Hochschule Mittweida
University of Applied Sciences

MITTWEIDA
ROSSWEIN

Faculty of Media

Frank Friemel

Design and implementation of the H.323 video conferencing standard and development of a secured video conferencing kiosk with the Google Voice and Video plug-in

- Submitted as Diploma Thesis –

Hochschule Mittweida – University of Applied Sciences

First examiner:	Prof. Dr. Michael Hösel
Second examiner:	Dipl.-Ing. Peter Lubosch

Bibliographic description

Friemel, Frank:

Design and implementation of the H.323 video conferencing standard and development of a secured video conferencing kiosk with the Google Voice and Video plug-in. – 2009. – p. 104. Mittweida, University of Applied Sciences Mittweida, faculty of media, diploma thesis, 2009

Abstract

The H.323 umbrella standard describes audiovisual communication over packet-switched networks. This thesis illustrates the standard in detail with regards to architecture and implementation.

The second part of this dissertation is dedicated to examining the Gmail Voice and Video plug-in, an Internet-based audiovisual communication platform. In the course of this thesis a secured kiosk environment for the Gmail Voice and Video plug-in is being developed.

Keywords: video conferencing, packet-switched networks, circuit-switched networks, H.323, H.320, ITU-T, Google, Gmail Voice and Video plug-in, IP, ISDN

Table of Contents

Table of Figures

List of Tables

List of Abbreviations

AAC	Advanced Audio Coding
ADPCM	Adaptive Differential Pulse Code Modulation
BRI	Basic Rate Interchange
CSS	Cascading Style Sheets
CCD	Charged Coupled Devices
DCT	Discrete Cosines Transformation
DPCM	Differential Pulse Code Modulation
DSP	Digital Signal Processor
EMEA	Europe, Middle East and Africa
FASIC	Function and Application Specific Integrated Circuits
FECC	Far End Camera Control
GIPS	Global IP Solution
HTML	Hypertext Mark-up Language
IEC	International Electrotechnical Commission
IP	Internet Protocol
ISO	International Standardisation Organisation
ISDN	Integrated Services Digital Network
MPEG	Motion Picture Expert Group
OS	Operating System
PCM	Pulse Code Modulation
PRI	Primary Rate Interchange
PSTN	Public Switched Telephone Network
RAS	Registration, Admission and Status
RRQ	Registration Rejected
RTP	Real-time Transport Protocol
RTCP	Real-time Transport Control Protocol
VoIP	Voice over IP
VC	Video Conferencing

Preamble

The term video conferencing derives of the two latin words videre „I see“and confere „to bring together“(Wilcox and Gibson 2005 p. 3). In our modern society most of the people are associating with the word a telephone with a screen and a camera.

However video conferencing has the ability to deliver more than just being a telephone with video. It is a multi-media collaboration experience, which makes long-distance communication a more interactive experience.

Video communication is real-time exchange of digitized video images and sounds between conference participants at two or more separate sites. Video conferencing needs to be interactive and doesn’t limit itself to the transmission of locally recorded images and/or audio. Anything as diverse as video clips, still pictures or recorded audio can be transferred. The difference from television is the interaction between the sites involved (Wilcox and Gibson 2005 p. 4).

Words like distributed workforce, global markets and supply chains have changed the nature of communication in many industries. Companies have international offices in many different countries. Instant and effective communication is necessary to react quickly to changing markets and demonstrate leadership. Video communication is the communication method of the future, because it combines versatility with increased quality of communication.

Manufacturing companies using video conferencing for quality assurance and real-time process and equipment monitoring. Healthcare providers deploy it to enhance patient care and legal teams use it to streamline casework. In fast moving society the new technology is embraced not only by businesses, but also more and more by private persons. Often family members are dispersed around the globe and use technologies like Skype[1] and Google Voice and Video to reconnect with each other.

Video conferencing technology offers the ability to make long-distance communication more natural.

[1] Skype® is a free communications and collaboration tool for voice and video conferencing.

1 Introduction

The following diploma thesis is designed to give an overview over the H.323 standard of video conferencing for packet-switched networks. The reader will get an in-depth knowledge about video conferencing over LAN based network architecture. The H.323 design is looked at, as well as the practical implementation of video conferencing solutions.

The goals of my thesis are:

- Overview over the H.323 standard and presentation of its associated components and protocols in detail.
- Presentation of audio- and video compression standards that are comprised in the H.323 standard.
- Comparison of a H.323 compliant video conferencing system with the Google Gmail Voice and Video plug-in
- Development of a video conferencing kiosk, to enable Google employees from the London office to make video calls from the phone booths in the office.

My thesis is structured to give at first a theoretical overview about the H.323 standard with regards to its architecture and protocols. The associated network elements are discussed in detail in terms of usability and core functions.

The theoretical overview is followed by a detailed description of a H.323 implementation at Google and in particular at the Google office in London.

The second part of this dissertation is dedicated to the development of a unique video conferencing solution adapted for the phone booths in the Google London office. This was project was undertaken to promote the newly developed Gmail Voice and Video plug-in among the Google staff and encourage them to use this video conferencing feature in corporate and private meetings.

I was given the project to design a video conferencing solution for the phone booth's installed on in the London office. Google was at the time developing the Google Voice and Video plug-in, which was required to be at the heart of the implementation.

Google is a good example for an international operating company that uses video conferencing to engage in information exchanges that span the globe in real-time. With offices all around the world and teams dispersed over different continents video conferencing offers the solution for instant and effective multi-media collaboration. By engaging audio and visual senses, communication resembles face-to-face communication.

Video conferencing in general is in the transition period at the moment where its extensive application in corporate environments has lead to adoption in the private space.

In 2004 a study by Wainhouse Research found out that 31% of the workers had already participated in interactive multimedia communication (Wilcox and Gibson 2005 p. 3). Nearly three quarters of the respondents of the survey declared that video conferencing use had increased in the last year and is expected to increase even further in the years to come.

In the financial year 2008/2009 BT Conferencing sold world-wide video equipment worth of £6.072.492. In comparison to the previous financial year this meant a plus of 228%[2].

The grim economic outlook causes companies to cut travel and other expenses for its employees and invest into video conferencing solutions that promise quick return-of investment (ROI). At the end of the first decade of the 21st century video conferencing is a growing market, despite economic crises and worldwide recession.

Video conferencing has become a mass-market, which will be growing for years to come. It will enrich and alter our fashion to communicate.

I would like to thank Professor Hösel for his help and assistance during the writing process. I would also like to express my gratitude towards Alberto Martin, James Bond and Rick Earnshaw from Google for their continuous support during the project phase. Google offered me at all times the best support for the successful completion of my thesis.

[2] Source: UKROW_Sales_Report_P12/ BT Conferencing

2 Videoconferencing over IP (LAN) and ISDN

A video conferencing signal consists of several streams: audio, video and data streams. Almost two decades ago a video conference was conducted over two separate networks: One ISDN network carried just the audio information and the second ISDN network carried the video signal (Firestone, Ramalingam et al. 2007 p. 17). The transmission of each signal over separate lines was long a common procedure in the world of long-distance collaboration (Wilcox and Gibson 2005 pp. 6-8) and constrained the success of the technology. Three obstacles had to be overcome to make video conferencing easy, reliable and cheap enough for mass deployment.

- The invention of digital transmission to allow complex signal treatment.
- The development of powerful compression algorithm to transport real-time media over channels with limited bandwidth. An uncompressed digitized television signal requires at least 45 Mbps bandwidth (Wilcox and Gibson 2005 p. 7).
- The design of networks that could transfer different signal types (voice, video and data) concurrently.

It wasn't until the early 1990's that powerful compression standards allowed video conferencing over only one ISDN line and the technology began to grow. Two network types are in use today to transport video conferencing signals: packet-switched networks and circuit-switched networks.

2.1 Packet switched networks

Today large organisations manage multiple independent networks, each optimized to carry unique types of data: voice, data and video. In the past the characteristics of each traffic type made it impossible to transfer different types of data over the same transmission line. You could not use the phone lines to transmit the phone signal and television signal as well. Many enterprises examined the existing and often separate data, voice, and video network infrastructures to determine the most efficient way to bring these networks together. Internet Protocol (IP) based LAN infrastructures seemed to be the solution. In these converged networks different digital data is transmitted over one channel.

The Internet Protocol (IP) is used for communicating data across a packet switched network using the Internet Protocol Suite known as TCP/IP[3]. Each computer (known as a host) on this network has at least one IP address that uniquely identifies it from all other computers. IP networks are built on the principle of packet switching networks.

In packet switched networks streams of data are broken into small subunits. These subunits are transmitted separately over the network. In order to be able to reassemble the stream at its destination additional information about the content of the subunit are added in front (header) and after (footer) the subunit. The whole structure is called packet. Each packet contains, along with the data load, information about the IP address of the source and destination, sequence numbers and some other control information. A packet can also be called a segment or datagram. Packet-switched networks allow the transmission of multiple streams between two communication bodies at the same time. There is no predetermined path. Every packet entails the information of the destination and can be routed different network paths (Tanenbaum 2003 p. 20). IP is a connection-less protocol, which means that there is no continuing connection between the end points that are communicating. Each packet that travels through the Internet is treated as an independent unit of data without any relation to any other packet.

The advantages of having a LAN that runs on the IP technology:

1. TCP/IP is the common worldwide standard for networking and supported by all vendors. All modern operating systems support TCP/IP, and most large private networks rely on TCP/IP for most of their traffic.
2. IP networks are omnipresent. In order to be able to gather, process, and distribute information, companies have built up large computer networks. Those networks can be connected to from anywhere in the world over the Internet, the global public network. Everything is connected and is uniquely addressed by its IP address.
3. Different devices (video conferencing units, personal computers, printers, servers) can all be connected to the same network without the need to change the network infrastructure.

2.1.1 Challenges of real-time communication over packet switched networks

The challenge of providing adequate real-time transmission in packet switched networks is the difficulty to manage latency, jitter and congestion.

