Design and implementation of a VoIP gateway

Arild Hansen

Department of Computer Science,

University of Tromsø,

Norway

Abstract

We describe the design and implementation of a Voice over IP gateway permitting the routing of voice traffic and control information between circuit switched telephone networks and packet based computer networks. The use of inexpensive hardware, a high-level programming language, and a flexible and modular design are key elements of the implementation. We identify the requirements for such an infrastructure, discuss the design choices and describe its implementation.

1 Introduction

The success of the Internet has stimulated the development of new technologies. Realtime voice transmissions over IP networks, commonly known as Voice over IP (VoIP), is one such service that is gaining the public's acceptance as an alternative to the telephone services provided by the Public Switched Telephone Network (PSTN). However, through its evolvement, the PSTN has manifested itself as an extremely reliable service. The rapid growth of packet based network technologies with promises of realtime services such as VoIP, will consequently not pose an immediate threat to the PSTN. The coexistence of these two infrastructures is thus required, and mechanisms are needed to combine these two different technologies into one unified and functional infrastructure. The objective of our effort is to identify the protocols, requirements, and design choices necessary to realize a VoIP gateway bridging a circuit switched network and a packet switched network. This gateway will enable voice data and signaling information to be transported across these different network technologies. The use of inexpensive hardware, a high-level programming language, and a modular and flexible design are all key elements of the implementation.

2 Media transmission and signaling

2.1 Circuit switched networks

Circuit switched networks such as the PSTN, use dedicated physical or virtual links to set up a connections between end-points. These circuits are logically divided into several channels using multiplexing techniques permitting one single physical link to transmit multiple media streams. This has the advantage of providing a

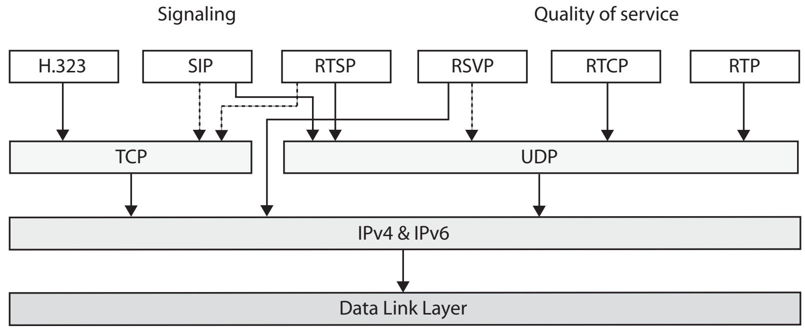

Figure 1: Internet telephone protocol stack.

guaranteed quality of service, but this assigned bandwidth comes at the cost of inefficient use of resources as the dedicated link occupies a relatively large amount of hardware independently of the actual traffic transmitted through the circuit. The establishment of a physical or virtual connection across the PSTN requires some sort of logical sequence of events performing the call-setup and "clear-down" procedure. These operations are generally referred to as *call control* functions [1] and include operations such as generating dial tones, performing number analysis and routing. R2 and the more recent Signaling System 7 (SS7) [4] are examples of signaling systems in use today.

2.2 Packet switched networks

The transport of data in packet switched networks is a complex interaction between different elements of the network. Data are divided into packets which are sent independently of each other, and the responsibilities of each network layer as depicted by the TCP/IP protocol suite, assure a sane delivery of packets at the receiving end. However, a telephone conversation contains large amounts of futile information such as silence and noise, and this information need not and should not be transported over the network. Packet switched networks allow valuable resources to be acquired only when the data being transmitted are important to the semantics of the realtime transmission.

Moreover, whereas the PSTN uses separate networks for signaling and media transmission, the packet switched networks must transport both these types of information using the same network. Another issue that must be handled is the heterogeneity of packet switched networks as opposed to the PSTN. IP networks require extended signaling due to the variation in end-point characteristics such as bandwidth requirements and media capabilities. Since the inherent diversity of IP networks allows for such variations, the signaling process must ensure that the end-points share the same capabilities before communication takes place.

3 Protocols related to VoIP

VoIP is an emerging technology where every standard ranging from signaling to media transmission and processing has yet to be agreed upon. The protocols shown in figure 1 are among the most prevailing today.

The application layer contains the higher-level protocols handling the communication details for isochronous VoIP applications. These protocols include mechanisms for setting up and releasing calls, call control, media transmission as well as media processing. They reside on top of the transport layer protocols such as TCP and UDP and provides end-to-end communication between VoIP applications and gateways.

