Voice Over IP

Marko Leppänen
Helsinki University of Technology
Department of Computer Science
Marko.Leppanen@hut.fi

Abstract

Voice Over IP (VoIP) has been in heavy investigation during recent years. VoIP means that voice is transferred over the Internet instead of Public Switched Telephone Networks (PSTN). VoIP offers cheaper prices, but less quality than PSTNs. The aim of this paper is to give several viewpoints of VoIPs possibilities and restrictions in technical manner.

1 Introduction

Packet based networks have developed rapidly during last decades. This has caused the Internet to expand to new areas. At present time people use more and more data transfer as our digital world is developing. Data transfer traffic is multiple compared with traditional voice traffic. This progress has led to the convergence of telecommunications and datacommunications networks. This also means that telecommunications networks are replaced with datacommunications networks and this has led to the need of transmitting voice calls over IP.

Voice over IP means that voice calls are transmitted over an IP network such as the Internet instead of Public Switched Telephone Networks (PSTN). Since access to the Internet is available at more and more places in the world, it's possible to use VoIP in a higher degree.

The focus on this paper is to introduce technical aspects from the viewpoint of VoIP. VoIp's Benefits and disadvantages are considered and future visions are given. Different implementations are also studied.

However, firstly an overview to VoIP is given: history, protocols and standards, technical aspects, advantages and disadvantages. Secondly future visions and few implementations is described. Finally, some conclusions are made.

2 Overview of VoIP

VoIP converts standard telephone voice signals into compressed data packets that can be sent over IP. The audiosignal is first captured by microphone or received from line input.

The analog representation is converted to digital representation at the audio input device. The digital samples are copied into a memory buffer in blocks of frame length. A silence detector decides whether the block is treated as silence or part of talk. If the block is a part of talk it is encoded with the selected codec. Header information is added to block. The block is written into the socket. The packet is transferred over the Internet and received by another VoIP terminal. The header information is removed, the block of audio is decoded using the same codec and the samples written into a buffer. The block of samples is copied from the buffer to the audio output device. The audio output device makes the digital to analog conversion and outputs the signal.

VoIP can be used with either a telephone or PC as the user terminal. This gives different modes of operation : PC to PC, PC to telephone, telephone to PC and telephone to telephone (via the Internet).

The background of VoIP lies in the late 70's when there was discussion and experiments with packetized voice over ARPANET. VocalTec launched the VoIP market in 1995 with a Internet phone software, that made possible a voice connection between two PCs over an IP-based network [4]. The product was ideally suited for the Internet. After that, several other competing software packages were launched consecutively. In 1996 first internetworking trials between IP network and PSTN were made. In 1997 the Delta Three launched the first phone to phone service for commercial use [6].

3 VoIP standards

There are a number of standardisation bodies working on VoIP. The most important ones are ITU-T, IETF, IMTC and ATM Forum. In addition, there are a couple of smaller organizations working on VoIP such as MIT Internet Telephony Consortium, Technical Advisory Committee and ECTF [1].

There are two VoIP signaling protocols that are competing against each other : the H.323, an ITU standard and SIP, an IETF standard.

3.1 H.323

In ITU-T, VoIP work has mainly been done in Study Group 16. Their responsibility is multimedia standards for terminals, modems, protocols and signal processing [2]. ITU-T's VoIP standard is H.323, that is a multimedia conferencing protocol. It includes voice, video and data conferencing, for use over packet based networks.

The H.323 recommendation was accepted as a standard in October 1996. The H.323 standard was developed for LANs that don't provide guaranteed QoS. Version 2 was approved in January of 1998 and there were enhancements in security, performance, supplementary services and scalability. Version 3 was approved in September of 1999 and introduced small improvements concerning mainly PSTN integration and scalability. The newest version of H.323 protocol is version 4 and was approved in November of 2000. Enhancements in this version was focused in scalability, services and generic framework.

