Global Serverless Videoconferencing over IP[U]

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.

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 complicated 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.

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

1. Introduction

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

In adopting the Internet protocol standards as the underlying, commonly available communication infrastructure, Videoconferencing over IP (VCoIP) can soon be expected to be widely available. When heading towards VCoIP as a standard Internet service, important steps have to be taken to ensure usability on a global scale. Any requirement of specific hardware or dedicated networking infrastructure is

[U] This work was supported in part by the EFRE Programme of the European Commission.

likely to hinder VCoIP roll-out and should therefore be avoided.

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

In contrast, VCoIP is embedded into general Internet-connected working devices and is today oriented towards more or less public conference groups. As employment of VCoIP grows more mature, the need for meeting-oriented, private sessions has to be met urgently. Since it addresses people rather than 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 serverless, highly efficient VCoIP software, implementing our information strategy. Our solution rigorously aims for ease and functionality at the price of loss of generality.

This paper is organised as follows. In section 2 we discuss communication strategies, introducing our basic ideas and examples of related work. Section 3 presents the daViCo videoconference software and its core technologies. Finally, section 4 is dedicated to conclusions and a look at practical experiences of the solution.

2. A Distributed Global Communication Framework

2.1. VCoIP Architectures and Related Works

Videoconferencing 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.

Videoconference communication is a person-oriented service. As the Internet in general accounts for location independent access to roaming 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 is presently organised in a similar fashion, with the significant difference that mail is a lightweight, 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 video streams 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 common MCU and Gatekeeper servers which, 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 centre on the ideas of telephone-based wide area connectivity and are made obsolete by the simple observation that the use of videoconferencing via telephony is not growing. Consequently, attempts are made to overcome local restrictions in addressing by interconnecting Gatekeepers via meta-directory servers, as done in the Video Development Initiative [3].

H.323 terminals may be used independently of servers for bilateral conferences. MS Netmeeting and others operate in this way. The serverless extension to multipoint capabilities in the IP world is 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, for example, 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 similar behaviour. However, the communicative aspect of these services remains far from a self-steering and is comparable to chat groups.

The problem yet to be solved concerns strategies of locating appropriate services and contacting a communication partner at will on a global scale. Thereby, in order to ensure short-term success, no solution should involve changes to the present Internet information structure.

A fairly general attempt has been made with the Session Initialisation Protocol (SIP) [8]. SIP covers beside user localisation negotiations about user capabilities, user availability, the call set-up by SDP and the handling of the calls itself. SIP introduces its own infrastructure of servers which actively communicate by using SIP-URLs or other network protocols such as ICMP. SIP is open to store persistent information in common databases such as LDAP directories, but adheres its own server communication layer.

SIP does not prescribe a specific addressing scheme but proposes addresses 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. If addresses not of the SIP-server-type are used, the server will perform a user address based routing throughout the distributed SIP databases (see fig. 2). SIP does not provide mechanisms to ensure success in locating a user or a SIP-server present on the network. It should be noted that SIP-server addresses cannot be guessed from mail addresses as soon as virtual users tables names without reflecting underlying infrastructure are used.

The SIP concept proposes either a significant roll-out of SIP self-learning, interrelated infrastructure or just the presence of single, isolated information servers. In the latter case, strategies to locate these information servers remain vague. Both SIP and H.323 have the drawback of exchanging addresses within the protocol payload and are thereby severely hindered in NAT traversal as well as in migration to IPv6.

In the following section we will introduce a mechanism covering the user location part of SIP that precisely specifies location strategies and operates without inventing new addresses or protocols.

2.2. A Global User Location Scheme

Videoconferencing is a heavyweight, 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 the system introduced here, we denote this by a User Session Locator (USL) and store appropriate session information in an LDAP directory server.

The videoconference 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 authentication procedures for user session registration: private communication channels are directed by advertising user session data. Also, authenticated user session data may serve as a weak mechanism of identification: a callee may verify an agreement of IP and the user address of the caller by searching the USL session registration in a trusted domain. 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 is reduced to deciding on unique user addressing and discovering the appropriate directory server for a given address. Without further constrains on addresses or names this problem is equivalent to performing an Internet wide user based routing as is the purpose of the SIP server infrastructure (see fig. 2).