[3] TCP/IP – Transport Control Protocol/ Internet Protocol is the suite of communications protocols used to connect hosts on the Internet.

Latency is the delay between the processing and sending of information from an endpoint and the reception of that information at the far endpoint. Typical A/D and D/A converters induce about 15 ms delay on a video signal[4]. A consistent latency will result in a delay in reception and response. In applications with a closed feedback loop (i.e. video conferencing), latency plays a very important role. Both conference participants are in constant connection during the conference. A long delay impedes the natural communication, as participants cut each other off.

Jitter describes the variance in latency during a videoconference (Tanenbaum 2003 pp. 395 - 396). In IP-based communication, packets are not guaranteed a particular transmittal order or throughput. Therefore, while one packet may be transmitted with minimal delay (latency), the following packet may be transmitted with large latency. When audio and video signals are reconstructed from the received packets to a continuous stream, the stream becomes interrupted with each variance in latency. This results in jumpy or dropped video sequences, and scratchy or unintelligible audio.

When too many packets want to be transmitted at the time over a pipe with limited bandwidth, congestion occurs (Tanenbaum 2003 pp. 384-385). The result of congestion is dropped packets. In a video conference dropped packets result in image errors and unintelligible audio. In the worst case the conference is interrupted.

In the next chapter another network type is presented that is used for the transmission of video conferencing signals: circuit-switched networks.

[4] http://www.salyens.com/latency/index.html, accessed the 06.06.2009

2.2 Circuit switched networks

Circuit switching and packet switching both are used in high-capacity networks (Tanenbaum 2003 pp. 147-151). In circuit-switched networks, electronic signals pass through several switches before a connection is established. These electronic signals travel along a fixed path. The decision on which route to follow is based on a resource-optimizing algorithm, and then all the transmission goes along that path. For the whole length of the communication session the route between the two communicating bodies is exclusive. Nobody else can use this route until the communication of the participants is terminated and the path is released. The information is sent as a single bit stream managed by a clock signal.

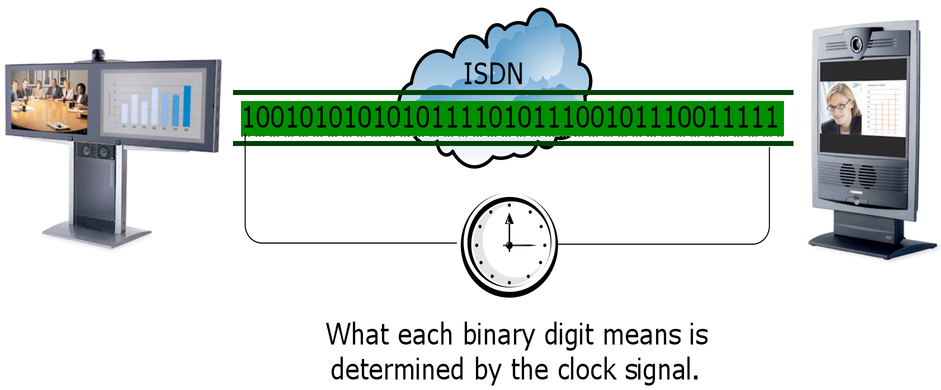


Figure 1: single bit stream with ISDN

The advantage of this type of connection is that the bandwidth is constant and the bandwidth is only used for your call and not shared. The main disadvantage is that the call fails when the clock signal is disrupted. Traditionally, this communication channel has been provided by a telecommunications infrastructure called ISDN.

ISDN stands for Integrated Services Digital Network. It is basically a digital circuit-switched telephone network system, designed to allow digital transmission of voice, video and data over ordinary telephone copper wires, resulting in better voice quality than possible with an analogue connection. Telephone numbers are the unique addressing scheme. Each device on a circuit-switched network has a telephone number which uniquely distinguish itself from all the other devices.

ISDN runs across the same network as analogue phones. Once the connection between to entities is established the single bit stream runs from the videoconferencing unit over the local loop to the local exchange. The local loop is the line between the the video conferencing unit and the local telephone exchange (Wilcox and Gibson 2005 p. 189).

The signal is then sent across the world from telephone switch to telephone switch. It may go through international switches, across transatlantic cables or even satellite links before it reaches the local telephone exchange of the remote location.

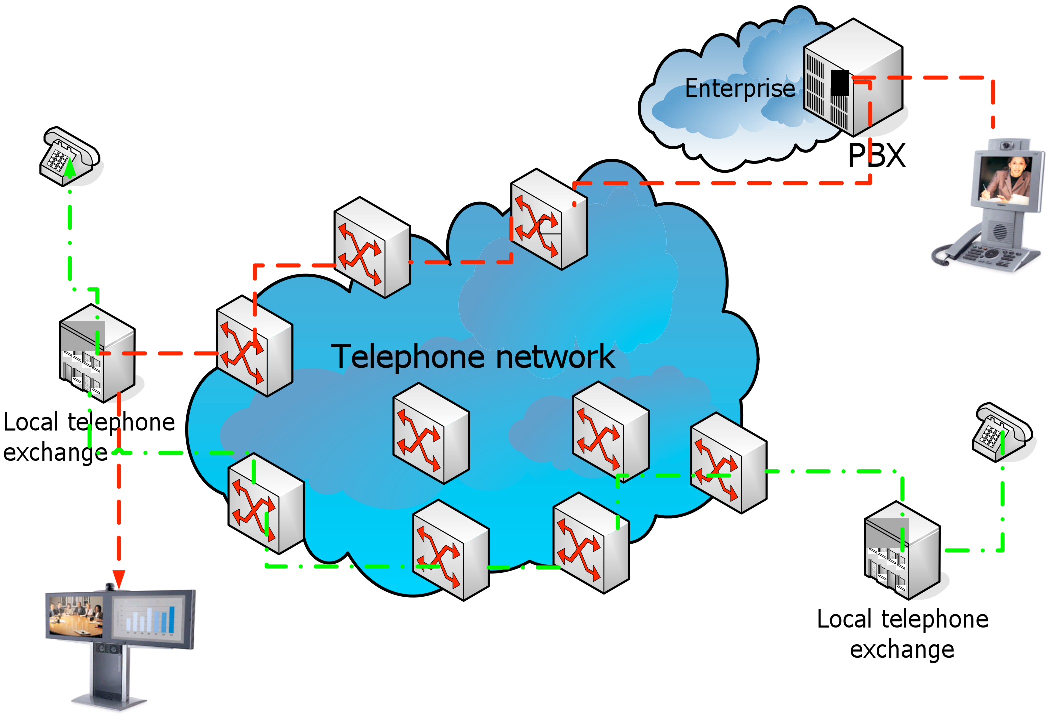


Figure 2: Telephone network

Telephone switches belong to telephone telecommunication companies, like "Deutsche Telekom".

ISDN comes in two sizes. BRI (Basic Rate Interface) and PRI (Primary Rate Interface) (Wilcox and Gibson 2005 pp. 201 - 204). A BRI ISDN installation contains 2 channels of ISDN, each capable of carrying 64Kbps of data which is 64 Kilobits per second. This means each second a channel is sending and receiving 65536 bits of information. Digital information is represented by a 1 or a 0. Most endpoints use BRI's while network equipment tends to use PRI's.

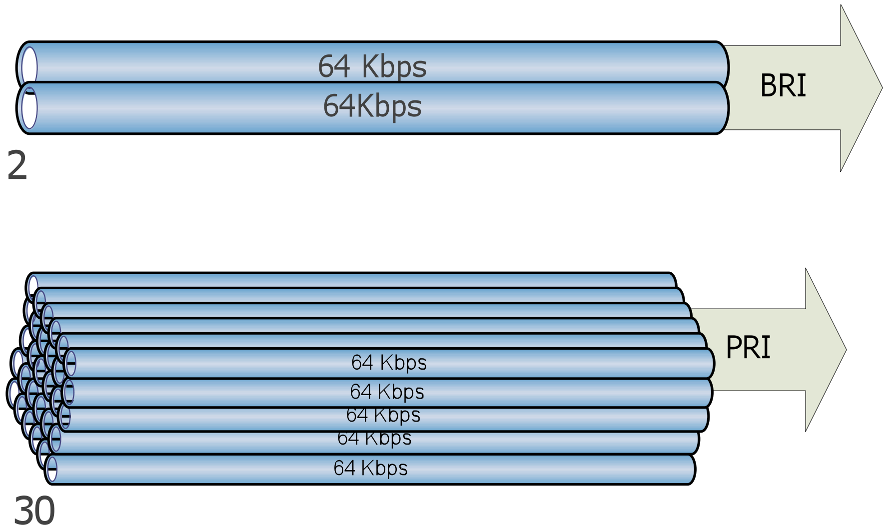


Figure 3: ISDN BRI and E1 PRI

PRI's can have unlimited amount of numbers associated with it. PRI's can come in two types depending on where you are located. An E1 PRI has 30 channels equalling 1920 Kbps of data and is found in most places around the world. A T1 PRI has 24 channels equalling 1472Kbps and is found in North America and Japan. The more channels are used for a videoconferencing call, the more expensive the call is.

2.3 H.323

The ITU published a standard of sub protocols under the name H.323 in the year 1996 to enable video conferencing over packet switched networks (LAN). The standard derived from the already existing H.320 standard which was developed for voice and video communication over circuit switched telephone networks (Tanenbaum 2003 pp. 686-688).

2.3.1 H.323 components

The H.323 specification defines a number of elements that are required for multimedia transmission (ITU-T 2007).

It is important to distinguish between the following H.323 components (ITU-T 2007 pp. 26-58):

- H.323 Codec/ Terminal

  Terminals are the client endpoints on the LAN that provide real-time bidirectional multimedia communications. An H.323 terminal can either be a personal computer (PC) or a stand-alone device, running an H.323 stack and the multimedia applications. It supports audio communications and can optionally support video or data communications. H.323 specifies the modes of operation required for different audio, video and/or data terminals to work together.

- H.323 Multipoint Control Unit (MCU)

  An MCU maintains the exchange of media in conferences with more than two participants. An MCU handles call control and the media exchange (voice, video and data) between the different participants during a conference.