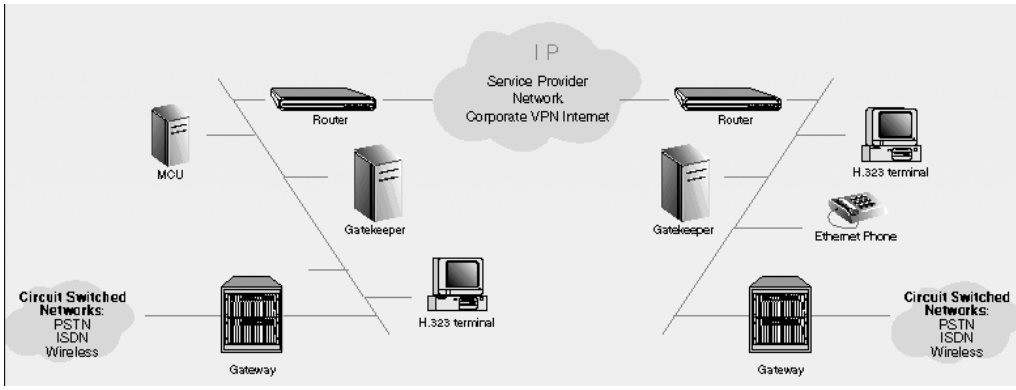

Figure 1: H.323 Architecture

H.323 utilizes the Real-Time Protocol (RTP/RTCP) from the IETF, along with internationally standardized codecs. With the ratification of version 2, H.323 is also being used for video and other communications, over the Internet.[7]

H.323 is based on reliable and unreliable communication and both communication types must be provided by the network. Reliable transport, provided by TCP, is required by control signaling and data. Unreliable transport, provided by UDP, is used for audio, video, and RAS channel. H.323 is independent from the underlying network topology. H.323 defines the system, control procedures, media descriptions and call signaling.

H.323 network architecture consist of four major components that interact on a packet network (Fig. 1). These components in H.323 architecture are :

- H.323 Terminal
- Gateway
- Gatekeeper
- Multipoint Control Units (MCUs)

H.323 Terminal is endpoint on a LAN. It supports real-time, full-duplex communications with another H.323 entity. H.323 Terminal support audio codecs and signaling (Q.931, H.245, RAS).

Gateway is an interworking device, which acts as interface between the LAN and the circuit-switched network. It translates communication procedures and formats between networks. Gateway controls call setup and clearing and perform compression and packetization of voice. Typically gateway converts IP packets to circuit switched networks like PSTN, ISDN or to wireless networks, and vice versa.

Gatekeeper is an optional device in H.323 environment. Its functions can be implemented partly in terminal or in gateway. Gatekeeper manages a zone (collection of H.323 devices). It provides the connectivity between ISDN endpoints calling into the LAN to reach an H.323 endpoint and perform network management. Gatekeeper takes care of certain functions like address translation, admission control, bandwidth control, call authorization,