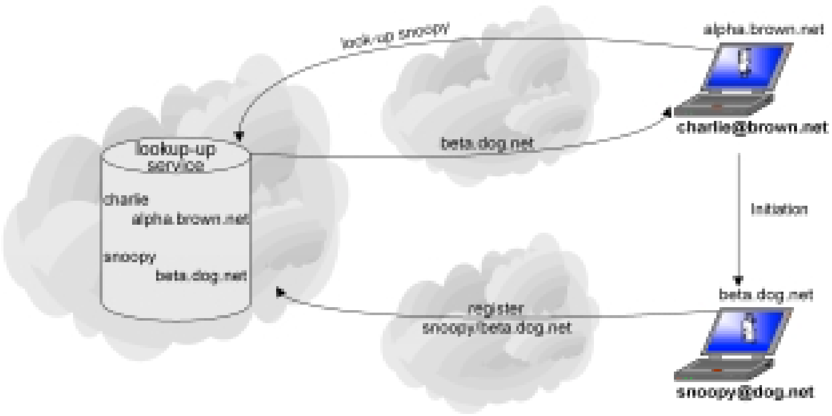

Figure 1: Centralised user look-up

Currently, the only uniformly available user addressing scheme on the Internet is given by mail. Mail addresses are not only globally unique but also device independent, commonly known or easily retrieved. Several vendors have noticed the uniqueness and popularity of mail naming , so that calling a videoconference user by his mail name has gained some popularity. Our system restricts user addressing to mail addresses because of its convenience and ease of use. In adopting this restriction we radically break with telephone compatibility.

But the Internet mail system provides a mechanism for resolving user location through 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 is referred to in [8]. However, it requests a change in Internet information structure at present stage and remains a proposal. 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].

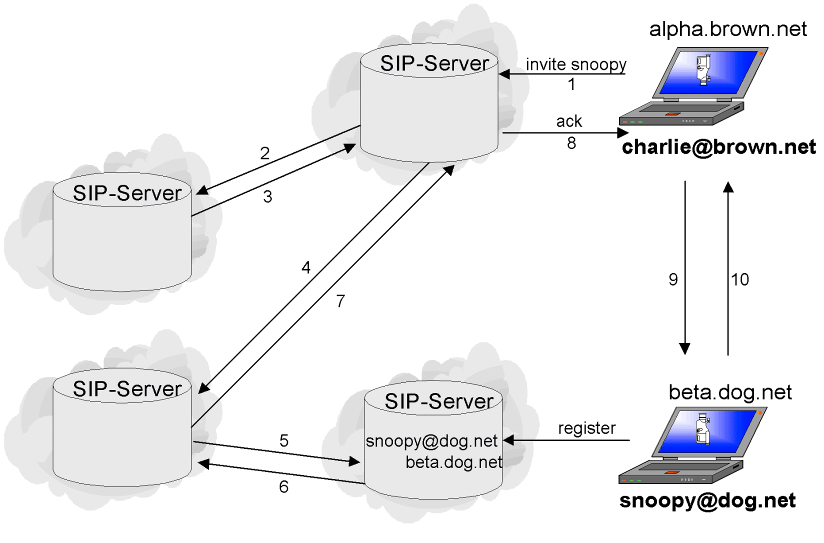

Figure 2: SIP user address based routing

Because these two approaches, despite their straightforwardness, imply global modifications on DNS content structure that cannot be easily achieved, we chose a much simpler strategy. DNS data provided today are ready to cope with it: because the mail exchange record indicates a physically present 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, the look-up server can be identified by the common approach of a naming convention, i.e. *usl.<mailexchanger-domain>* [17]. Consequently, a global user look-up proceeds in two steps. Firstly, the MX record for the target user is requested, and secondly, the directory server hostname formed from the above naming convention is resolved (see fig. 3).

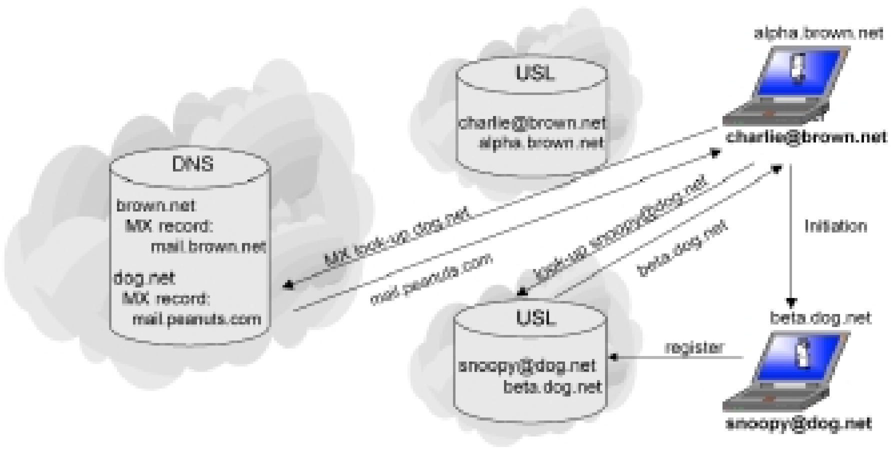

Figure 3: Distributed User Location Scheme

Though simple, this user session information architecture neither relies on infrastructural changes nor requires dedicated user knowledge on the application side. Note that in contrast to H.323 gatekeepers or SIP servers the USL server consists of a passive session record store and can be realised by an unmodified standard LDAP server such as OpenLDAP. It is easily integrated into existing local infrastructure and may establish videoconferencing as 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 issues:

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

By following the lookup strategy defined in the previous section, we omit issues 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 appears as follows:

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 would thereby be excluded from any peer-to-peer video communication system. Of course, this is equally true for H.323 or SIP-based solutions which carry an additional burden in signalling connection data within

```
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
      )
)
```

separate control sessions. To correct the way that 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 the change to the IPv6 protocol, discussions on how to overcome NAT-GWs are increasing throughout the Internet community [16].

