[APWS99] *Ralf Ackermann, Jörg Pommnitz, Lars Wolf, Ralf Steinmetz:* **MBone2Tel - ein Gateway für die Teilnahme von Nutzern konventioneller Telefonendgeräte an MBone-Konferenzen;** ITG/TKTG-Fachtagung Multimedia: Anwendungen, Technologie, Systeme, 8. Dortmunder Fernsehseminar, Dortmund, September 1999, S. 237-240.

MBone2Tel – Telephone Users Meeting the MBone

Ralf Ackermann[1], *Jörg Pommnitz*[2], *Lars C. Wolf*[1], *Ralf Steinmetz*[1,2]

1 KOM
Darmstadt University of Technology
Merckstr. 25 • D-64283 Darmstadt • Germany

2 GMD IPSI
Dolivostr. 15 • D-64293 Darmstadt • Germany

{Ralf Ackermann, Lars.Wolf, Ralf.Steinmetz}@kom.tu-darmstadt.de
Joerg.Pommnitz@darmstadt.gmd.de

Abstract: The integration of the Internet with the existing PSTN found much interest recently. While in many cases the aim is to use the Internet as a carrier medium for voice phone calls, we will present a proposal to extend classic Internet multimedia services to users on the PSTN.

We introduce an architecture and prototype implementation that allow users on the PSTN to actively participate in MBone audio conferences. We present a set of building blocks for applications using voice access via telephones and give an overview of how these can be combined to meet individual requirements. Finally, we discuss application scenarios, usage experiences and potential future enhancements to improve usability and to extend the range of available services.

Keywords: Multimedia Gateway Technology, MBone, Public Switched Telephone Network and Internet Internetworking.

1 Introduction

The MBone forms the IP Multicast Backbone [9] of the Internet. In operation since 1992, it uses multicast directly and connects networks capable of supporting it via unicast IP tunnels. Originally meant to be a research tool in the multicast area it quickly became a valuable resource for users outside the network research area as well.

Today it is widely used to distribute multimedia contents like audio and video to recipients in diverse environments ranging from research and education to business and entertainment all around the world. Though it has been used for some years now, IP multicast is still considered a somewhat uncommon feature and is not yet available to all users on all platforms. One of the technical reasons for this is the difficulty to send multicast traffic over point-to-point dial-up links to end nodes. Since many people access the Internet over point-to-point dial-up links, they have to establish IP-IP tunnels across their access link. This imposes an additional burden on the Internet Service Provider (ISP) and the end user.

Additionally, in many situations, for instance for mobile users, access bandwidth, hardware resources or the available software for the used devices are not sufficient for using attractive MBone services directly, though the user may have – at least temporarily – some sort of connectivity to the Internet. In these cases, even applications like mTunnel [10] which allow to easily build up and use multicast tunnels are no adequate means.

One of the most attractive MBone services is audio (and video) conferencing using tools such as the "Robust Audio Tool" rat [11] or vat. Our work was initially inspired by the lack of an adequate access facility for our research staff members. When they worked at home but tried to take part in a scientific conference transmitted via the MBone, we actually had to attach the handset of a conventional phone to a workstation loudspeaker manually.

The idea came up, that if it would be possible to access MBone audio conferences from standard Public Switched Telephone Network (PSTN) terminals (i.e. simple phones or video phones) in a convenient and comfortable way, the possible audience for MBone distributed content could be significantly increased. Hence, the MBone2Tel Gateway (Figure 1) described in this paper is designed to bridge the gap between the MBone and the PSTN.

After a discussion of related issues in the next section, we will describe the concept and architecture of the gateway in Section 3 before we focus on both the implementation of audio forwarding (Section 4) and the control facilities (Section 5). Finally, usage scenarios and possible extensions of the Gateway are shown.

2 Related Work

Over the last years many MBone based applications have been developed, covering the domain of audio and video conferencing including archival of MBone sessions and the reliable and efficient distribution of data to many customers.

Much of that work has been done as part of the MERCI (Multimedia European Research Conferencing Integration) [2] project which resulted in the availability of powerful applications like the *Robust-Audio Tool* (rat). These tools can cope with varying network conditions causing jitter, corruption or packet loss and still deliver a reasonable good quality to the end user and interoperate with other solutions. Their main drawback is that they are limited to a pure MBone environment and cannot be used by plain PSTN users.