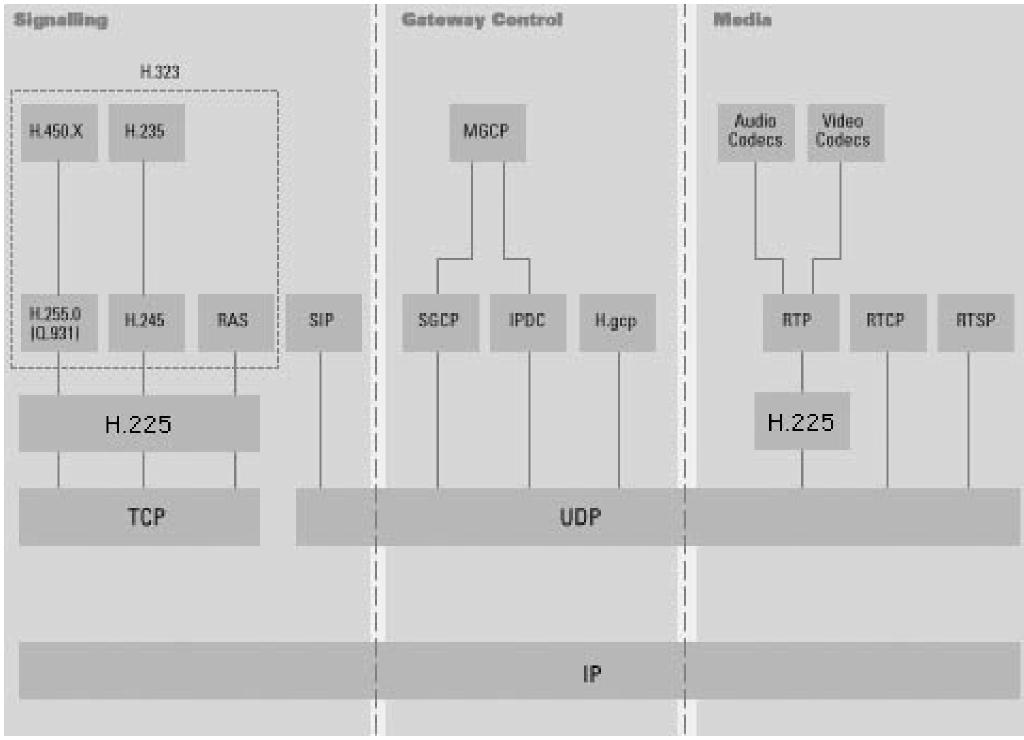

Figure 2: VoIP protocol Stack

bandwidth management, directory services and call management services. Gatekeeper is typically a software application, that is integrated in a gateway or terminal.

MCU supports sessions between 3 or more endpoints. MCU contains multi-point controller (MC) for signaling and controlling. multi-point processor (MP) in MCU processes media streams. MCU can be a separate device or integrated into a gateway, gatekeeper or terminal. Conferences can be centralized or decentralized. In the centralized conference MCU handles signaling and stream processing. In the decentralized conference MCU handles only signallig, streams go directly to endpoints.

H.323 protocol architecture consist of several protocols such as H.245, H.225, H.450, Q.931, RAS, H.235 and RTP/RTCP (Fig. 2).

signaling is transported reliably over TCP. RAS protocol handles the registeration of the terminal with the gatekeeper and is responsible for opening RAS channel in the call establisment. RAS channel acts as status channel between terminal and gatekeeper. RAS protocol also makes necessary bandwidth changes during a call.

Call setup, termination and signaling is handled by Q.931 protocol.

H.245 protocol is responsible for negotiating capabilities, that mainly include description of client's ability to transmit media streams. H.245 protocol also carries control messages concerning the operation of the H.323 entity. There are four different H.245 messagetypes: request, response, command and indication [3].

H.225.0 protocol specifies packetization and synchronization of media streams. It also specifies call control messages, signaling, registration and admission [7]. H.225.0 layer is the interface to the LAN side, so all H.323 traffic goes through H.225.0 layer (Fig. 2).

H.235 protocol defines the security requirements. H.235 provides authentication, integrity, privacy and non-repudiation. Authentication is provided by admission control that is handled by the gatekeeper that controls the zone. Data integrity and privacy is provided by encryption. Non-repudiation ensures that H.323 endpoint can't deny that it took part in a call. This is handled by gatekeeper logging services.

H.450.X series of recommendations cover supplementary services. There is a hierarchical architecture for new services. The H.450.1 includes a general framework for supplementary services. Ratified supplementary services by ITU-T include call transfer, call diversion, call hold, call park/pickup, call waiting, message waiting, name identification and call completion on busy subscriber.

Supported codecs by H.323 are ITU-T standards. There is a series of audio codecs ranging in bit rates 5.3 to 64 kb/s. The only mandatory codec is G.711, that transmits voice at 56 or 64 kbps. Other supported codecs are G.722, G.723.1, G.728, G.729, MPEG 1 audio and GSM [3]. The different codecs reflect different tradeoffs between speech quality, bit rates, CPU load and signal delay.

3.2 RTP

IETF audio-video transport group was formed to specify a protocol for real-time transmission of audio and video over UDP and IP multicast. This protocol was named as Real-time Transport Protocol (RTP). Many kind of applications can benefit from RTP: multimedia conferencing, interactive media, distributed simulation etc.

RTP consists of the actual Real-time Transport Protocol which is used to transfer data with real-time properties and RTP Control Protocol (RTCP) which is used to monitor data delivery information about the participants in an on-going conference.

RTP implementation will often be integrated into application rather than being implemented as a separate protocol layer (Fig. 3). In applications RTP is typically run on top of UDP to make use of its port numbers and checksums[8]. One RTP connection uses two ports, one port is used for data stream, and the other is used for control (RTCP) packets. The RTP framework is not so strict allowing modifications and tailoring depending on application. A complete specification for a certain application will require a payload format and profile specification. The payload format specifies how a certain payload is to be carried in RTP. A payload specification defines how a set of payload type codecs are mapped into payload formats.

