDAP directory services then can be easily achieved through a server referral.

2.4. A Word on NAT

Many potential users may be located behind a Network Address Translation Gateway, NAT-GW, and thereby be excluded from any peer-to-peer video communication system. This of course is equally true for H.323-based solutions which carry an additional burden in signalling connection data within separate control sessions. To cure how NAT breaks such applications, an Application Layer Gateway, ALG, commonly needs to be implemented directly on the NAT-GW. Even though major vendors offer H.323 ALGs, much of the sustainable success of Internet applications is hindered if they cannot run on endpoints without first requiring upgrades to infrastructure components. Even though NAT-GWs are expected to disappear with a change to IPv6 protocol discussions on how to overcome NAT-GWs increase throughout the Internet community [16].

To keep up with our goal of spreading out VCoIP on the given, unmodified infrastructure even in the presence of NAT, we proceed as follows: Working behind a NAT-GW the USL needs to be installed outside the NAT range. Since our system signals and receives media streams on a single network port, which can be tcp or udp with similar qualitative performance, we proceed through the NAT to contact the USL via tcp. We then preserve this connection in order to restrain the NAT-GW from dropping its state information, extract address and port from the packet headers and publish them to the USL directory. By following this procedure infrastructure remains completely untouched while any caller from the public Internet will meet addressable connection data to initiate a video conference session. Note that this NAT work-around could be achieved for udp-based communication in a similar fashion.

3. The daViCo Video Conferencing System

3.1. Overview

The digital audio-visual conferencing system daVico [11] forms a serverless multipoint video conferencing *software* (see fig. 3). It has been designed in a peer-to-peer model as an preferably lightweight Internet conferencing tool aiming at effortless use. Guided by the latter principle daViCo abstained from implementing H.323 client requirements.

The system is built instead upon a fast, highly efficient video codec, based on a wavelet algorithm. Exploiting specific properties of the coding scheme the software permits a scaling in bandwidths from 64 to 4000 kbit/s. Audio data is compressed using an MP3 algorithm with latencies below 120 ms depending on buffersize. Audio and video streams can be transmitted as unicast as well as multicast. An application sharing facility is included for collaboration and teleteaching.

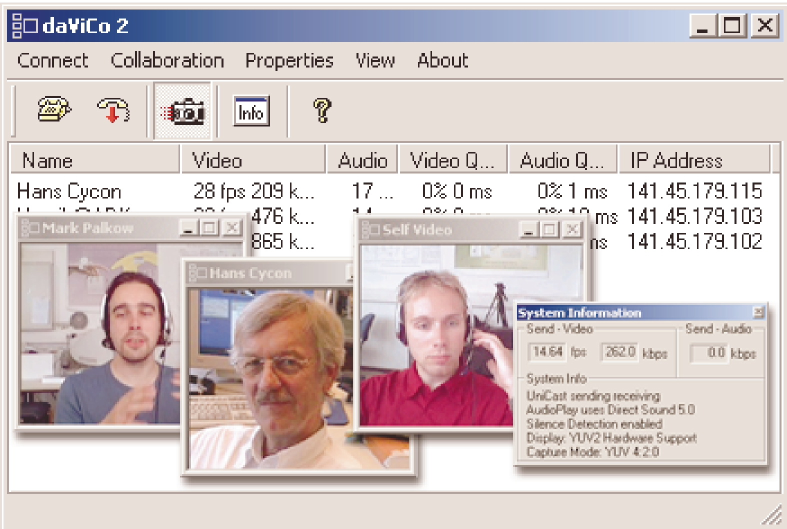

Figure 3: The daViCo Conferencing Tool

Due to low bandwidths requirements daViCo is well suited for long distance video conferences on best effort basis. To strengthen its global usability the user location scheme described above has become part of the software.

3.2. Wavelet-Based Real-time Video Codec

Transformation and Quantizer

The real-time video codec is based on a fast low complexity wavelet transformation. Transformation coding usually consists of three modules: A lossless transformation which decorrelates the signal, a quantizer and a lossless entropy coder which compactifies the data produced by the quantizer (see fig. 4). The transformation we use is of wavelet type transforming the image as a whole. Thus no blocking artefacts occur.

Filtering is done in a low complexity implementation with a 5/3 tab convolution – subsampling on three levels. As quantizer we chose a simple uniform scalar with enlarged dead zone. The third module is a highly efficient fast entropy codec scheme consisting of a precoder (PC) and a set of Golomb Rice codecs. To reduce the temporal redundancies in a video sequence we use DPCM coding, i.e. the difference from a frame to the next will be coded only.