The interworking between MBone tools and applications implementing the ITU recommendations of the T.12x and H.3xx series is regarded to be of high importance, as stated e.g. in [5]. The ITU recommendations describe the implementation of multipoint multimedia conferences both for circuit switched networks as well as for packet switched ones. Especially the H.323 standard [8] is one of the most important protocol candidates within the evolving field of IP telephony.

To provide interoperability of the MBone and the PSTN world, work has been carried out as part of research projects such as MERCI as well as by individual commercial initiatives [18]. These efforts resulted in the design, development and public availability of applications and services [19] which are capable of forwarding multimedia content between the MBone and the PSTN H.3xx domain while also providing means for multiple access and an adequate mapping of control semantics. While this already broadens the coverage of potential users, it still leaves those without access to H.3xx/T.12x equipment without support. These can be users who are either not attached via the necessary access lines, such as mobile users with a GSM phone or persons who – probably even more often – may not use the appropriate applications due to hard- or software restrictions. Therefore, our implementation does not compete with approaches that bridge into the T.12x or H.3xx world directly. Our work is orthogonal and covers a large and yet still mostly unsupported area.

Actually there are many efforts to use the Internet as a carrier medium for telephone calls and to integrate the usage of the PSTN with Internet applications. The IETF working group "Internet Telephony" (iptel) [12] concentrates on the engineering of protocols to be used for IP telephony applications and signalling as well as on the deployment and interaction of gateways between the Internet and the PSTN. This primarily focuses on the transport of telephone traffic via the IP infrastructure.

The "PSTN-Internet Internetworking" (pint) [13] working group describes application scenarios and frameworks such as "Click to dial" and "Voice access to Web content". Many of the building blocks we implemented can easily be adapted to be used for providing a selection of these services as well.

In summary, while these approaches provide access to MBone sessions from H.32x devices resp. use the Internet as transport medium for phone calls, our work adds the support for phone users to access MBone sessions.

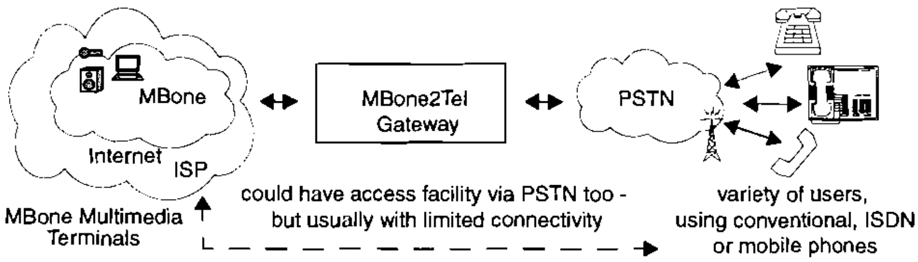

Fig. 1. MBone2Tel Gateway

3 The MBone2Tel Gateway

3.1 Intended Features and Design Considerations

The design of the MBone2Tel Gateway was driven by the following goals:

- implement the basic service of forwarding audio content bidirectionally between MBone audio conferencing sessions and the PSTN,
- do not restrict the intended audience by means of special hardware or software requirements
- keep the operational overhead for a regular service as small as possible while still retaining a maximum of configurability,
- develope components which can easily be adapted and combined to support further scenarios such as dial-up voice access to Email or active notifications on the occurrence of certain events.

There are various well established MBone and other audio conferencing tools. We decided to use already available and proven components and concentrate on enhancing and integrating them in a new way. By keeping the interfaces towards these applications as small and universal as possible, we can also benefit from advances in these base tools.

As shown in Figure 1, the MBone2Tel Gateway is located between the Internet and the PSTN and is connected to both of them. A regular phone user makes a standard PSTN call to the gateway and can participate in a MBone session. Figure 2 shows a more detailed view of the gateway and its building blocks as well as their interactions.