3.3 SIP

The Session Initiation Protocol (SIP) is a text-based lightweight protocol, similar to HTTP and SMTP, for initiating interactive communication sessions between users. SIP is an application-layer control (signaling) protocol for creating, modifying and terminating sessions with one or more participants. These sessions include Internet multimedia confer-

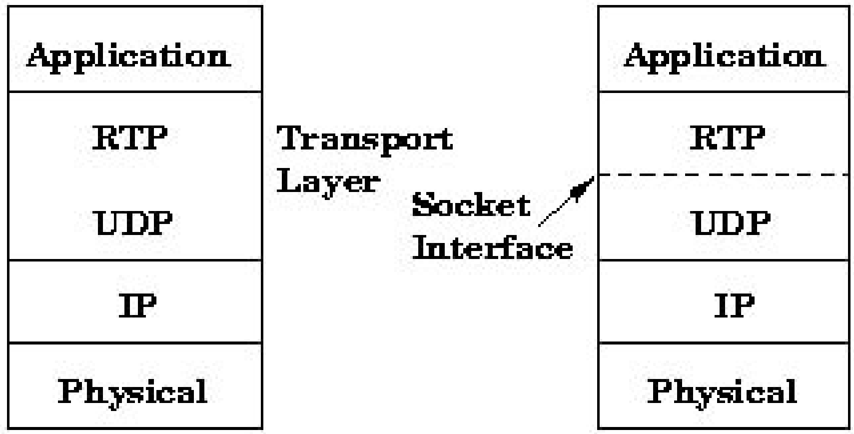

Figure 3: RTP in protocol layer

ences, Internet telephone calls and multimedia distribution [9].

SIP is developed by the Multiparty Multimedia Session Control (MMUSIC) working group of the IETF. SIP is still under development and is not so known as H.323.

SIP can be used with several transport protocols. Indeed any datagram or stream protocol that delivers a whole SIP request or response in full can be used. SIP doesn't even require reliable transport protocol and simple client can be implemented using only UDP transport.

SIP can be used to initiate sessions as well as invite members to sessions that have been advertised and established by other means. Sessions can be advertised using multicast protocols such as Session Announcement Protocol (SAP), electronic mail, news groups, web pages or directories (LDAP), among others. Members in a session can communicate via multicast or via a mesh of unicast relations, or a combination of these.

SIP uses the Session Description Protocol (SDP) in describing the capabilities and media types supported by terminals.

SIP defines four types of entities in its architecture : Terminals, proxy servers, redirect servers, location servers (Fig. 4).

Terminal provides 2-way communication with another SIP entity. Terminal supports both signaling and media, like H.323 terminal. Terminal initiates and sends SIP requests.

Proxies are like intermediary programs that acts as both a server and a client. Requests are forwarded, possibly after rewriting the request message. Proxies maintain the state of call.

Redirect server advertises the caller to contact another server directly. A redirect server can leave behind the call request after it has been processed.

Location server contains the information about callee's possible location. Location server is usually integrated in redirect or proxy server.

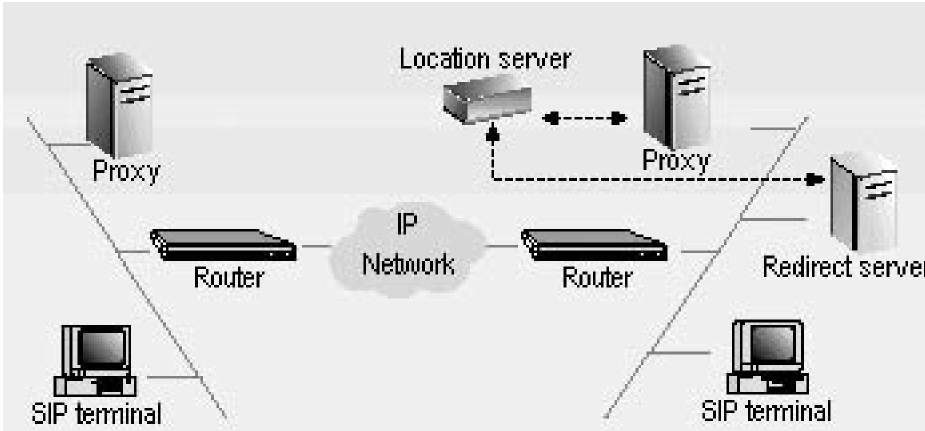

Figure 4: SIP architecture

4 Technical perspective of VoIP

VoIP sets certains technical requirements to work satisfactory. The basic requirements are devices needed to capture voice and code digital representation in someway efficient manner. Digital speech stream must be transferred over network and the flow of bits must be stable, so that packets arrive constant speed at destination. Problem in IP-networks are, that they are best-effort networks, so there is no guarantee of constant bit flow. There are delays and a lot of variation about delays. Additionally, packets sometimes vanish in their way to the destination. Moreover, processing the streams in terminal takes some time and adds delay.