- H.323 Gatekeeper

  A gatekeeper is responsible for address translation, bandwidth control and security of a video conferencing network. It provides addressing schemes: (E.164 numbers) and alphanumeric names (H323 ID's).

- H.323 Gateway

  A gateway is any device that allows calls to be established between networks, whether of the same or different types. It may also provide protocol conversion between H.323 endpoints and endpoints that do not support H.323. For example: An ISDN gateway does the protocol translation between a H.323 protocol stack to a H.320 protocol stack and vice versa.

The ITU introduced H.323 to standardize protocol stacks for the usage in video conferencing solutions. H.323 should guarantee interoperability between products from different manufacturers (ITU-T 2007 p. 4).

2.3.2 Video conferencing, networking and telephony protocols

H.323 comprises a number of protocols, including audio (voice) and video codec's, and standards for passing data and control structures (Tanenbaum 2003 pp. 686-688).

- G.711, G.722, G.728, G.729, AAC-LC, and AAC-LD are audio codec's
- H.261, H.263(+) and H.264 are video codec's. H.264 is the most recent video compression standard. It offers approximately twice the quality of H.263 at the same bandwidth, or the same quality at half the bandwidth. H.264 is based on the MPEG-4 AVC standard of the ISO/IEC[5] (Firestone, Ramalingam et al. 2007 p. 342). However H.264 encoders may have a CPU load four times as much as inferior codec's.
- H.239 is the international standard for the use, control and labelling of two simultaneous channels of video in a single video conference. It defines how PC desktop graphics are converted into a separate media stream and transmitted in parallel with the video stream.

- H.225 and H.245 are signalling protocols
- The Real-time Transport Protocol (RTP) defines a standardized packet format for delivering audio and video over a network. RTP is usually used in conjunction with the Real-time Control Protocol (RTCP). RTCP monitors the transmission and quality of service (QoS) information. When both protocols are used in conjunction, RTP is usually originated and received on even port numbers, whereas RTCP uses the next higher odd port number (Firestone, Ramalingam et al. 2007 pp. 105-107).

[5] International Organization for Standardization/ International Electrotechnial Commission

3 Endpoints/ Terminals

A video conferencing endpoint is a device, which serves as input for the audio, video and data signals. It is basically the most important component of a video conferencing network. The endpoint receives the analogue signals from the input devices camera and microphone.

The word codec is synonymously applied and points out the function as encoder and decoder of audio, video and data streams.

The codec's function is the recording of image and sound on the near end, and the reproduction of the far-end[6] signal. Furthermore as the name already indicates, a codec is also responsible for encoding the data stream and subsequently sending it over the network. Concurrently it receives the media packets from the other communicating body and decodes the compressed data stream to reproduce audio and video information accordingly (ITU-T 2007 p. 15).

An endpoint is usually equipped with a camera, a microphone and a screen, as well as DSP for processing of the media streams.

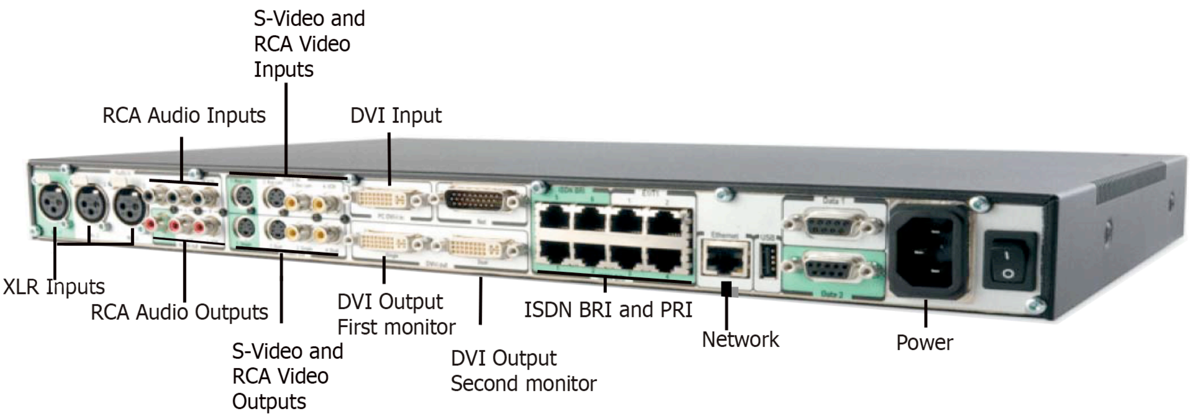


Figure 4: Inputs and outputs of an endpoint

[6] Far-end describes the person sitting in front of the codec on the distant side as in opposite of the codec locally situated.

3.1 Audio

3.1.1 Microphones

The microphones generally deployed in endpoints are boundary microphones that make use of the effect that there will be a doubling of acoustic pressure at the diaphragm of the microphone, when sound encounters a hard unyielding surface (Henle 2001 p. 169). In a conference room where the participants sit sometime several meters apart from each other, this is the superior technology as the microphone delivers a higher sensitivity. Depending on the size of the room several microphones can be used. Due to interference problems, usually only one microphone is deployed.

It is possible to use other kinds of microphones for the codec’s like lapel microphones and even hand-held microphones.

Figure 5: Boundary microphone as audio input

Once the audio signal is recorded and reaches the codec it is digitized and compressed.

3.1.2 Audio compression

Digitizing audio is a multi-step process (Wilcox and Gibson 2005 p. 93). The first step is sampling. Sampling is the process of measuring analogue slices over time, but at regular intervals. Depending on the frequency range of the sampled material: The sampling frequency must be twice as high as the highest analogue frequency of the sampled signal[7] .

The human voice typically produces frequencies between 50 Hz and 4 kHz (Wilcox and Gibson 2005 p. 93). In this frequency range most of the information of the human speech is concentrated. After the Nyquist-Shannon sampling theorem (Karrenberg 2005

p. 303) the signal needs to be sampled at minimum 8000 samples per seconds to reproduce the frequencies correctly.

PCM is the most common method to convert an analogue audio signal to a digital version in video conferencing. The incoming audio from the microphone is sampled 8000 times per second. When the sample is taken, it is compared to a set of numeric values. Those numeric values derive from the bit rate the audio is encoded with. If the sample is being encoded with a bit rate of eight bits, then 256 integers can be used to encode it. The integer most closely resembling the value of the sample is used to represent it. This step is called quantization.

After the quantization, the next step following is encoding. In this step the selected integer is expressed as an eight-bit word. Eight bit is one BYTE. Every 125 microseconds a stream of one BYTE is transmitted across the network until it reaches its destination and a decoder reconverts it into its analogue form.

PCM requires a transmission speed of 64 Kilo-bits-per-second (Kbps), because the signal is sampled 8000 times per second and each sample is encoded using one BYTE equalling eight bits (8000 * 8= 64000). 64 Kbps are the fundamental building block of digital networks. One ISDN bearer channel entails exactly that bandwidth.

Table 1: Audio codec's in video conferencing

Standard	**Basis algorithm**	**Audio bandwidth in kHz**	**Required transmission line speed**
G.711	PCM	3,1	64
G.721	Adaptive DPCM	3,1	32
G.722	Sub Band Adaptive PCM	Until 7	48, 56 or 64
G.727	Adaptive DPCM	3	32 or 40
AAC-LD	MPEG4- AAC	11	24, 48 or 96

3.1.2.1 PCM

In Table 1 are the audio codec's displayed in use for video conferencing. G.711 is most basic standard using PCM and a bandwidth of 3.1 KHz, which equals the frequency spectrum the human voice is operating in. G.721 and G.722 use an advanced form of PCM the adaptive DPCM. Instead of quantizing the speech signal directly, the differential PCM quantizes the difference between the current signal sample and its predecessor.

[7] Nyquist-Shannon sampling theorem

Therefore it requires a lower bit-rate to encode the samples as the difference between two samples vary little and so only few bits are needed for quantization. The adaptive PCM uses an algorithm to predict the likely value of the next information signal. Once the signal is predicted only the difference between the predicted and the real signal sample is transmitted. Due to that process the bit-rate of the audio signal is further reduced, which then has impact on the required bandwidth.

3.1.2.2 AAC

The advanced audio coding (AAC) is a lossy encoding scheme and is meant to be the successor of the mp3 format[8]. AAC has been standardized by the ISO[9] and IEC[10] as part of the MPEG-2[11] and MPEG-4 specification. Additionally to reducing the necessary bandwidth by only transmitting the difference between two samples and the use of predictive coding (compare to adaptive DPCM) the AAC codec removes signal components that cannot be perceived by the human ear. After encoding the audio signal all redundancies are being removed.

- It supports sampling frequencies from 8 kHz to 96 kHz
- Up to 48 channels can be encoded into one audio stream.
- Short time processing delay of only 20 ms ((VCC) 2008)

After having discussed the audio signal processing in video conferencing, the next chapter is dedicated to look into image recording and processing.

3.2 Video

3.2.1 Charge-coupled devices (CCD)

An important component of a video conferencing system is the camera. Cameras are responsible for recording the image from the near end and sending the video signal to the codec for processing.

Most of the cameras are based on a technology called charge-couples-devices (CCD). Instead of focusing the scene sample on an imaging tube, the lens directs the photons to strike the surface of a CCD chip (Wilcox and Gibson 2005 p. 404).

---

[8] http://www.apple.com/quicktime/technologies/aac/, accessed the 25.06.2009

[9] International Standardization Organization

[10] International Electrotechnical Commission

[11] MPEG-2 and MPEG-4 are video compression standards published by the Motion Picture Expert Group

When the light beam hits a capacitor, it accumulates an electrical charge proportional to the light intensity at that location. A CCD usual consists of a two-dimensional matrix, where each individual element defines the camera's spatial resolution capacity. Each matrix element samples the light that strikes it, converts that light to a charge and then transports the charge to the camera circuitry (Schmidt 2003 pp. 286 - 287).