To achieve our goal extending VCoIP on the given, unmodified infrastructure, even with 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, the infrastructure remains completely untouched, while any caller from the public Internet will obtain addressable connection data to initiate a videoconference 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 serverless multipoint video conferencing *software* (see fig. 4). It has been designed in a peer-to-peer model as a lightweight Internet conferencing tool aimed at effortless use. Guided by the latter principle, daViCo refrained 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 scaling in bandwidths from 64 to 4000 kbit/s. Audio data is compressed using an MP3 algorithm with latencies below 120 ms depending on buffer size. Audio and video streams can be transmitted as unicast as well as multicast. An application- sharing facility is included for collaboration and teleteaching.

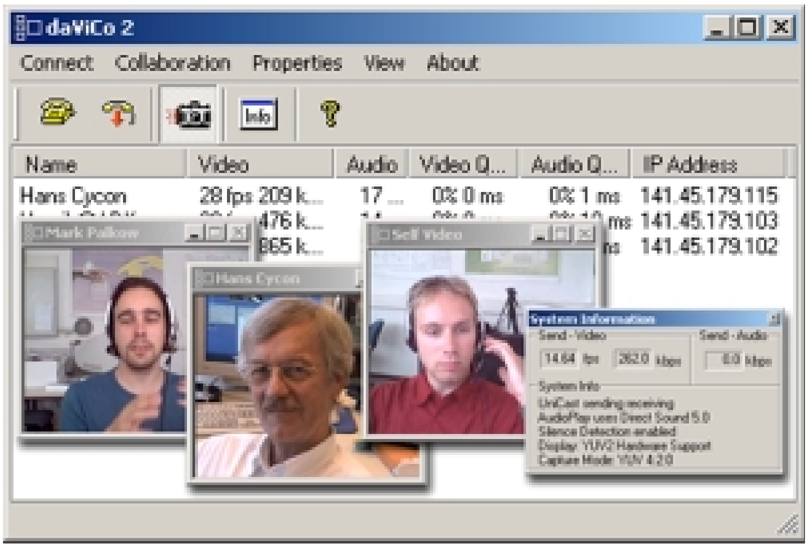

Figure 4: The daViCo Conferencing Tool

Due to low bandwidth requirements, daViCo is well suited to long distance video-conferences on a 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 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 compacts the data produced by the quantizer (see fig. 5). The transformation we use is wavelet-type, transforming the image as a whole. Thus, no blocking artefacts occur.

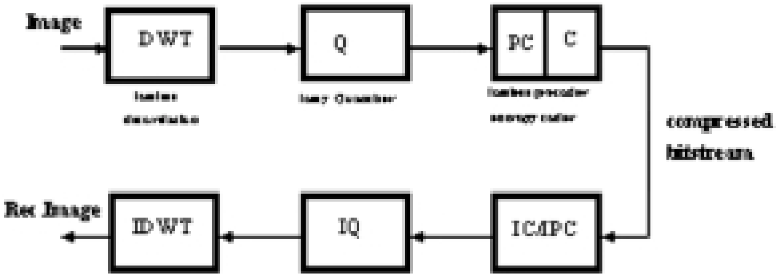

Figure 5: Transformcoding

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 an 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.,only the difference from one frame to the next will be coded.

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 frame rate is expected to increase with forthcoming algorithmic improvements. The image quality is better or comparable with MPEG 4 / H.263 Coders. At moderate motion complexity, this frame rate produces a bit rate of ca 200 kb/s while sustaining very good visual quality.

The codec has also 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

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

We present a proposition, both communication framework and conferencing software, to overcome these obstacles with a lightweight solution. The current solution has been recently rolled out within our institution. First experiences support our conjecture of sustaining acceptance by ease of use.

The future development of our system will evolve according to standards. The advancement of our video codec will be part of the ITU-T standard H.264 or the MPEG standard 'Advanced Video Codec' (AVC), respectively. As soon as the DNS service record [9] is broadly established, user service locators will be denoted therein. Currently the application is ported to IPv6.

Acknowledgement:

We would like to thank Stefan Zech for his cheerful 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.
[17] M. Hamilton, R. Wright: *Use of DNS Aliases for Network Services,* RFC2219, October 1997.