Delay is essential issue concerning VoIP, because real-time voice conversations are delay sensitive. If one-way delay exceeds a quarter of a second (250 ms), the probability that parties will talk one upon another increases. Each component in the transmission path - encoding, packetization, output queuing, packet transmission, input queuing, jitter buffer, decoding - adds delay. ITU-T G.114 recommends 300mSec round-trip delay so the maximum One-Way delay is 150 mSec to achieve high-quality voice [10].

Jitter quantifies the effects of network delays on packet arrivals at the receiver. Packets transmitted at equal intervals from the left gateway arrive at the right gateway at irregular intervals (Fig. 5). Unreasonable jitter makes speech cracked and difficult to understand. Jitter is calculated based on the inter-arrival time of successive packets. For high-quality voice, the average inter-arrival time at the receiver should be nearly equal to the inter-packet gaps at the transmitter and the standard deviation should be low. Jitter buffers (packet buffers that hold incoming packets for a specified amount of time) are used to counteract the effects of network variation and create a smooth packet flow at the receiving end.

Packet loss typically occurs either in bursts or periodically due to a consistently congested network. Periodic loss of 5-10 % of all voice packets transmitted can downgrade voice quality significantly. Occasional bursts of packet loss can also make conversation difficult.

Blocks in packet based networks can cause packets to take different routes to reach the

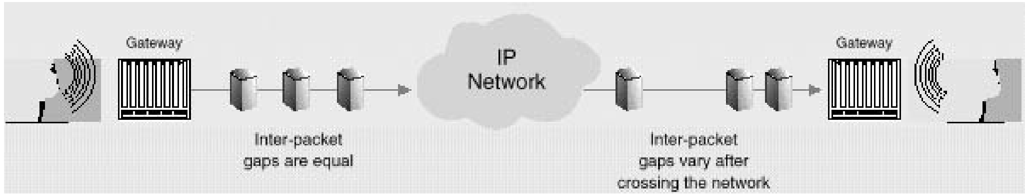

Figure 5: Jitter

same destination. Packets may arrive in different order than they were transmitted resulting in confused speech.

4.1 Quality of Service

The ultimate objective of VoIP is reliable, high-quality voice service, the kind that users expect from the PSTN. The quality issues of VoIP are very complicated and challenging. It's hard to achieve same level of Quality of Service as in PSTN. The main QoS issue is speech quality. Also issues like service availability and usability include in QoS. ITU-T has defined QoS as : "The collective effect of service performance which determine the degree of satisfaction of a user from the service."[11]

The use of priorities for VoIP packets can minimize the effects of serialization delay and the resulting jitter. There is no common standard to adopt these priorities, and most major router vendors have made their own ways of prioritizing IP traffic. The IP packet header itself has a TOS (type-of-service) field that you could be used to ensure VoIP packet priority. However, routers have never consistently used this field, and many router vendors simply ignore the TOS field or use a vendor-specific implementation that is worthless unless every router in the network looks at the TOS field in the same way. This tangle is one of the keys to priority setting. Each element of the network must understand how to read the priority and then know what to do.

There are several deployed QoS technologies nowaday. For example, the Resource Reservation Protocol (RSVP) is a true reservation setup-and-control protocol designed to achieve the characteristics similar to circuit-switched networks on an IP network. The RSVP protocol is used by a host to request specific Qos from the network for particular application data streams or flows. RSVP is also used by routers to deliver QoS requests to all nodes along the path of the streams and to establish and maintain state to provide the requested service. RSVP requests will generally result in resources being reserved in each node along the data path [12].

Unlike RSVP, Differentiated Services (DiffServ) is an evolving standard that assigns QoS classifications to traffic based on service-level agreements between users and service providers. DiffServ specifies assured service levels that you can apply to network traffic according to predefined policy criteria. As such, it is considered more scalable than RSVP but offers no standardized QoS classifications.