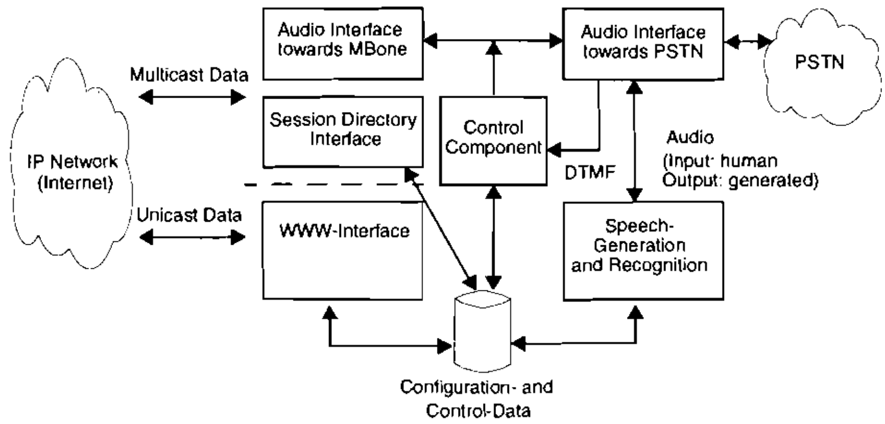

Fig. 2. Architecture of the MBone2Tel Gateway

3.2 Basic operation

When a user with a conventional phone dials up the gateway, the Audio Interface towards the PSTN accepts the call and presents a short introduction message and a description of the available services is generated on the fly by a speech generation component. All the control information is locally stored and is dynamically updated either by the system's operator or automatically through interacting with the Session Directory Interface that permanently receives and decodes Session Announcements from the MBone. The caller is then guided through an audio menu and may select his further operation either by Dial Tone Multi-Frequency (DTMF) tones, by giving single word answers to questions or by naming a MBone session directly through its multicast address and port.

When a valid session is chosen, an enhanced "rat" program is started with the corresponding parameters and audio data is forwarded bidirectionally through the Audio Interface towards the MBone. After the connection is set up that way it can be used to take part in the selected MBone audio conference as passive listener or even as an active talker and the configuration database is updated to represent the current state of the gateway.

All the control and management information can also be accessed via a WWW and a Java based interface that is used locally as well as from the Internet via a unicast connection. The management of the system is also done this way.

4 Audio Forwarding

The prototype implementation of our MBone2Tel Gateway has been done on the Linux operating system and largely benefits from the features the system supports for accessing the PSTN via ISDN-cards or voice modems.

The part responsible for audio forwarding (see Figure 3) is basically split into two components. One component, the audio interface towards the MBone, receives and decodes the incoming data stream from the network in the receiving case or encodes and sends it when the phone user speaks. It is based on the popular MBone "Robust Audio Tool" (rat) which has been extended with a generic audio interface that is able to forward and receive audio data using the systems IPC mechanisms, namely Named Pipes. By using "rat" we profit from the work already done in the area of jitter compensation and redundant audio transmission as well as from its adaptivity to varying network conditions.

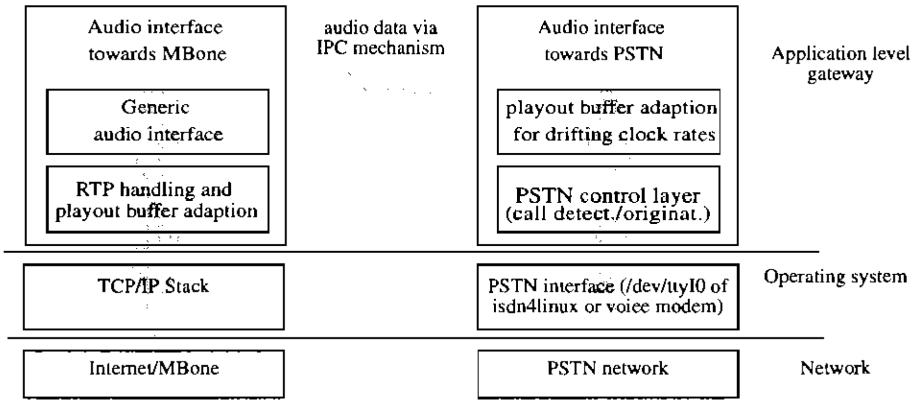

Fig. 3. Audio Forwarding – Components and Data Flow

The other component, the audio interface towards the PSTN, waits for incoming calls or originates them, establishes a connection and sends the audio data to the PSTN. Doing our experiments in Germany, where the ISDN system has a widespread deployment, we used the functionality provided by the ISDN kernel extension (isdn4linux) [14] of the Linux operating system. It provides a modem emulation accessible over a TTY-like interface. The code we implemented has been largely inspired by the "vbox" package [15], which implements a telephone answering machine on top of the isdn4linux facilities.