During the making of this thesis the best cameras available were capable of recording an image in HD at either 1080p50 or 1080p60, meaning an image resolution of 1920x1080 pixels and a maximum frame rate of 60 frames per second. Video signals are digitized within the camera and then transferred over a HDMI connection to the codec.

Figure 6: Tandberg PrecisionHD camera[12]

3.2.2 Camera control

In large meeting rooms, where participants sit around the table and different people have the word, an important feature of a video conferencing system is camera control. Camera control means the ability to change the camera position and to refocus.

The remote control that comes with each codec offers basic camera movements to pan and tilt as well as zoom in and out. Modern HD cameras have a HDMI connection to transport the video signal and a separate RS-232[13] (e.g. Tooley, Winder 2002 pp. 72-73)

[12] www.tandberg.com, accessed the 25.07.2009

[13] The RS-232 interface is widely used for serial communication between peripheral devices. A RS-232 serial port is usually implemented using a standard 9-pin interface connector.

nine PIN connection for transmission of the control commands from the codec to the camera.

Figure 7: RS232 Serial 9 pin

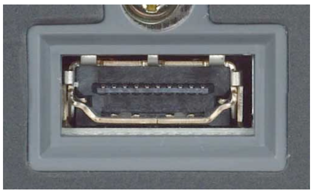


Figure 8: HDMI interface

Far-End Camera Control (FECC) is a protocol within the H.323 standard to allow a person on one end of the video conference to control the camera on the other end. This only works for a direct connection between two endpoints. FECC used to be a feature only limited to high-end room systems, but is now available to most systems with a movable camera (Firestone, Ramalingam et al. 2007 pp. 17-18).

3.2.3 Video compression

3.2.3.1 The human eye

Human eyes and ears are less sensitive in certain frequency ranges. For instance the eye adjusts very well to variations in relative and absolute brightness, but not to the lack of focus (Wilcox and Gibson 2005 p. 104). Viewers need the part of the picture that convey critical information clearly visible – for example numbers, letters and the eyes of other people. A blurry background does not perturb our viewing experience, but illegible text does. Modern compression algorithms use this as the basis for lossy quality reduction.

The human visual system is three-dimensional. We must use three different components to describe a video signal: hue, saturation and luminance (Wilcox and Gibson 2005 pp. 95-97). Hue refers to the complete spectrum of colours. There exist three primary hues red, green and blue. Saturation describes the pureness of the colour and brightness is a measurement for whiteness in the colour.

The three different components[14] of colours need to be taken into account when it comes to encoding. Colour encoding systems used for video must transform colour space. They must change it into a different format so it can be encoded and compressed more easily.

The human eye accommodates brightness and colour differently (Wilcox and Gibson 2005 p. 97). Separation of the luminance (brightness) signal from the chrominance signal (hue and saturation) is the first step in video encoding. Luminance signals are usually recorded at a higher frequency because the eye is more sensitive to luminance than to chrominance. The codec's used in video conferencing represent each pixel by its luminance value (Y) along with the chrominance values (Cb and Cr) (Firestone, Ramalingam et al. 2007 p. 52).

3.2.3.2 Compression steps

An uncompressed PAL television signal with a resolution of 720 x 576 and 25 fps requires a transmission bandwidth of 20 Mbps (CCIR [ITU-R] 601). Media streams of that size would lead in a LAN network immediately to congestion. Bandwidth is always limited and therefore needs to be thoughtfully administrated. Compression is a means to reduce the amount of bandwidth needed, but on the other hand requires powerful and especially fast encoder and decoder at both ends of the communication channel. Today it is possible to transfer audio and video information over two ISDN B channels with a bit rate of 64 Kbps due to elaborate compression algorithms (Wilcox and Gibson 2005 p. 120).

A digital image is a grid that contains a two-dimensional array of values representing luminance or chrominance. To achieve this format an analogue image must be first converted to the two-dimensional matrix, using pixel. For a pure black and white signal, without difference in saturation two values (0 and 1) would be sufficient. However most of the images are coloured and therefore require at least eight bits for the Red, Green and Blue signal.

Video coding involves four major steps: pre-processing, encoding, decoding and post-processing.

1. Preprocessing: Reduction of video noise and removal of information that the human visual system is not sensitive to.
2. Encoding
3. Decoding

[14] http://www.ncsu.edu/scivis/lessons/colormodels/color_models2.html, accessed the 14.06.2009

4. Post-Processing: Decoded pixel slightly deviate from the original pixel due to the lossy encoding scheme. Post processing includes deblocking filters to remove block artefacts (visible pixel block borders between two adjacent blocks).

3.2.3.3 Encoding

Compression can be achieved by two primary methods: Redundancy elimination and quality reduction.

A frame of motion video is full of duplication. The difference between one frame and its successor is usually very little. The background remains static and the action changes incrementally. A form of redundancy reduction uses the same methods as in audio compression (cf. chapter 3.1). Instead of describing the signal full each time, only the changes between successive frames are encoded. Efficient algorithms were developed to detect repetitive patterns and describe how they repeat (Wilcox and Gibson 2005 p.98).

3.2.3.3.1 Interframe encoding

In most cases subsequent pictures differ only in the position of the displayed objects. Usually the movement is consistent over a certain period of frames and does not change from frame to frame. If an object on an image moves in one direction over the course of several frames prediction methods can calculate the position of the object in subsequent frames. Object-oriented methods are not in use yet as the encoding process cannot distinguish between different objects in an image. Currently the so-called bock matching technology is deployed, where blocks of pixel (usually 16 x 16 frames) are compared on two adjacent images. Blocks that are very similar to each other are compared in there position and then the offset is expressed as a movement vector (Schmidt 2003 p. 128).

3.2.3.3.2 Intraframe encoding

Video codec's that compress data within a single frame are called intraframe codec's. They take a frame and divide it into individual blocks. Within one block the codec looks for repetitive information and removes it. In any given frame of television, there is a lot of redundant information – for example large areas of a single colour. The result is that a block can be described with less bits than before the compression was applied.

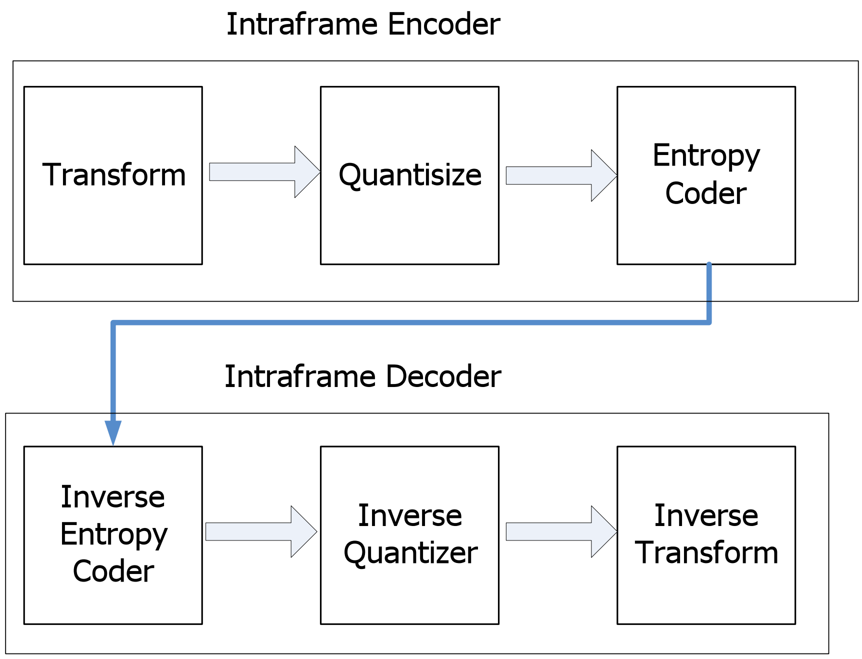


Figure 9: Encoder and Decoder Processes

Modern Intraframe codec's are using transform algorithms like discrete cosine transform (DCT). The H.261, H.263 codec's divide a frame into blocks of 8x8 pixel and then apply the DCT to each individual block. The DCT transforms the spatial domain of a codec in to a frequency domain (Firestone, Ramalingam et al. 2007 pp. 56-57). That means a conversion of pixel intensities into their frequency-based equivalents. After the DCT is complete a block of video data is described in terms of frequencies and amplitudes instead as a series of pixels. A frequency spectrum of a video signal can be displayed three-dimensional. Apart from the frequency for the time (one dimension), there exist frequencies for the structure of the image. Low frequencies indicate large image structures and modest transitions in luminance whereas high frequencies indicate fast small structures with abrupt changes in luminance (Schmidt 2003 p. 130). Low band filtering is a means for further data reduction by cutting off high frequencies in the image.

Intraframe compression removes spacial redundancies whereas Interframe compression removes temporal redundancy.

3.2.3.4 Quantisation

Quantisation is the process of reducing precision of the frequency domain. The frequencies and values that were determined by applying the DCT are divided by a fixed value and then rounded to the nearest integer (Firestone, Ramalingam et al. 2007 p. 59). The result is less information is needed to represent the frequency domain values and the outcome is a lower bit rate. One the other hand the process of quantisation means removing detail forever.

3.2.3.5 Entropy coding

Entropy coding follows the quantisation. It is a lossless encoding scheme that seeks to reduce the bit rate of the bit stream by eliminating redundancies within the bit stream. In H.261 and H.263 run-length and Huffman encoding schemes are used (Wilcox and Gibson 2005 p. 103). In run-length encoding, the codec compresses a string of identical bits by a token that indicates how many time the pattern repeats. Huffman encoding converts the most frequently repeated tokens into shorter bit strings.

3.2.3.6 H.264

H.264 is a standard for video compression, and is equivalent to MPEG-4 Part 10, or MPEG-4 AVC (for Advanced Video Coding) (Firestone, Ramalingam et al. 2007 pp. 342-348). The H.264 standardization effort is to provide significantly enhanced compression performance and provision of a “network-friendly” packet-based video representation. H.264 also delivers high video quality across the entire bandwidth spectrum—from 3G to HDTV and everything in-between (from 40 Kbps to upwards of 10 Mbps). All modern video conferencing endpoints are working with the H.264 standard.