3.1 Signaling using H.323

H.323 is part of a family of ITU-T recommendations called H.32x describing the architecture and operation of multimedia communication services over a variety of networks [2]. The protocol is built up of already defined ITU protocols and specifies the extended signaling and infrastructure required to establish multimedia sessions over packet switched networks. The protocol itself does not define any standards dealing with the actual media transmission and depends on other protocols to transport isochronous data streams and to get feedback on network quality. H.323 is independent of the packet switched network and the transport protocols over which it runs.

3.2 Signaling using SIP

The Session Initiation Protocol (SIP) is an application layer control protocol developed by the Multiparty Multimedia Session Control (MMUSIC) working group of the IETF [3]. This protocol is part of a complete framework of protocols intended to support Internet teleconference sessions. The protocol handles basic call signaling, user location, registration and call handling.

3.3 Media transmission using RTP

The Realtime Transport Protocol (RTP) developed by the IETF, is designed to provide end-to-end transmission of realtime data such as audio and video [6]. RTP is used in coherence with the RTP Control Protocol (RTCP) providing mechanisms for network monitoring including packet loss and jitter. The main features of RTP are the use of sequence numbers, time-stamps, loss detection and content identification to enable synchronization and correct playback of realtime data at the receiving end.

3.4 Resource management and QoS

The application layer also contains protocols providing resource management. These protocols are intended to give applications providing isochronous services a certain control over network resources. The Integrated Services (Intserv) and Differentiated services (Diffserv) of the IETF are examples of two such protocols.

4 Requirements

We aim at designing a VoIP gateway being as transparent as possible to the terminals connected to it. This necessitates an interface to both networks coherent with the media and signaling protocols of those networks. Moreover, we aim at designing a

VoIP gateway providing a framework compatible with a broad range of protocols and hardware. This envisions a gateway framework that separates the various functions into modules and third-party applications. New modules can thus be added as the requirements change. The following list identifies a set of requirements providing a framework for implementing a VoIP gateway suitable for research related to the VoIP technology.

1. The VoIP gateway *must* provide translation of voice data between the PSTN and the packet switched network. This translation *must* occur without introducing any unreasonable amounts of delay.

2. The VoIP gateway *should* provide translation of non-voice data such as DTMF tones and facsimile data between the PSTN and the packet switched network. This translation *must* occur so that the original data can be correctly regenerated at the receiving end.

3. The VoIP gateway *must* provide support for various media codecs providing functionality to reduce bandwidth usage and to improve QoS issues.

4. The VoIP gateway *must* provide translation of signaling and control messages between the PSTN and the packet switched network. The VoIP gateway *should* provide support for different signaling protocols so that interoperability with VoIP terminals using different signaling standards is possible.

5. The VoIP gateway *must* provide address translation compatible with the signaling protocol in use.

6. The VoIP gateway *must* be inter-operable with a broad range of hardware and allow for easy addition of new hardware.

7. The VoIP gateway *may* provide support for other network technologies such as Asynchronous Transfer Mode (ATM), Frame Relay and Integrated Services Digital Networks.

8. The VoIP gateway *should* be platform independent or at least easily portable to other operating systems and hardware platforms.

9. The VoIP gateway *should* have a modular design and be implemented using a high-level scripting or interpreted language to allow for easy modification and extension with respect to protocols and standards relevant to VoIP. This includes voice codecs, signaling and transmission protocols as well as network protocols.

10. The VoIP gateway *should* have support for or at least be easily modified to give support for mechanisms dealing with system management as well as advanced end-user services.

A VoIP gateway implemented according to these requirements will provide the necessary functionality to interconnect the PSTN and packet switched networks. Whereas some requirements are an absolute necessity, others are important to provide flexibility rather than functionality. However, to construct a gateway suitable for research purposes, all requirements should be included.

5 Design

The requirements specification outlined above encourages a framework for a VoIP gateway that separates the various functions into modules and third-party applications. This allows for easy extension and modification as the requirements change. Such a structure consists of the following logically separated components: a *media module*, a *signaling module*, a *system management module* and a *network protocol module*. This structure is shown in Figure 2. This model has been chosen because it logically categorizes all the major functions of a VoIP gateway. To fully realize the modular design depicted above, the VoIP gateway should also be implemented using an object-oriented programming language. In addition, to facilitate future modifications and extensions, the programming language should be a high-level scripting language with inherent support for the functionality and versatility outlined by the requirement specification.