To keep compatibility with conventional analog modems and to have a means of providing the service also without ISDN, we decided to use the isdn4linux modem emulation interface and to strictly follow a layered approach. This modem emulation is fairly complete and provides extended commands for voice processing compatible with the popular ZyXEL voice command pseudo standard. It supports features like voice activity detection and real-time in-band dial tone (DTMF) decoding which can be used to trigger control operations.

Within the whole system – which also has interfaces towards a synthetic speech generation and a speaker independent single-word recognition module, that we will describe later on – we use a uniform 16 bit signed little endian audio format. Thus the components which all have to implement an interface towards this format can be combined in a very easy manner and form a kind of a tool box.

The independent timing used by the two components of the gateway system was a subtle problem. Forwarding audio data from the MBone to the PSTN layer and vice versa is typically application driven. In our scenario data has to be sent at a rate that corresponds to the sampling rate of the PSTN side (e.g. 8 kHz in the ISDN system). This rate should be the same for both parts, but the sampling clocks may drift apart. If the "producer" is too slow, the audio stream gets interrupted which is audible by annoying clicks. When too much data is arriving, buffers may overflow and audio information gets lost. We cope with these problems by using a playout buffer adaption algorithm.

While the telephone line as a dedicated medium is used continuously, this is not the case for the transmission medium in MBone sessions. The approach there is typically to only emit audio data packets when a speaker is really active. This can be indicated either explicitly by pressing some kind of Push To Talk (PTT) button or implicitly by a voice activity detection mechanism. Due to the limited control facilities when using a conventional handset, we used the voice activity feature of the MBone tool "rat".

5 Control Operation of the Gateway

For widespread and convenient use of the gateway appropriate control mechanisms are of crucial importance. While the described forwarding of audio content is typically determined by the characteristics of the connected systems and the encoding they use and therefore straightforward, there is a wide spectrum of potential control mechanisms. Basically a user wants to decide which MBone session he gets connected to and whether the gateway should play a passive role by just accepting calls or an active one by calling him itself.

Typically, there is a varying number of different MBone sessions that can be received at the gateway location at a time and might be of potential interest for a caller. There are multiple ways for a user to select one of these sessions. The various possibilities to control the operation of the gateway can be categorized along the following parameters:

- the direction in which descriptive data or control information is transferred,
- the media and encoding used for the descriptive and control data,
- the time when the information is presented or the control operation takes place.

This leads to the matrix of usage scenarios shown in Table 1. We show the third dimension (time) by distinguishing between white (if the information flow or control operation takes place before the actual session) and shaded (if the operation takes place during the session) table fields.

white – before the session	shaded – during the session

	data presented to the user	descriptive and control data originated by the user
in-band (using the session channel)	(1) Audio description of available sessions and navigation support in a voice guided menu – but in a call before the actual session itself.	(2) Control managed via an initial call to a dedicated service telephone number (which could even be operated by a human), while the gateway's MBone audio data service is provided by another telephone number.
	(3) Equivalent to (1) but as part of the session.	(4) Navigation through a menu by means of DTMF touchtones generated either by the telephone itself or an adequate external dialer. Additionally speaker independent single word recognition can be used to choose alternatives presented by the voice menu.
out-of-band	(5) Calls can be originated by the system whenever they are scheduled and the MBone session actually starts. Additionally all the features described under (7) may be used.	(6) MBone sessions can be initiated and advertised via a proxy interface to a session directory tool, thus not requiring direct MBone connectivity. Additionally all the features described under (8) may be used.
	(7) The announcement of available services (dial-up number according to a certain MBone session) can be done by means of Electronic Mail or a WWW-Interface.	(8) User may actively manipulate the assignment of a dial-up number to a MBone session, may give additional information (such as his real name) and may change transmissions parameters or activation of its channel using a WWW browser interface enhanced with a Java Applet.