Transmission rates for H.264 content[15]

Use Scenario	Resolution & Frame Rate	Example Data Rates
Mobile Content	176x144, 10-15 fps[16]	50-60 Kbps
Internet/ Standard Definition	640x480, 24 fps	1-2 Mbps
High Definition	1280x720, 24 fps	5-6 Mbps
Full High Definition	1920x1080, 24 fps	7-8 Mbps

H.264 uses techniques fairly different from MPEG-2 and can match the best MPEG-2 quality at up to half the data rate. H.264 also delivers excellent video quality across the entire bandwidth.

[15] http://www.apple.com/quicktime/technologies/h264, accessed the 25.07.2009
[16] Frames per second

3.3 Digital signal processors

Video conferencing is a very time sensitive process. In order to replicate a real-time conference feeling audio, video and data streams have to be encoded/ decoded and compressed/ decompressed in a very short amount of time. The Digital signal processor (DSP) is a specialized microprocessor to work in real-time computing (Wilcox and Gibson 2005 pp. 107-108). The DSPs digitize the analogue media and compress it in very little time intervals. Most DSP have constraints on latency, meaning the operation has to be successful in a certain amount of time.

Figure 10: Processing chain of analogue source signals

There exist two different types: Function and applications specific integrated circuits (FASIC) and programmable DSPs.

Programmable DSPs are in multimedia processing engines that perform multiple tasks. They can be used to compress either audio or video using a wide variety of algorithms. Some programmable DSPs process millions of operations per second (MOPS) other process billions of operations per second (BOPS). Modern DSPs have the ability to process different media (audio, video and data) at the same time.

FASICs are used in video conferencing as well. They are usually referred to by the function they were designed for. There are MPEG-2 FASIC's or audio FASIC's (e.g. G.722 or G.728)

Due to the nature of video conferencing as a real-time, two-way communication process: The video codec must operate in low delay mode. As a result the DSP cannot use advanced coding methods that are used for example to encode material for one-way viewing like DVD's.

3.4 Types of endpoints

Video conferencing endpoint can be classified into different categories, related to their field of deployment (Firestone, Ramalingam et al. 2007 p. 16-17). In this chapter the different forms will be introduced.

3.4.1 PC video conferencing

There is a wide range of software available for video conferencing experience on personal computers or laptops. Tandberg for example offers a product called "Movi", Polycom offers the Polycom PVX software ((VCC) 2008 p. 44-45). Those solutions are targeted for travellers that want to be reachable over video.

Figure 11: Emblaze VCON vPoint HD 8.0 desktop videoconferencing system[17]

Most modern laptops offer integrated microphones and cameras as well as the ability to connect wired or wireless to the Internet.

The en- and decoding of the media streams is a very resource intensive activity and a fast processor is needed to deal with the highly compressed audio and video streams. Other than in hardware solutions for video conferencing where dedicated DSP are ensuring the en- and decoding, in a PC the CPU takes over that task. Modern video conferencing clients offer the ability of sending and receiving images in HD quality. This

[17] http://visibleprocrastinations.files.wordpress.com/2008/04/vpointhd-01.jpg, accessed the 25.07.2009

requires both computers to have the latest spec in terms of processing technology, to be able to encode and decode the high resolution signal in time.

Current software-based video conferencing solutions offer Whiteboards[18] and ApplicationSharing[19] to enhance the collaboration experience. Whiteboards allow participants of a conference to write and draw images on a simulated canvas and develop schematics or take public notes during a video conferencing session. ApplicationSharing goes one step further by allowing the participants to use programmes installed on one of the participating machines in close collaboration.

More important for today's corporate meetings is the use of a presentation sharing functionality, where a host can share his screen via a second channel with the other conference participants.

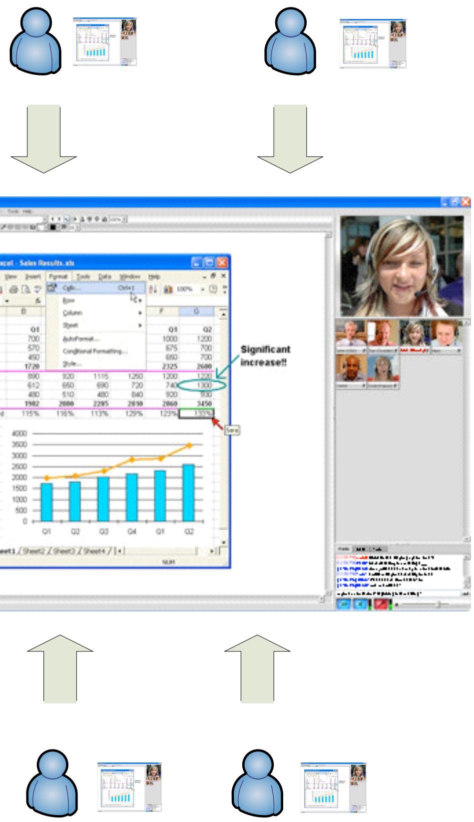


Figure 12: ApplicationSharing in Marratech's video conferencing client[20]

[18] White boards allow participants of a conference to write or draw images on a simulated canvas and develop schematics or take notes during a video conferencing session

[19] Application sharing is the ability to share a computer program with the different participants during a video conferencing session.

[20] http://1.bp.blogspot.com/_XYizHvjgLrA/RihS6t2sRpI/AAAAAAAAA8E/78XUV_h9X7k/s400/google+web+conference+software.jpg, accessed the 25.06.2009

3.4.2 Desktop video conferencing systems

Desktop conferencing systems are designed for single person or little group usage only. The codec, the screen and the audio and video inputs are generally integrated in one shell. The advantage is that the desktop systems are built compact and therefore easily transportable. They fit on the desk, next to telephone and computer. However, desktop systems are mostly limited in their functionality and are not scalable. It is not possible to attach a second screen or another microphone.

Figure 13: Tandberg 150MXP desktop video conferencing system[21]

3.4.3 Room conferencing systems

Room conferencing systems are designed for high demand in quality and scalability. Room systems can be individually configured to the needs of the customer. For example it is possible to use multiple microphones and cameras with one room system codec and you can usually connect at least two screens to the endpoint. The microphones are typically boundary microphones that are placed on the table in the middle of the room. The bigger the room the more microphones are used. Normally only one or two microphones are used due to the increasing problem with interference.

Because the microphones and speakers are in very close proximity feedback reduction plays a very important role in video conferencing. Elaborate echo cancellation algorithms are integrated in the codec and remove the loudspeaker signal portion from the recorded audio signal. Echo cancellation is included in every video conferencing solution. Some people prefer to use headsets and so eliminate the need for signal treatment.

Room conferencing systems are designed to be used by multiple persons at the same time. The camera can be mounted to the wall and can be tilted and panned to focus on different speakers in the room. In comparison to desktop system is the camera of higher quality with an all-in-all increased image quality.

In order to easily operate the equipment Room conferencing systems come with a remote control that offers camera, codec and sound control.

Figure 14: Tandberg remote control[22]

In general room systems are designed to be built-in racks. However some manufacturer like Tandberg and Polycom offer room systems that can be moved very quickly. Tandberg’s Maestro codec is built on wheels and can be transferred to a new location immediately. The codec lacks big screens and it is therefore necessary to provide external screens or projectors.

[21] www.tandberg.com, accessed the 25.07.2009

Figure 15: Tandberg Maestro MXP codec[23]

Modern room systems are HD capable and offer a high-resolution camera. In meeting rooms video conferencing units are often used in conjunction with projectors to display the images on the wall.

Room conferencing systems provide a higher bandwidth than desktop systems and they support different protocols. Most of the codec's are not only equipped with RJ45 sockets to support the H.323 standard but also provide sockets for ISDN lines and therefore are H.320 compliant.

Another important feature of meeting room codec's is the support of presentation sharing. At the back of the codec there are either digital inputs (DVI) or analogue inputs (VGA) for attaching computers to the codec and sharing the screen with the other participants.

3.4.4 Telepresence systems

Telepresence is the endeavour to make video conferencing experience as realistic as possible. The goal is that all participants in the meeting have the impression they are actually in the same room just a few feet away from each other. The idea behind is that normal video conferencing systems can only replicate a face-to-face meeting to a cer-

---

[22] www.tandberg.com, accessed 25.07.2009

tain extend. The user always has the feeling that there is the medium in between the participants which affects the way communication is held.

In video conferencing meetings local coalitions can form, in which participants tend to agree more with those in the same room than those on the other end of the line. This behaviour is encouraged through technology like the mute button, where one side can block the audio channel to the other side.[24]

Telepresence tries to limit the presence of the communication channel so that participants don’t feel the separating element anymore.

This is achieved by

- HD cameras which can produce images at 1080p60
- High quality speaker systems and spatial audio codec’s
- Very large displays with screen sizes around 65”
- Very powerful codec’s that can process HD video and audio in real-time

HD cameras are installed just above the screen and create an atmosphere as if you are looking your counterpart straight in the eyes.

Microphones and speakers work together to create spatial audio. Instead having a mono signal coming out of all speakers with the same intensity, stereo effects are created, so that the conference participants are able to locate the source of an audio signal.

Manufactures of Telepresence systems don’t sell only the equipment. They sell a room fitted exactly to the needs of a Telepresence collaboration environment. One room design costs around £170.000[25].

Each major manufacturer of video conferencing equipment has released its own Telepresence solution package. The rooms usually consist of a special composition of chairs and a table. Special lightning is in place as well as acoustic wood walls for sound improvement. Tandberg Telepresence solution with three screens is called T3 and can deliver 1080p HD quality on an 18 Mbps bandwidth[26].