Beyond these, Multiprotocol Label Switching (MPLS), IPv6, and 802.1p each have their own implementation of QoS. Certainly, these various QoS standards and technologies will evolve and merge over time.

MPLS is a versatile solution to address the problems faced by present day networks speed, scalability, QoS management, and traffic engineering

IEEE 802.1p standard is for prioritizing network traffic in the Ethernet. It can be defiend as Quality of Service protocol on MAC level.

4.2 Security issues

The internet, being an open network where everyone can receive and transmit packets relatively easily, demands advanced mechanisms to secure communications. Eavesdropping of calls in IP networks is definitely easier than in PSTN. Not only the audio stream itself needs protection. Signaling requires security, primarily authentication, to prevent spoofing of calls and denial-of-service attacks. Security includes protecting call setup, call management and billing. In order to charge for VoIP calls the billing mechanism must be properly secured and implemented.

4.3 Firewall interoperability

Nowadays, almost all intranets are protected by firewalls. To make use of VoIP possible H.323 and firewalls have to be able to collaborate. Both H.323 client and firewall need changes. A H.323 proxy is required and it works like other proxies. It controls calls and decides which are allowed to let go through the firewall. A proxy is like a special gateway executing access control policies in extension to bandwidth control.

H.323 uses dynamic port allocation. When the call is set up a port number for the H.245 connection is named, and a new TCP connection will be set up to the one port. Also media channels use dynamic ports. Each channel calls for two UDP connections for the RTP streams and one bi-directional connection for the RTCP stream. A VoIP session requires two TCP connections and four UDP connections and only one of them is static. Dynamic addresses and port numbers are exchanged within the data stream.

With a packet filtering router all TCP and UDP ports above 1024 must be open for two-way traffic due to dynamic port numbers. This approach breaks the whole idea of the firewall. There can be some port address space in which the port numbers are reversed for H.323 calls, but the bigger the port address space is the bigger vulnerability. The can also some mechanisms which disassembles the H.323 packets in gateway to examine the used ports and open these ports, but is quite hard to implement and that kind of implementation increases the delay.

In SIP firewalls doesn't cause that much worries. Only one UDP or TCP connection is needed and it is quite easy to configure the firewall. Because of similarities to HTTP, SIP enables reuse of proxies and security mechanisms without modification.

5 Benefits and disadvantages of VoIP

Advantages of VoIP include lower costs of calls and possibilities of making different and versatile new services. For consumers the VoIP calls are cheaper and even free sponsored by big world wide companies.

From the operator's point of view the business advantage is that VoIP reduces network operating costs because of managing only one network. IP networks are needed anyway, but using VoIP PSTN networks aren't needed anymore. Also network building costs are far more inexpensive in the case of VoIP compared to PSTN networks. Moreover, IP networks can be utilized much more wide-ranging applications, not just voice calls and services.

The obvious benefit in VoIP is that it uses the bandwidth much more efficient than PSTN. PSTN networks reverse one pipe for every call as in IP networks many calls can use the same pipe simultaneously.

The disadvantages of VoIP focus in the quality of voice and in the service availability. The voice quality in public IP networks are not in the same level as in PSTN's but is evolving all the time. PSTN also assure better availability, you can make almost all the time a call. If IP networks are heavy jammed, then VoIP calls are not possible to offer.

6 Future aspects of VoIP

Several factors contribute future VoIP products and services. The most promising areas for VoIP are corporate intranets and commercial extranets. Because of their infrastructure, operators can control who uses network and who doesn't.

Another factor influencing VoIP evolution is VoIP gateways. Corporations are developing robust embedded systems, that will be able to handle thousands of calls simultaneously.

VoIP products and services transported via the public Internet will suffer from the varying performance levels of that transport medium. So there must be significant improvements in the basic infrastructure to guarantee constant network transport performance. There must be an increase in magnitude of backbone bandwidth and access speeds. It might also be necessary that users have to pay for the specific service levels they require.