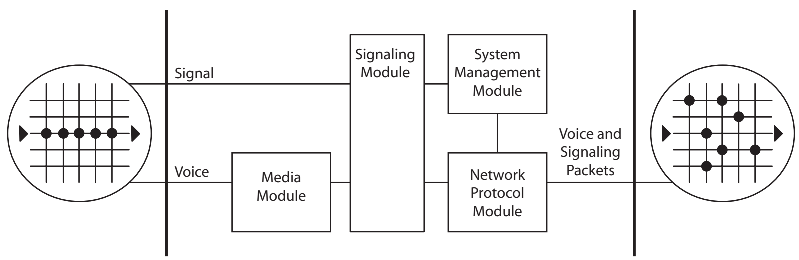

Figure 2: VoIP gateway architecture.

5.1 The media module

The media module constitutes the first logical entity of the VoIP gateway architecture. This module is responsible for translating voice, and optionally other types of data, between the two dissimilar networks connected to the gateway. Processing to improve voice quality and to adapt the media stream to the different networks is also performed by this module. These functions can be divided into several smaller operations handled by various components as shown in Figure 3.

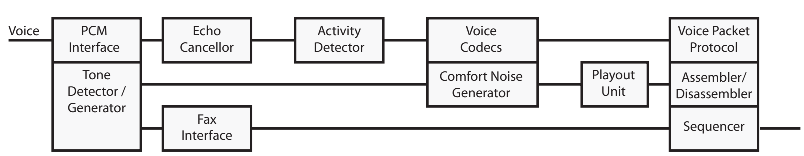

Figure 3: Media module architecture.

The PCM interface constitutes the interface toward the circuit switched network. It is responsible for converting voice signals to binary data represented in PCM format before relaying this to other voice processing components. The analogue to digital conversion is usually performed by a Digital Signal Processor embedded in specialized hardware. The tone detector and generator produces DTMF tones and call progress tones and discriminates between voice and non-voice data. This component is located close to the PCM interface since it does not require the processing provided by other components of the media module. This allows the VoIP gateway to produce DTMF tones under command of a VoIP terminal. This is useful in situations where user-input is required such as for automated telephone information systems.

The voice processing units of the media module comprise codecs for echo cancellation as well as various voice codecs for compression and decompression of voice data. To reduce bandwidth, periods of silence should be suppressed by the VoIP gateway. When no activity is detected, the network protocol module is notified thus preventing data packets from being sent across the network in periods of silence. To compensate for normal background noise, the gateway can reproduce the idle noise characteristics of the PCM interface and relay this information to the VoIP terminal on the packet switched network. Optionally, a fax interface can be implemented. When facsimile data are detected by the PCM interface, they are forwarded to the fax interface for special handling. Since most voice codecs do not allow proper compression and decompression of non-voice data, facsimile transmissions originating from the PSTN are converted to a representation internal to the packet switched network. This involves demodulating the facsimile, extracting the data and representing the scan line fax data as packets. This saves bandwidth but most importantly avoids errors resulting from voice codec processing.

The Playout Unit provides buffering of data packets from the packet switched network. Whereas the look-ahead delay introduced by the voice codecs is required for inter-frame correlation, the buffering provided by the Playout Unit assists in eliminating errors resulting from packet loss, out-of-order arrival, delay and jitter. By buffering voice packets arriving from the packet switched network, necessary corrections can be made to eliminate these faults before the data are depacketized and decompressed by the appropriate codec and transmitted to the PSTN. Since buffering introduces an additional delay to the total transmission time, the buffering must be carefully tuned to the rest of the system. Finally, the voice packet unit encapsulates the compressed voice data for end-to-end transmission over packet switched networks. This unit is implemented using protocols for realtime transmission such as RTP and RTCP and thus constitutes the interface toward the IP network.

5.2 The signaling module

The signaling module connects directly to the circuit switched network and provides translation of protocols for call signaling and control between the two networks. This module translates call control-functions and end-to-end signaling provided by the circuit switched network into state changes used by the network protocol module to set up and manage connections. The state changes are ultimately forwarded to the network protocol module using these messages to set up and administer connections according to the protocol specifications. The signaling module and its components are shown in Figure 4.