[23] www.tandberg.com, accessed the 25.06.2009

[24] Lynne Wainfan, P. K. D. (2004). Challenges in Virtual Collaboration, National Defense Research Institute

[25] Excerpt TANDBERG pricing Q4 2008

[26] Tandberg Telepresence 2009 www.tandberg.com/totaltelepresence accessed the 24.06.2009

Figure 16: Tandberg Telepresence solution T3[27]

Telepresence systems are the non-plus-ultra solution for a realistic video conferencing experience. The latest technology is capable of creating the impression of indifference in space. However, this technology targets big or medium international companies with the necessary funds to operate such technologies. You need at least two rooms to experience the full capability. Further costs result of the provisioning of a dedicated high speed connection between those two rooms with low latency.

[27] www.tandberg.com/totaltelepresence, accessed the 25.06.2009

4 Multipoint Control Unit (MCU)

In a video conferencing network a Multipoint Control Unit, or bridge administrates and controls multipoint conferences. Two endpoints can make a point-to-point call but when three or more endpoints want to hold a video conference, a MCU is required. The MCU receives all incoming media streams from the endpoints, processes them and sends to each participating endpoint a new individual stream.

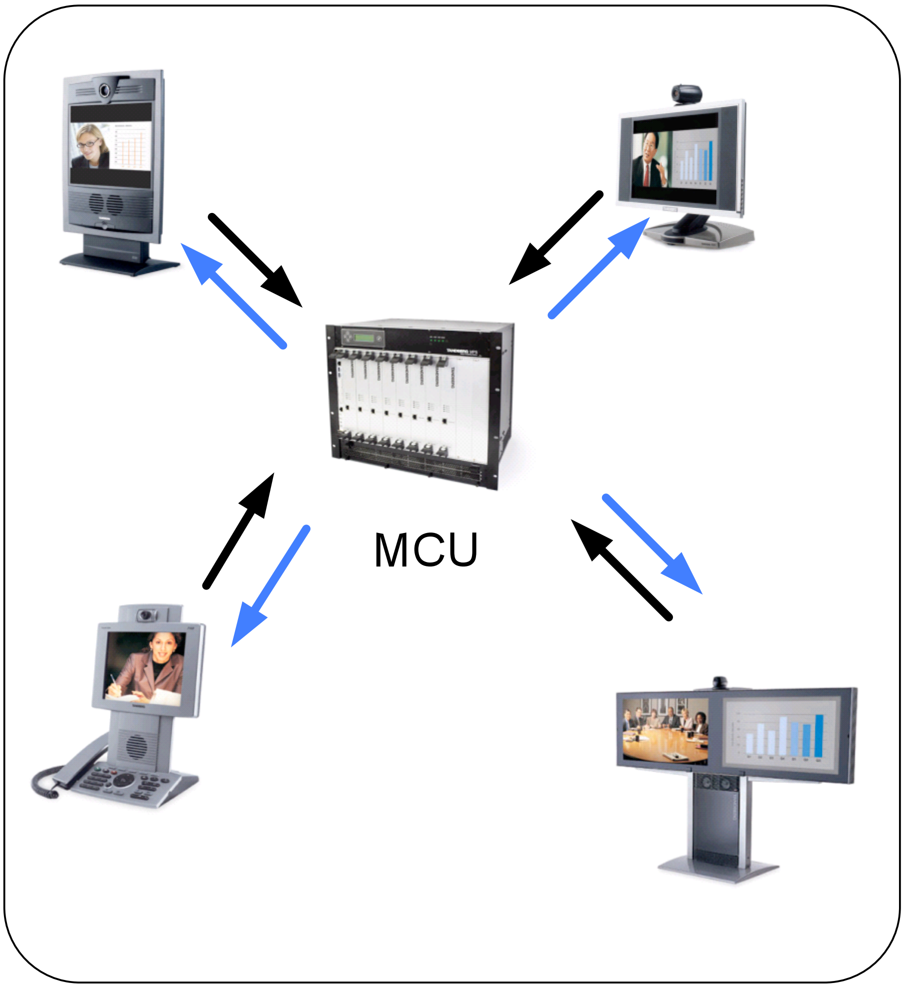


Figure 17: Multipoint Control Unit (MCU)

The MCU can be implemented in the VC network as a software or hardware solution.

Before a conference can start on a MCU, virtual rooms have to be created and configured. A virtual room is comparable to a chat room, where the different conference participants meet and exchange audiovisual information. For each individual conference a virtual room is needed. On the MCU you can create and configure virtual rooms and specify certain characteristics for example: maximum call duration, amount of participants allowed, and maximum bandwidth etc.

Each virtual room is represented by a dial-in number, which follows the E.164 addressing scheme[28] and consists of numeric values. Prior to the conference the participant are given the number of the virtual room and they can connect by dialling the number from their endpoint.

When you dial in a MCU instead of only seeing and hearing one person like in a telephone call you can see and hear different persons on screen. The more endpoints are participating in the conference the smaller the space for each participant on your screen. In the image below gives an example where four people hold a video conference.

Figure 18: Continuous presence mode (2x2)[29]

4.1 Conference modes

With modern MCU it is possible to have conferences with sixteen or more connected users. With so many people on one screen it would be hard to distinguish each individual. Therefore the conference administrator has additionally the possibility to change the conference layout and conference mode.

The conference layout determines how the different video signals are displayed on the screen of each individual conference user. In the graphic below you see possible layouts for conferences with up to six participants.

There are two main conference modes a MCU can operate in (Firestone, Ramalingam et al. 2007 pp. 11-13):

[28] E.164 is an ITUT recommendation which defies the international telecommunication numbering plan. It defines the format of telephone numbers.

[29] (VCC), M. d. K. f. V. (2008). Videokonferenz Handbuch, TU Dresden.

4.1.1 Voice switching mode

In voice switching mode the MCU displays on all the receiving endpoints the person that is currently speaking full screen. The MCU monitors audio levels and automatically displays the person who sends the highest voice energy level, because it assumes that this person is speaking. When another person produces higher audio levels, the audio and video automatically switch to the new speaker.

The person who has the floor sees the previous speaker. This has two reasons. First each endpoint has a “self view” function built-in, so if the speaker wanted to see himself speaking, he could switch to “self view” mode. Secondly and more important is the image that is sent to and from the MCU is being processed. Sending the media stream back and forth and processing adds noticeable delay to the signal. This can confuse the speaker who is seeing a delayed picture of himself on the screen.

4.1.2 Continuous presence mode

As the name of this mode already implies the participants are constantly seen on the screen. The MCU tiles the images of multiple participants into a single composite video. This composite video is then sent to each individual participant. How the MCU divides the sub pictures on the output stream decides the layout chosen by the conference administrator.

Mixed modes between the voice switching mode and continuous presence mode are possible.

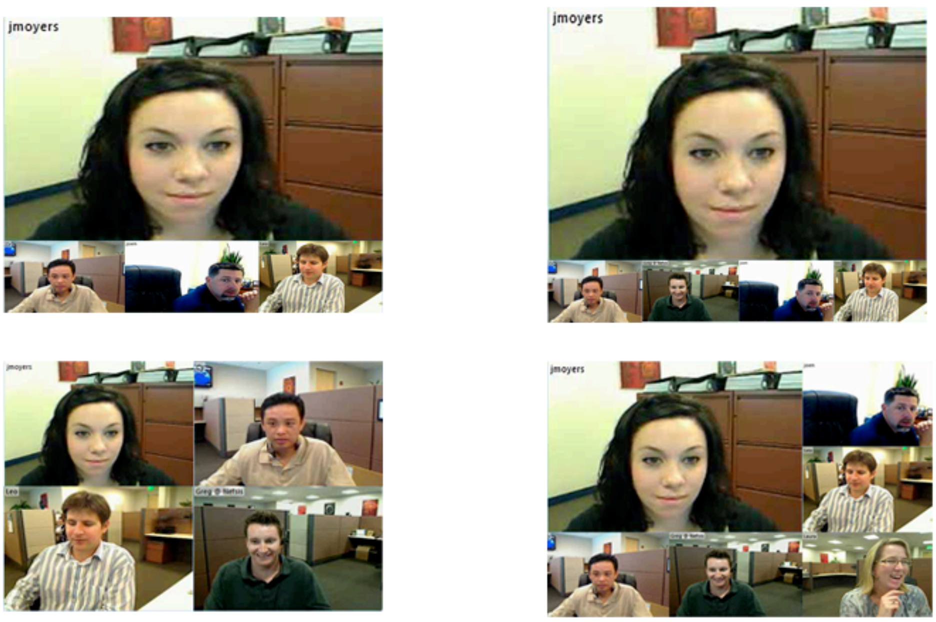


Figure 19: Different conference layout in continuous presence mode

4.2 Components of a MCU

A MCU has to coordinate different audio and video streams as endpoints provide media with different characteristics (codec's, bit rate, frame rate, picture size). Furthermore has a MCU the capability to convert video and audio streams to the need of the other participating endpoints.

The MCU consists of different components, which are crucial to establishing a successful multipoint conference:

4.2.1 Control plane

The control plane is responsible for establishing a connection between the MCU and the endpoint, which wants to participate in the conference. It listens on the network for incoming connections. When an endpoint wants to connect itself to a new or existing conference the control plane provides it with the necessary information about the audio and video capabilities of the MCU. Then both sides negotiate the type of media they are going to use and open logical channels for media streaming. This process is very important as both sides determine which media, encoding and decoding schemes and bandwidths are supported. Afterwards the control plane connects the endpoint with the mixer and the media plane.

A basic MCU is able to work with a control plane only. All endpoints would have to send their audio and video streams with the same characteristics and the MCU would operate in voice switching mode. No media would actually being processed.

Due to the difference in the media characteristics of today's meetings (codec, RTP payload type, picture size, frame rate), the MCU deploys a media plane which is able to transform audio and video streams and change them to the different requirements of the endpoints in a conference.

4.2.2 Media plane

The media plane contains infrastructure that processes media streams and includes audio and video mixers. The media plane manages Real-time Transport Protocol (RTP) and Real-time Control Transport Protocol (RTCP) port allocation and may control a DSP for setting audio and video stream characteristics (like codec, RTP payload type, picture size, frame rate).

The media plane is also responsible for detecting stream failures, like incoming RTP stream loss.