7 VoIP implementations

Nowadays there are many different VoIP client implementations. Vocaltec was the first in market to provide voice calls between two PC's via IP network [4]. Nowadays Vocaltec offers many kinds of solutions to VoIP service providers. Another well-known VoIP client implementation is Microsoft's Netmeeting. Companies like 3Com, Cisco, Intel, Net2Phone have their own solutions and imp- lementations. A bunch of companies are making VoIP gateways supporting both H.323 and SIP, best known are 3Com, Cisco, Lucent, Motorola, Nortel networks, Siemens and Vegastream. Proxies and gatekeepers are also implemented by several companies. Few companies have implemented protocol stacks for VoIP supporting H.323 and SIP. Moreover there are billing and managements software products for VoIP provided by several companies.

8 Summary

This document presented different facts of VoIP. First there were introduction to VoIP and then overview of VoIP. After that there were discussion of major standards related to VoIP, H.323, SIP and RTP. There were also considerations of technical issues in VoIP, like QoS, security and firewalls. Benefits and disadvantages were considered and also future visions. Major VoIP implementations and companies involved in were mentioned.

VoIP products are developing and nowadays there are lots of companies developing VoIP products.

There are two competiting VoIP standards: ITU-T's H.323 and IETF's SIP. H.323 is older standard and has spread wider. SIP is newer than H.323, but SIP is reaching H.323 lead, because of simplicity. H.323 has been examined to be inflexible and complicated compared to SIP. Simplicity and flexibility of SIP might help SIP to get as high appreciation as H.323.

There are numbers of essential issues influencing the succession of VoIP. Delays, jitter, packet loss and blocks in networks produce drawbacks to VoIP.

QoS is one of the major issues concerning VoIP. There must be certain level of QoS in VoIP that it would totally displace PSTNs. QoS issues are managed with several protocols and standards like RSVP, DiffServ, MPLS, IPv6 and IEEE 802.1p.

Security in the VoIP context is still not seriously considered. But there there are few standards providing VoIP security.

VoIP has a lot of advantages: cheaper calls, cheaper network building and management, efficient bandwidth usage. Disadvantages are in the field of QoS and availability.

VoIP has well-lighted future, because several companies are developing VoIP products and IP networks are developing fast.

References

[1] Jain R. *Voice over IP references* http://www.cis.ohio-state.edu/ jain/refs/ref_voip.htm

[2] ITU-T Study Group 16. *Area of Responsibility* http://www.itu.int/ITU-T/com16/area_resp.html

[3] ITU-T Study Group 16. *Contribution 54.* ITU-T recommendation H.323, October 1997

[4] http://www.vocaltec.com/about/aboutus.htm *VocalTec Communications Ltd.*.

[5] http://www.protocols.com/voip/index.html *Protocol.com*.

[6] Koistinen T., Haeggström J. IP Telephony. Nokia Telecommunications, October, 1998. http://keskus.hut.fi/opetus/s38130/s98/ip_tel/ip_tel.html

[7] IMTC. *H.323 Overview*. http://www.imtc.org/h323.htm

[8] Schulzrinne H., Casner S., Frederick R., Jacobson V. RTP: A Transport Protocol for Real-Time Applications, RFC 1889, January 1996. http://www.alternic.org/rfcs/rfc1800/rfc1889.html

[9] Handley M., Schulzrinne H., Schooler E., Rosenberg J. SIP: Session Initiation Protocol, RFC 2543, March 1999. http://www.alternic.org/rfcs/rfc2500/rfc2543.html

[10] ITU-T *recommendation G.114*

[11] Spergel, L. & Kimchi G. ETSI TIPHON, Project Overview, 1998.

[12] Braden R., Zhang L., Berson S., Herzog S., Jamin S. Resource ReSerVation Protocol (RSVP), RFC 2205, Semtember 1997. http://www.alternic.org/rfcs/rfc2200/rfc2205.html

Abbreviations

ECTF	Enterprise Computer Telephony Forum
ETSI	European Telecommunications Standards Institute
IETF	Internet Engineering Task Force
IMTC	International Multimedia Teleconferencing Consortium
ISDN	Integrated Services Digital Network
ITU-T	International Telecommunications Union - Telecommunications
LAN	Local Area Network
MCU	Multipoint Control Unit
MPLS	Multiprotocol Label Switching
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RAS	The Registration, Admission, and Status
RSVP	Resource Reservation Protocol
RTP	Real-Time Protocol
RTCP	Real-Time Control Protocol
SAP	Session Announcement Protocol
SDP	Session Description Protocol
SIP	Session Initiation Protocol
VoIP	Voice Over IP