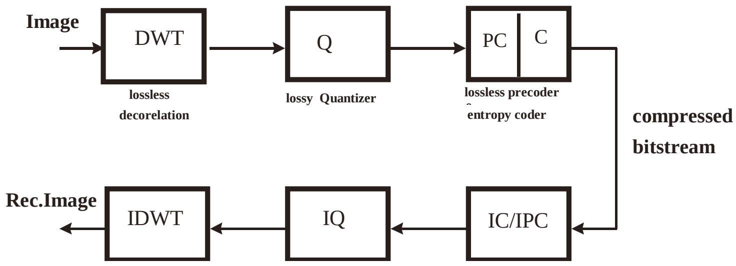

Figure 4: Transformcoding

For encoding the quantized wavelet coefficients, we follow the conceptual ideas presented in [12]. For more details, the readers are referred to [12], [13].

Results

In native implementations the video codec encodes and decodes 25 CIF frames (352 x 288 pixels) simultaneously on a 500 MHz Pentium machine. Alternatively 5 frames in PAL (720 x 576) resolution may be processed, where framerate is expected to increase with forthcoming algorithmic improvements. The image quality is better or comparable with MPEG 4 / H.263 Coders. On moderate motion complexity this frame rate produces a bit rate of ca 200 kb/s while sustaining very good visual quality.

The codec also has been ported to JAVA as part of a Web streaming system [14]. The JAVA codec running in an applet still decodes or encodes 5 CIF frames per second in real-time or - more appropriately - QCIF format with 25 frames.

4. Conclusions and Outlook

Video Conferencing over IP offers an opportunity beyond well known communication methods such as synchronous telephony or asynchronous mail. It thereby exhibits an enormous potential to transforming into a regular standard service throughout the Internet. However, the distribution of VCoIP presently is retarded through the fact that common approaches rely on significant changes of Internet infrastructure.

We presented a proposition, both communication framework and conferencing software, to overcome these obstacles from the lightweight side. The solution currently has been newly rolled out within our institution. First experiences support our conjecture of sustaining acceptance by ease of use.

Future development of our system will more closely evolve according to standards. The advancement of our video codec PACC will be part of the ITU-T standard H.264 resp. the MPEG standard 'Advanced Video Codec' (AVC). As soon as the DNS service record [9] will be established, user service locators will be denoted therein.

Acknowledgement:

We would like to thank Stefan Zech for most joyful collaboration: His tricks pushed some Windows into networking.

References:

[1] ITU-T Recommendation H.323: *Infrastructure of audio-visual services – Systems and terminal equipment for audio-visual services: Packet-based multimedia communications systems*. Draft Version 4, 2000.
[2] E. Verharen: *Development of a European Videoconferencing Service*. Proceedings of TERENA 2001 Networking Conference, http;//www.terena.nl/conf/tnc2001/proceedings, 2001.
[3] Video Development Initiative, homepage http://www.vide.net, 2002.
[4] Mbone tool download ftp://ftp.ee.lbl.gov/conferencing/, 1996.
[5] The VCON homepage: http://www.vcon.com, 2002.
[6] The IVISIT homepage: http://www.ivisit.com, 2002.
[7] *NetMeeting Resource Kit Contents*, Chapter 3, Finding People http://www.microsoft.com/Windows/NetMeeting/Corp/ResKit/Chapter3/default.asp, 2002.
[8] M. Handley, H. Schulzrinne, E. Schooler, J. Rosenberg: *SIP: Session Initiation Protocol.* RFC2543, March 1999.
[9] A. Gulbrandsen ; P. Vixie ; L. Esibov: *A DNS RR for specifying the location of services (DNS SRV).* RFC2782, February 2000.
[10] R. Rosenbaum: Using the Domain Name System To Store Arbitrary String Attributes. RFC1464, May 1993.
[11] The daViCo homepage: http://www.daViCo-gmbh.de, 2002.
[12] D. Marpe and H. L. Cycon: *Efficient Pre-Coding Techniques for Wavelet-Based Image Compression*, 1997, Proc. PCS '97, pp. 45–50.
[13] D. Marpe and H. L. Cycon: *Very Low Bit-Rate Video Coding Using Wavelet-Based Tech-niques*, IEEE Trans. on Circ. and Sys. for Video Techn., 1999, 9 (1), pp. 85–94.
[14] B. Feustel, T.C. Schmidt: *Media Objects in Time --- A Multimedia Streaming System*. Proc. of the TERENA Networking Conference '01. Computer Networks, 37/6, November 2001, pp 727--735, Amsterdam 2001.
[15] A. Sears: *A Scalable Directory Schema in LDAP for Integrated Conferencing Services.* Proc. of Inet97. http://www.isoc.org/inet97/proceedings, 1997.
[16] M. Shore et al.: *The Middlebox Communication (midcom) Group.* http://www.ietf.org/html.charters/midcom-charter.html, 11-Mar-2002.