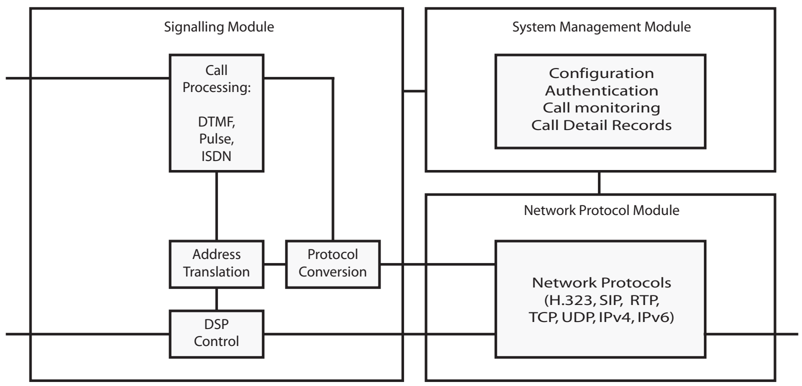

Figure 4: Signaling module architecture.

The call processing component receives signaling information from the circuit switched network and translates this information into state changes. The hardware and protocol specifications of this interface depend on the nature of the circuit switched network connected to it. Depending on network technology, signaling is relayed as DTMF or ISDN control messages. Through the modular design, support for multiple interfaces are possible to provide support for different circuit switched network technologies.

The signaling module detects incoming calls from the circuit switched network and immediately prepares for address translation. Since network addressing is protocol-dependent, this operation is divided into two stages where the first stage is completely independent of the packet switched network technology. The first stage consists of identifying a terminal located on the packet switched network using the information forwarded from the call processing unit. This includes using tables mapping virtual numbers into protocol-independent names uniquely identifying a terminal on the packet switched network. These tables are either created statically or dynamically when the terminals register with the gateway. The second stage is protocol-dependent and consists of locating the terminal using a network address conforming with the specifications of the signaling protocol used. Telephone calls directed from the packet switched network to the PSTN follow an identical two-stage translation. The caller located on the packet switched network provides the E.164 formatted telephone number which is relayed to the gateway in a message using the semantics of the specific signaling protocol. The Protocol Conversion component then extracts this number and forwards it to the Address Translation component. The call processing component finally calls the appropriate E.164 formatted number using DTMF or ISDN control messages.

5.3 Network protocol module

The network protocol module shown in Figure 4 contains all protocols required to represent voice data and signaling information from the PSTN on a format recognized by the packet switched network. For VoIP, this includes protocols for signaling (H.323, SIP), protocols for end-to-end transmission of realtime media (RTP, RTCP), transport protocols (TCP, UDP) and network protocols (IPv4, IPv6). The network protocol module operates in close conjunction with the signaling module. The signaling module connects directly to the circuit switched network and translates call control-functions and end-to-end signaling into state changes used by the network protocol module to set up and manage connections according to the appropriate signaling protocol. Media streams arrive from the media module and are encapsulated and prepared for end-to-end transmission. Finally, transport and network protocol headers are added to both voice and signaling packets before being transmitted into the packet switched network.

5.4 System management module

The system management module shown in Figure 4 provides mechanisms for authentication, registration and call monitoring as well as more advanced services such as encryption, voice messaging and voice conferencing. As a bare minimum, the system management module must be able to authenticate and register VoIP terminals and manage calls.

6 Implementation

Python has been chosen as programming language for implementing the VoIP gateway. Python is a byte-compiled, interpreted language. It is multi-threaded, object-oriented and portable with interpreters available for all common operating systems such as Linux, FreeBSD, NetBSD and Windows. Python is built around small components called modules. These modules are easy to design and use, and this encourages creation of formal and informal code libraries. Use of modules, class-support and embedded code make Python apt for the modularity requested above. However, as an interpreted language, Python is slower than compiled languages and is therefore not suited for computationally intensive applications such as voice codecs. But it constitutes an excellent framework to be used around such applications, and performance-critical components are easily replaced by pre-compiled third-party applications.

6.1 Processing of voice data

Instead of providing audio functionality itself, the media processing components of gateway are implemented using *Robust Audio Tool v.4* (RAT) [5]. This application handles unicast and multicast media streams relying on the RTP/UDP protocol suite without providing call signaling and control beyond that provided by the RTP protocol. RAT was developed by the Department of Computer Science, University College, London, and is available through an open-source license. RAT provides all the functionality outlined above including echo cancellation, activity detection, a broad range of voice codecs, silence and noise substitution as well as mechanisms for voice encapsulation and playout of RTP packets. Thus, all voice processing components of the media module can be replaced using RAT. The VoIP gateway communicates with RAT using an application launcher responsible for executing and terminating third-party applications. The application launcher also provides RAT with the parameters required to set up a connection.