4.2.2.1 Video Mixer/ Compositor

The video mixer is part of the media plane.

After the control plane has successfully negotiated the video stream characteristics and conference management has determined the type of video presentation required. The video mixer/ compositor is responsible for creating the overall video experience.

It receives and decodes incoming streams in various formats and creates appropriate output streams for the endpoints based on conference policy (bandwidth/ frame rate set by the administrator and supported by the participating endpoints).

A video mixer can receive a wide range of streams in various bit rates and picture formats and compression schemes and it must be able to transcode or transrate those streams in the output streams manageable by the endpoints.

Transrating is downspeeding a certain video stream from a high bandwidth like 768 Kbps to a lower bandwidth (e.g. 384 Kbps). Transrating is applied when an endpoint with a very good network connection and an endpoint with a slower network connection participate in a MCU controlled conference.

Transcoding is the process of re-coding the audio or video information using another codec. If all participating endpoints in a conference support H.264 except of one, the MCU needs to transcode all video streams to and from that codec (Firestone, Ramalingam et al. 2007 p. 229).

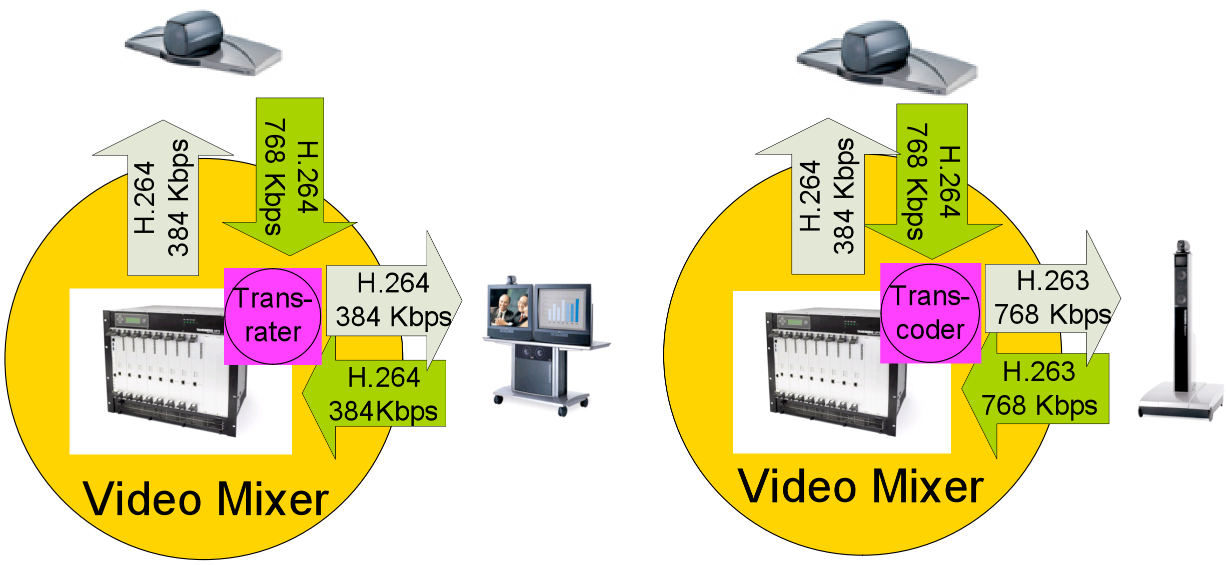


Figure 20: Transrating and transcoding in a MCU

4.2.2.2 Audio mixer

The audio mixer is responsible for selecting the input streams and summing the streams to an output stream for each participant.

The audio mixer like the video mixer belongs to the media plane.

In large conferences the audio mixer typically only selects the three or four loudest input streams and sums them up, because the human ear is only capable of differentiating between three or four distinct talkers (Firestone, Ramalingam et al. 2007 p. 31).

Conference participants must not hear themselves back from their codec. This would very much disturb the video conferencing experience as each speaker heard himself talking with a few milliseconds delay.

To avoid self-echo, every endpoint that contributes a stream for the audio mix receives a unique audio stream, which doesn't contain the audio from the very same participant.

The approach is called N-1 summation, where N is the number of mixed streams and the stream returned to a conferee is the summation of the mixed streams, minus the one contributed by that individual.

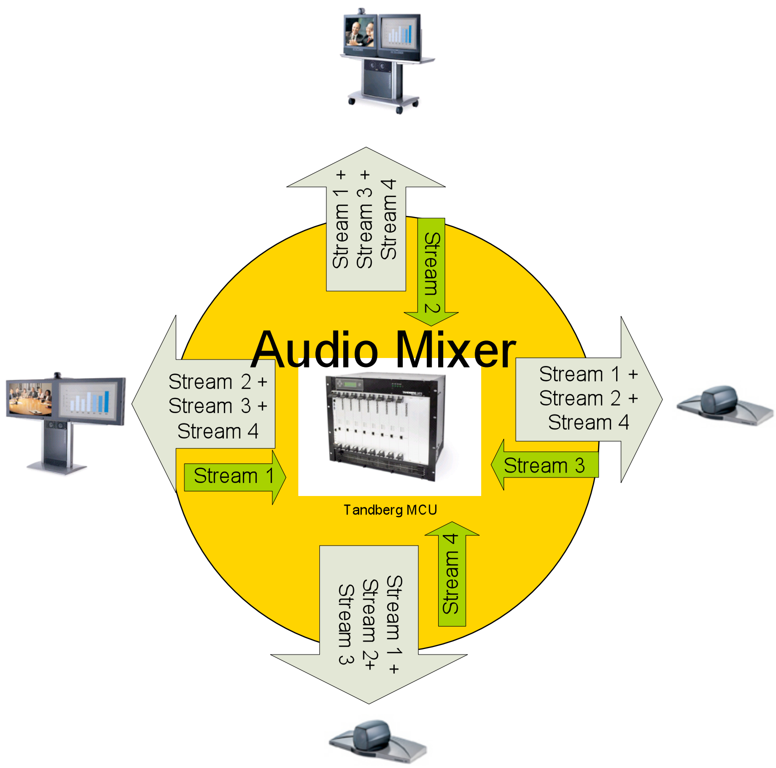


Figure 21: Audio mixer in a MCU

5 Gatekeeper

The gatekeeper comprises different functionalities in a H.323 video conferencing network. You can distinguish three main responsibilities of a gatekeeper: address translation, bandwidth management and security. If a MCU is deployed in a VC network, a gatekeeper is mandatory. Furthermore all units which are participating in a MCU controlled conference have to be registered with the gatekeeper.

Figure 22: Tandberg gatekeeper[30]

Gatekeepers are optional in a video conferencing network. However, if a gatekeeper is implemented all endpoints of the zone have to be registered with it.

Address translation: A gatekeeper enables different endpoints on a video conferencing network to be addressed by a unique E.164 number or H.323 ID. Without a Gatekeeper it is only possible to dial another endpoint via its IP address.

The E.164 Dialled Digits addressing scheme assigns a numeric string to each device and is the most common way of endpoint aliasing (Firestone, Ramalingam et al. 2007 p. 187). E.164 digits may include any digits between 0 and 9 and have a recommended maximum of 17 digits. VC administrators assign numbers according to the dial plan. The idea behind is that the E.164 dialling scheme is similar to the phone system and so easier to use than dialling full IP addresses. When an endpoint wants to call another endpoint, it presents the address it wants to call to the gatekeeper using a protocol knows as RAS. The gatekeeper tries to resolve this address to the corresponding IP address and supplies the calling endpoint with information about the called endpoint. If the gatekeeper does not have the information it automatically queries other gatekeeper in the network.

A dial plan makes sure that the same E.164 numbers are not assigned multiple times to different endpoints. In the end it is up to the VC administrator to ensure that the dialling plan is consistent. Each endpoint should have a unique H.323 ID and E.164 number.

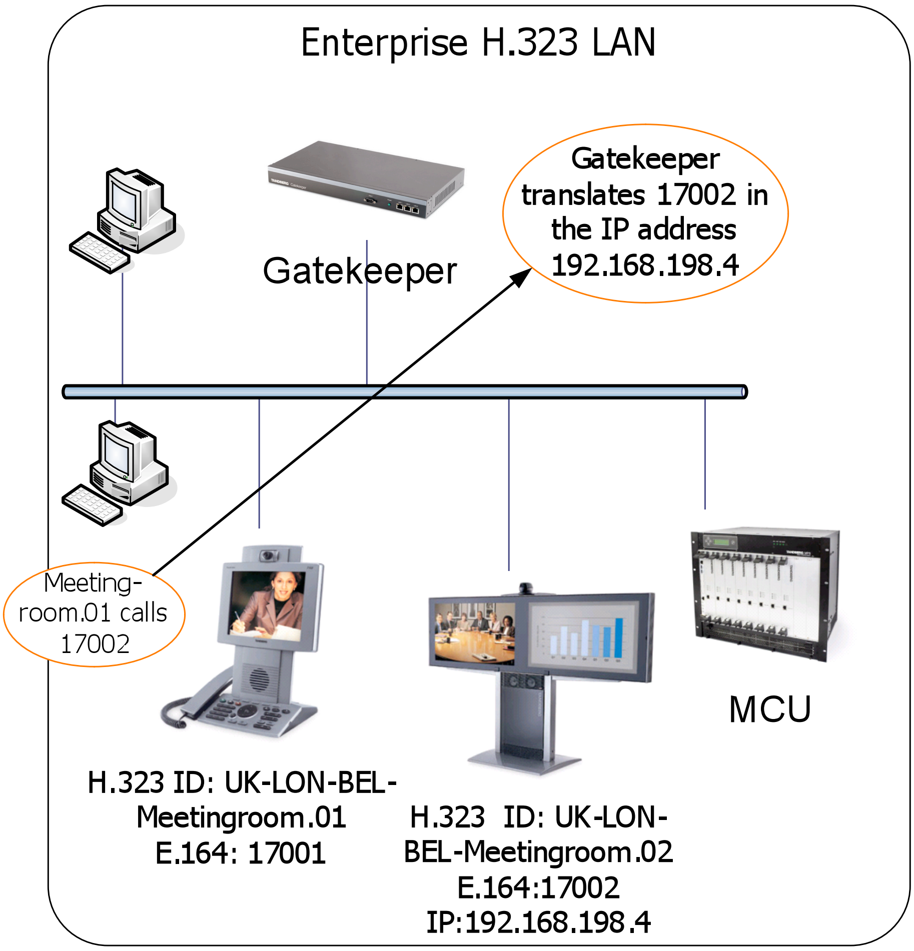


Figure 23: Address translation of the gatekeeper

The H.323 ID is a string-based alias assigned to each endpoint, such as Last-name.Firstname (Friemel.Frank). The advantage of an H.323 ID to an E.164 alias is: A H.323 ID can have a meaningful name, which relates to the location of an endpoint.