Tab. 1. Systematization of Control Modes

All of the described cases are supported by the gateway and have been evaluated. We now describe the most common ones in more detail and give an overview of their implications and the software components that have been developed to support them.

5.1 Static Mapping between MBone Sessions and PSTN Numbers

In the case of a static mapping from MBone sessions to PSTN numbers, the session to be accessed through the gateway is fully determined by the telephone number the user dials up to and the time he does so. In a very basic implementation the provider of the gateway may thus statically assign a MBone session to a certain number and – as a first additional service – reject calls or present a corresponding informative audio message during the time the session is not yet or no longer active. This configuration, while usable and implementable without much additional effort, still has a considerable management overhead and an even more serious problem: How do potential users learn about this mapping? In order to solve this problem it is desirable to give users additional information and let them choose the session themselves.

5.2 User Controlled Session Mapping

With a user controlled session mapping the particular session which a user receives can either be predetermined before the call is done or may be chosen or even changed while the telephone connection is already established. In this approach descriptive information and control data may either be handled in-band during the connection or out-of-band through a call in advance or via a computer based interface. All variants will now be explained in more detail. We show that despite of our initial concentration on access by means of a simple telephone also the latter mode where a computer is used in addition to a phone may be very useful.

5.2.1 Integration of a Text-to-Speech, DTMF- and Single-Word-Recognition Component

For a regular service, it is desirable to

- have a means to inform the user about the session he will be attached to when he dials up a certain number,
- to guide him through navigation menus, where he can choose between several options
- or to inform him about errors that have occurred.

While some of this information is static and could be prerecorded and replayed whenever this is necessary, problems may arise for data that changes frequently such as, for instance, the MBone session announcements. Session announcements are encoded according to the rules of the Session Description Protocol SDP [4] and transmitted using the Session Announcement Protocol SAP [6]. A tools like the Session directory tool sdr [7] can be used to send and receive announcements.

We integrated a Text-to-Speech component which gives us the opportunity to generate audio information on the fly just by means of accessing text input files or commands. Since it is freely available and supports a rich set of languages and speaker voices, we use the Festival Speech Synthesis System [3] here. Due to its command-line interface, powerful Scheme-based scripting language and the free availability of all sources for research and educational purposes, it can be augmented to use the IPC mechanisms used in our gateway to attach various audio sources.

In our scenario, the speech synthesis component is used to generate audio from english text input and has been proven to work quite satisfying. It supports an operation mode that – after a short introduction message – guides a caller through a menu of available MBone sessions from which he may choose by signalling through DTMF tones.

Since DTMF tone generation is not always available, we incorporated the speaker-independent Single-Word-Recognition component "EARS" [16] into our system. It has been enhanced with the generic audio interface and may thus be trained and used via a telephone line. In the interest of a high recognition rate we limited the vocabulary to the answers "Yes" or "No" for decision questions, the ordinal numbers 0 to 9 and the word "dot". This way a user may choose to select a MBone session – that must be known to him – directly by naming its associated multicast address and port. If the recognition result is valid within the current context we trigger the selected operation, otherwise the question and its possible answers are presented to the user again. Doing both the training and usage experiments with a couple of different users we got an acceptable behavior of the recognition component though its utilization needs very disciplined callers and takes some time to get accustomed to.

After a session is chosen the gateway starts up the "rat" component and connects the audio input and output streams. This mode is suitable for users with a conventional phone and no further means of control. However, a selection-only mode requires the operator of the gateway to do a preselection of the MBone sessions presented, otherwise the number of alternatives is usually unacceptable large.

5.2.2 Management via a WWW-Interface

As stated before the number of MBone sessions available simultaneously at a site makes handling them just by means of DTMF or voice navigation quite uncomfortable or even impossible and typically requires a considerable human configuration effort for preselecting a reasonable subset. Therefore, in order to provide a regular service, we incorporated a WWW interface that allows to manage the system in an easy and yet very efficient way.

It reads session announcements that are received and decoded by the session directory tool sdr and enables a remote or local user to assign dial-in or dial-out connections to a certain MBone session by means of a Java applet.