The PCM interface and the fax interface are constructed using a voice modem with speakerphone capabilities[1] in combination with a soundcard. Analogue voice data from the PSTN are received by the modem and relayed to the soundcard where it is translated to binary form. The DSP of the soundcard performs sampling and digitization of sound. Data arriving from the IP network in binary form, are converted to analogue signals by the DSP on the soundcard before being relayed to the modem. The modem then routes these signals to the PSTN through its PSTN interface. Thus, in this scenario, the modem merely acts as a router relaying voice signals in either direction without interfering with the flow of data. The modem only comes into play when reading and sending signaling and control information to and from the PSTN under the control of the gateway. The DSP of a soundcard in combination with a modem provides smooth translation of voice data between digital and analogue representation without introducing noticeable delays.

6.2 Processing of non-voice data

Currently, the gateway does not support transport of facsimile data across IP networks. However, the PCM interface will recognize facsimile transmissions. This makes it possible to extend the gateway to also support Fax over IP (FoIP) services.

6.3 The signaling module

Since the operating environment of this gateway did not require a complex signaling protocol, a proprietary protocol was implemented. This protocol implements a subset of SIP. PSTN signals first arrive at the call processing component of the signaling module. If address translation is required, the address translation component is invoked; otherwise the signal is relayed directly to the protocol conversion unit. This module is responsible for translating raw PSTN signals to protocol dependent message that can be used by the gateway to initiate some type of action. The gateway also implements a local table containing information about the VoIP terminals currently registered with the gateway as well as a unique identifier associated with each terminal. The table is stored as a plain text file which allows for easy monitoring. The table assigns unique identifiers to all registered VoIP terminals and keeps track of which terminals are currently connected to the gateway.

6.4 The system management module

The system management module comprises the main part of the gateway as it is responsible for registration of VoIP terminals and call monitoring. When the system management module is started, it first prepares the hardware interfaces to both the PSTN and the packet switched network before setting up the signaling channel using TCP/IP on a well-known port. The gateway is now ready to accept calls from the two networks. During the initialization of the gateway, all PSTN devises are made available through an internal list. Whenever a VoIP terminal wishes to make a call, the gateway first verifies whether there are available resources or not. If no resources are available, the VoIP terminal is notified. If there is an available PSTN

[1]This implies that the modem is capable of being connected to an external microphone and external speakers.

interface (i. e modem), the PSTN interface is reserved. The VoIP terminal is at all times informed about the status of the gateway through messages complying with the devised signaling protocol.

The system management module also provides access to additional services. A text file contains bindings between various events and actions. Thus, for each event an appropriate action can be specified. As new service are added, they only need to be included in the list specifying the event-action bindings. This gateway implements a voice-mail service where a caller may choose to send a voice-mail when the called party cannot be reached. The caller is then asked to leave a message which is recorded to file and eventually sent to the receivers E-mail address.

6.5 Summary

The design depicted above provides the functionality and flexibility outlined by the requirements. The system management module supervises call requests from the PSTN and the packet switched network, registration and authentication of VoIP terminals, logging facilities as well as resource management. The media module, implemented using RAT, provides the necessary audio processing tools, and the signaling module translates between PSTN control information and signaling protocol messages used by the packet switched network.

7 Testing

The gateway was tested using a qualitative approach where telephone calls routed through the gateway were compared to traditional telephone calls. Functionality and sound quality were consequently compared to a standard well-known to most people.

To test the gateway, a VoIP terminal had to be devised and implemented. The VoIP terminal was also implemented using Python. Besides using the same signaling stack as the VoIP gateway, it also relies on RAT for media processing. The terminal has a graphical user interface through which its various functions can be accessed. In addition to include functions for initiating and answering phone calls, the terminal allows for network configuration and also also provides information on network status and available gateway resources. Figure 5 shows the graphical user interface of the terminal in a call session.

The test platform consisted of a VoIP terminal and a VoIP gateway connected to the same LAN as well as a conventional telephone connected to the PSTN. Various call sessions were set up. The test procedures confirmed that the gateway worked according to the requirements and the design depicted above. In addition, the gateway performed well in a user-perspective except a noticeably long call setup time primarily caused by the execution of the third-party application used for voice processing (i. e. RAT). However, this can be avoided using a tighter integration between the gateway and RAT. Voice quality, depending on the chosen codec, was acceptable and echo and latency were perceptible but not annoying. Inspired by the extensive research into the IPv6 technology at the University of Tromsø, the gateway and VoIP terminals were also successfully adapted to and tested on a network running IPv6 only. Being simple and versatile, the VoIP gateway is highly applicable when intended to be used for research related to the VoIP technology.

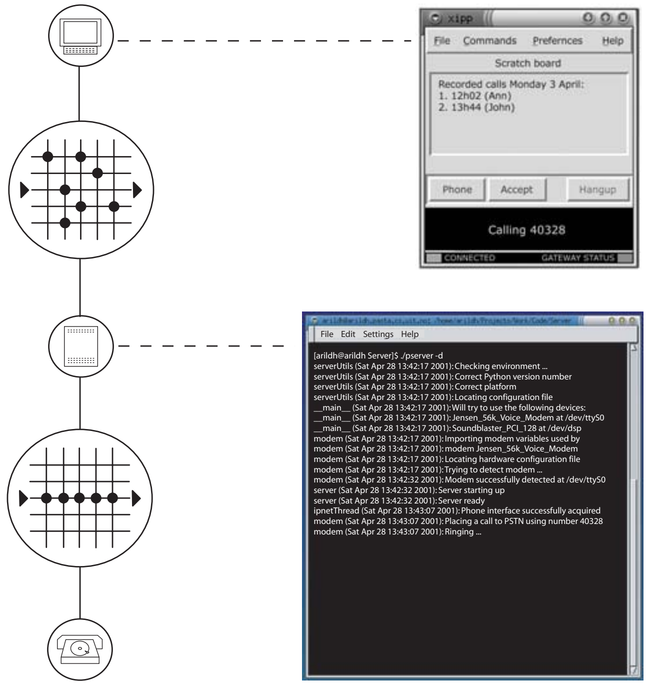

Figure 5: VoIP terminal trying to call a PSTN phone.

8 Discussion

This article has briefly investigated the challenges that must be surpassed to successfully implement a VoIP gateway bridging the PSTN with packet switched networks. Voice processing, system management, call control and signaling constitute the core services that must be provided by a VoIP gateway bridging circuit switched and packet switched networks. Through the design and implementation of a gateway, a solution to the challenges have been devised and substantiated. Though not being fit for large-scale use, the gateway provides a framework for further experimenting and research. The resulting gateway constitutes what can be defined as the least common denominator providing the flexibility needed to include current and future standards. It offers a generic reference design from which new solutions can be devised. Through its modular structure, protocols, audio software and hardware can be added to the gateway with ease.

Isaac Newton once wrote that: "If I have seen further, it is by standing on the shoulders of Giants."[2] The words of a humble student should not equal those of a genius, and I will therefore rephrase this quote with the words of a fellow colleague and friend who once claimed that: "If I haven't seen further, it is by standing in the footprints of Giants."[3]. The VoIP arena is characterized by a multitude of competing and often conflicting protocols and standards. This is not eased by the presence of the conservative and powerful telecommunications industry facing the younger and

[2]Isaac Newton in a letter to Robert Hooke, 1675.

[3]Kjetil Malde in an E-mail to the author of this article.

more open-minded Internet standardization bodies. However, the author believes that what has been referenced and outlined in this article constitute standards and protocols that will survive the preliminary phase of the battle of the species. The challenges of the VoIP technology have been identified, and solutions to some of these problems have been devised.

9 Acknowledgments

Thanks to Tage Stabell-Kulø for comments and suggestions throughout the work on this article.

References

[1] Martin P. Clark. *Networks and Telecommunications, Design and Operation.* John Wiley and Sons Ltd, Chichester, West Sussex, England, second edition, 1997. ISBN 0-471-97346-7.

[2] H.323. Recommendation H.323 (11/00) - Packet-Based Multimedia Communications Systems. *International Telecommunication Union*, 2000.

[3] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg. SIP: Session Initiation Protocol, Request for Comments (RFC) 2543. *Internet Engineering Task Force*, 1999.

[4] Q.700. Recommendation Q.700 (03/93) - Introduction to CCITT signalling system No. 7. *International Telecommunication Union*, 1993.

[5] RAT. Robust Audio Tool (RAT) Homepage http://www-mice.cs.ucl.ac.uk/-multimedia/software/rat. 2001.

[6] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications, Request for Comments (RFC) 1889. *Internet Engineering Task Force*, 1996.