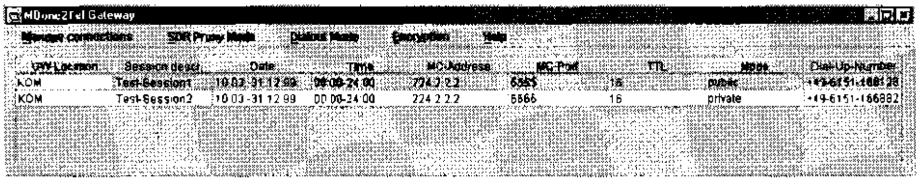

Fig. 4. Controlling the Gateway via a Java Interface

Though it may not be obvious from the very beginning, why a user who is directly connected to the Internet – which is a necessary prerequisite for using this management facility – should use the gateway, there are several reasonable scenarios, where this is quite desirable. Examples are mobile users, who scheduled the configuration in advance or are using a device such as a personal communicator via a GSM data connection, but have no means for taking part in the audio conference otherwise. The interface also allows the initiation of MBone sessions by means of a proxy session directory mode. This is important for users who want to originate a MBone session but have no means of accessing the MBone directly. Users may also be called by the Gateway at a preconfigured time when using the "Dialout Mode".

6 Conclusion and Future Work

The gateway has been in experimental operation for a few months at our site and has proven to work stable. It is mainly used in the operation mode incorporating the WWW control component that allows to configure it using the information that is dynamically generated by the session directory tool sdr.

We also envision a couple of other interesting application scenarios. Universities or local ISPs could establish dial-up nodes that allow their students or customers to follow lessons distributed over the MBone. This would be useful for tele-learning applications.

With an adequate deployment of gateways it is also possible to arrange ad-hoc multiparty telephone conferences by just using conventional phones and having the MBone build a kind of virtual Multipoint Control Unit (MCU) (Figure 5). Such a scenario is otherwise still difficult to set up since it usually needs to be planned and arranged with a MCU provider in advance.

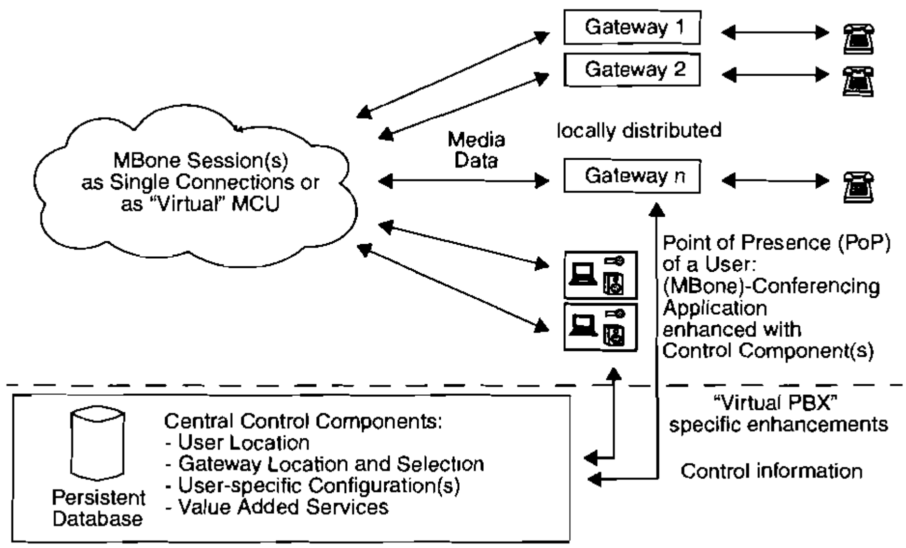

Fig. 5. Gateways Forming a Virtual MCU for Multiparty Telephone Conferences

The configuration can even be adapted to be used as a "Virtual PBX" [1], which in combination with an existing conventional telephony exchange provides additional services such as local user mobility.

When using a packet switched network for transmitting telephone calls, security considerations play an inherently important role for ensuring the callers privacy. This applies even more in our scenario where any user or malicious attacker can subscribe for receiving the corresponding multicast packets by just sending adequate IGMP messages. The audio tool "rat" supports closed group sessions by means of encrypting its audio data streams, whereas session keys have to be agreed upon, exchanged and typed in externally and all communication partners in a session have to use the same key.

Therefore, we suggest to incorporate public key mechanisms for determining and exchanging session keys to authorized components which have to prove their authenticity in advance.

We consider parts of the implementation as appropriate building blocks for future applications spanning a wide range and covering for instance audio access to mailboxes or web content. With the emerging use of IP telephony, the realization of value-added applications becomes a very reasonable challenge. So it is quite reasonable to extent the accessibility of the gateway to users of IP telephony applications as well.

The MBone2Tel Gateway is a convenient way to extend the range of potential participants in MBone audio conferences. We identified a number of additional usage scenarios where the gateway has experimentally been proven to be of great value. Currently we try to extend the possible range of interaction by incorporating an implementation of the H.323 protocol [17] developed as part of the Linux VOXILLA telecom project.

References

[1] R. Ackermann, J. Pommnitz, L. Wolf, R. Steinmetz. "Eine Virtuelle PBX", 1. GI-Workshop "Multicast-Protokolle und Anwendungen", Braunschweig, Mai 1999, S. 187-197

[2] R. Bennett, P. T. Kirstein. "Technical Innovations Deployed by the MERCI Project", Proc Networkshop 25, Belfast, March 1997, pages 181-189

[3] A. Black and P. Taylor. "Festival Speech Synthesis System: system documentation (1.1.1)" Human Communication Research Centre, Technical Report HCRC/TR-83, 1997, http://www.cstr.ed.ac.uk/projects/festival/festival.html

[4] A. Black and P. Taylor. "Festival Speech Synthesis System: system documentation (1.1.1)" Human Communication Research Centre, Technical Report HCRC/TR-83, 1997, http://www.cstr.ed.ac.uk/projects/festival/festival.html

[5] S. Clayman, B. Hestne, P. T. Kirstein. "The Interworking of Internet and ISDN Networks for Multimedia Conferencing", IOS Press 1995

[6] Mark Handley. "SAP: Session announcement protocol", Internet Draft, Internet Engineering Task Force, Nov. 1996. Work in progress

[7] Mark Handley. "The sdr Session Directory: An Mbone Conference Scheduling and Booking System", Department of Computer Science University College London Draft 1.1, 14th April 1996, http://www-mice.cs.ucl.ac.uk/mice-nsc/tools/sdr.html

[8] International Telecommunication Union. "Visual telephone systems and equipment for local area networks which provide a non-guaranteed quality of service", Recommendation H.323, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, May 1996

[9] Hans Erikson. "MBONE: The Multicast Backbone", Communications of the ACM, August 1994, Vol. 37, No. 8, pp. 54-60

[10] P. Parnes, K. Synnes, D. Schefström. "mTunnel: A Multicast Tunneling System with a User Based Quality-of-Service Model", 4th International Workshop, IDMS '97, Darmstadt, Sept. 97, pages 87-96

[11] Angela Sasse, Vicky Hardman, Isidor Kouvelas, Colin Perkins, Orion Hodson, Anna Watson, Mark Handley, Jon Crowcroft, Darren Harris, Anna Bouch, Marcus Iken, Kris Hasler, Socrates Varakliotis and Dimitrios Miras. "Rat (robust-audio tool)", 1995
http://www-mice.cs.ucl.ac.uk/multimedia/software/rat/

[12] "Charter of the IETF Working Group IP Telephony (iptel)"
http://www.ietf.org/html.charters/iptel-charter.html

[13] "Charter of the IETF Working Group PSTN and Internet Internetworking (pint)"
http://www.ietf.org/html.charters/pint-charter.html

[14] "FAQ for isdn4linux" – Version pre-1.0.5
http://www.lrz-muenchen.de/~ui161ab/www/isdn/

[15] "vbox – Anrufbeantworter für Linux"
ftp://ftp.franken.de/pub/isdn4linux/contributions/vbox-1.1.tgz

[16] "EARS: Single Word Recognition Package"
http://robotweb.ri.cmu.edu/comp.speech/Section6/Recognition/ears.html
(no longer maintained or supported by the original author)

[17] "OpenH323", Part of the Linux VOXILLA Telecom Project
http://www.openh323.org/

[18] "RADVision Products for IP Telephony and Multimedia Conferencing"
http://www.radvision.com/

[19] "The JANET Videoconferencing Switching Service"
http://www.ja.net/video/service/pilot_service.html