[30] www.tandberg.com, accessed the 25.07.2009

Table 2: Examples for E.164 and H.323 ID

E.164	H.323 ID	IP
17001	UK-LON-BEL-Meetingroom.01	192.168.198.3
17002	UK-LON-BEL-Meetingroom.02	192.169.198.4
17004	Friemel.Frank	192.168.198.7

A second very important functionality of a gatekeeper is bandwidth management. A gatekeeper can limit the overall bandwidth allowed in a video conferencing network and therefore help to prevent congestion of the LAN network.

Bandwidth management: The gatekeeper can allow or deny bandwidth request between endpoints (Firestone, Ramalingam et al. 2007 p. 209). In a highly congested network the video conferencing administrator can limit the overall bandwidth allowed by all endpoints on the gatekeeper. A requirement is the registrations of all VC devices to the gatekeeper. Furthermore a gatekeeper can segregate a video conferencing network into different subzones, and apply different bandwidth limitation per subzone. It is possible to grant endpoints that are used by higher management a higher bandwidth than VC units in normal conference rooms. Special subzones can also be applied to inferior network segments with congestions problems.

A lower bandwidth automatically reduces the audio and video quality of a call. If in a zone bandwidth limitation apply and a video conference already takes up all the available bandwidth, the gatekeeper can either reject additional conferences over that network segment or downspeed the existing conference (i.e. the bandwidth of 384 Kbps can be limited to 256 Kbps).

Bandwidth limitation can be very beneficial for the performance of the video conferencing network. However, the bigger the network is the more difficult it is to implement rules that remain simple and are effective. In large video conferencing networks with multiple sub networks bandwidth limitation becomes more difficult to manage.

Admission control: The admission control is a feature that allows the network administrator to restrict access to a H.323 video conferencing network (Firestone, Ramalingam et al. 2007 p. 210). When a new unit wants to register itself to a gatekeeper in order to make calls it sends a registration request (RRQ) to the gatekeepers IP address. The gatekeeper then looks up its admission policy and grants or denies the registration request. Rules in the gatekeeper can be set up to decide who gets access (including and excluding rules based on IP addresses). Admission control is especially important if the gatekeeper resides on the Internet to prevent unauthorised users access to the video conferencing network.

When an endpoint initiates a call to another endpoint, the gatekeeper can reject the call request based on different criteria (e.g. time of the day, access rules and bandwidth restrictions).

Zone management: Zone management is used for communication between the different gatekeepers on a network (Firestone, Ramalingam et al. 2007 p. 210). A zone is composed of VC infrastructure actively registered with one gatekeeper, comprising endpoints, MCU's and gateways. Each endpoint only belongs to only one zone. Gatekeepers are involved in call signalling and sometimes responsible for the routing of the media streams. For that reason it should always reside close to the endpoints participated in the call. If the endpoints are in Europe and the gatekeeper is situated in the US each time a call is initiated traffic goes all the way to the US and back. This adds unnecessary delay to the call setup. In large international video conferencing networks, there are several gatekeeper deployed in different zones which are linked (neighboured) to each other.

6 Gateways

“A gateway is a network point that acts as an entrance to another network.”[31]

Next to the H.323 standard exists another standard H.320 which describes video conferencing over an ISDN (Integrated Services Digital Network) based network (Firestone, Ramalingam et al. 2007 p. 208).

Figure 24: Tandberg gateway[32]

The main disadvantage of the H.323 protocol stack is that it is designed for LAN/WAN environments. H.323 heavily depends on the existing network configuration and is designed for the use within one network. Interconnection capabilities between multiple networks are limited. In order to get out of a corporate network firewalls have to be traversed, which heavily restrict packet exchange of network devices. E.164 numbers and H.323 IDs are not centrally administered like the phone numbers given out by the phone companies. Each video conferencing administrator organizes his or her dial plan individually. In most companies it is very easy to make calls to internal endpoints, but more difficult to connect to external endpoints or networks.

A gateway offers the possibility to break out of the H.323 environment and call external endpoints either direct (gateway to endpoint with ISDN connection) or connect to another LAN over a second gateway (gateway to gateway) (Wilcox and Gibson 2005 p. 124).

---

[31] Definition of a gateway by whatis.com, accessed the 25.06.2009

[32] www.tandberg.com, accessed the 25.07.2009

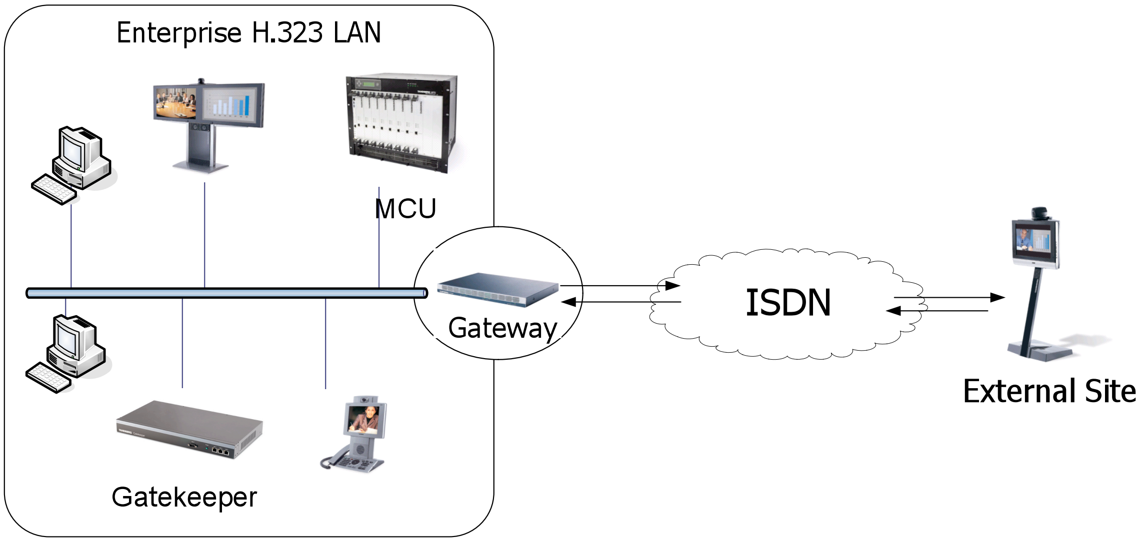


Figure 25: Gateway – transition between H.320 and H.323 networks

ISDN is available in most countries in the world.

The endpoints displayed in Figure 25 can call the external site, because the H.323 network has an integrated gateway. With that gateway it is possible to call any ISDN phone or ISDN video conferencing unit in the world.

A H.323 endpoint can communicate with another terminal on an H.320 network, over a H.323 gateway. The term gateway refers to the ability to bridging any two different networks. In most cases they are used to bridge IP and ISDN networks.. The number of simultaneous connections allowed through a gateway depends on how many PRI‘s are connected from the ISDN side to the gateway and the available bandwidth of the LAN network the gateway is connected to.

7 Video conferencing at Google

7.1 Hardware-based video conferencing at Google

The H.323 video conferencing system implemented in the Google offices is solely based on components of Tandberg. Tandberg is next to Sony, Polycom and RADvision one of the biggest manufacturers on the video conferencing market.

The aim after installing video conferencing infrastructure in the offices was being able to establish „human-like"-interaction even over long distances. Google offices are dispersed all over the world. In EMEA[33], Google has over 27 offices in 21 different countries. Globally the company deploys around 2500 Tandberg endpoints[34] and other H.323 infrastructure. To enable staff to communicate quickly and easy over long distances, Google decided to implement videoconferencing solutions in all of their offices.

At Google all VC units are connected over the internal corporate LAN network.

Main ideas behind implementing those technologies were cutting travel costs and providing immediate access to teams in other locations. Without any preparation time, meetings spanning remote locations are instantly possible and participants do not suffer fatigue from travel.

Another important aspect of modern video conferencing technology is the possibility to interview potential new employees from distant countries and save travel costs. Google is an international company with employees from all over the world. In the interview process the first stage is phone screening. After the applicant has passed this he is invited into one of the offices for a video conference. This way, the candidates are able to connect from any Google office to recruiters, prospective colleagues and managers.

The video conferencing system in the London office is quite diverse. Most of the meeting rooms have systems that are scaled to the size of the room. The smallest meeting rooms with space for only three or four people are fitted with a system that resides on the table and can be operated without remote control. Meeting rooms with more than four people have a big LCD screens mounted to the wall and microphones installed on the conference table.

---

[33] EMEA – Europe, Middle East and Africa

[34] Information retrieved from Google's Tandberg Management Suite (TMS)

Figure 26: Conference room in Google Haifa

The biggest meeting room is a room with seats for sixty people and four projectors hanging from the ceiling. This way simultaneously to a video conference two presentations can be locally shown. The room is equipped with three different cameras. Depending on the situation one of those three cameras can be selected as a video source for the video conference.

Figure 27: Newton – Biggest conference room in the Google London office

7.2 Implementation of a video conferencing solution

Every codec is connected to the LAN network over a RJ-45 socket in the back. A CAT5 cable connects the codec to a socket in the wall, which is connected to an access switch in a communication room.

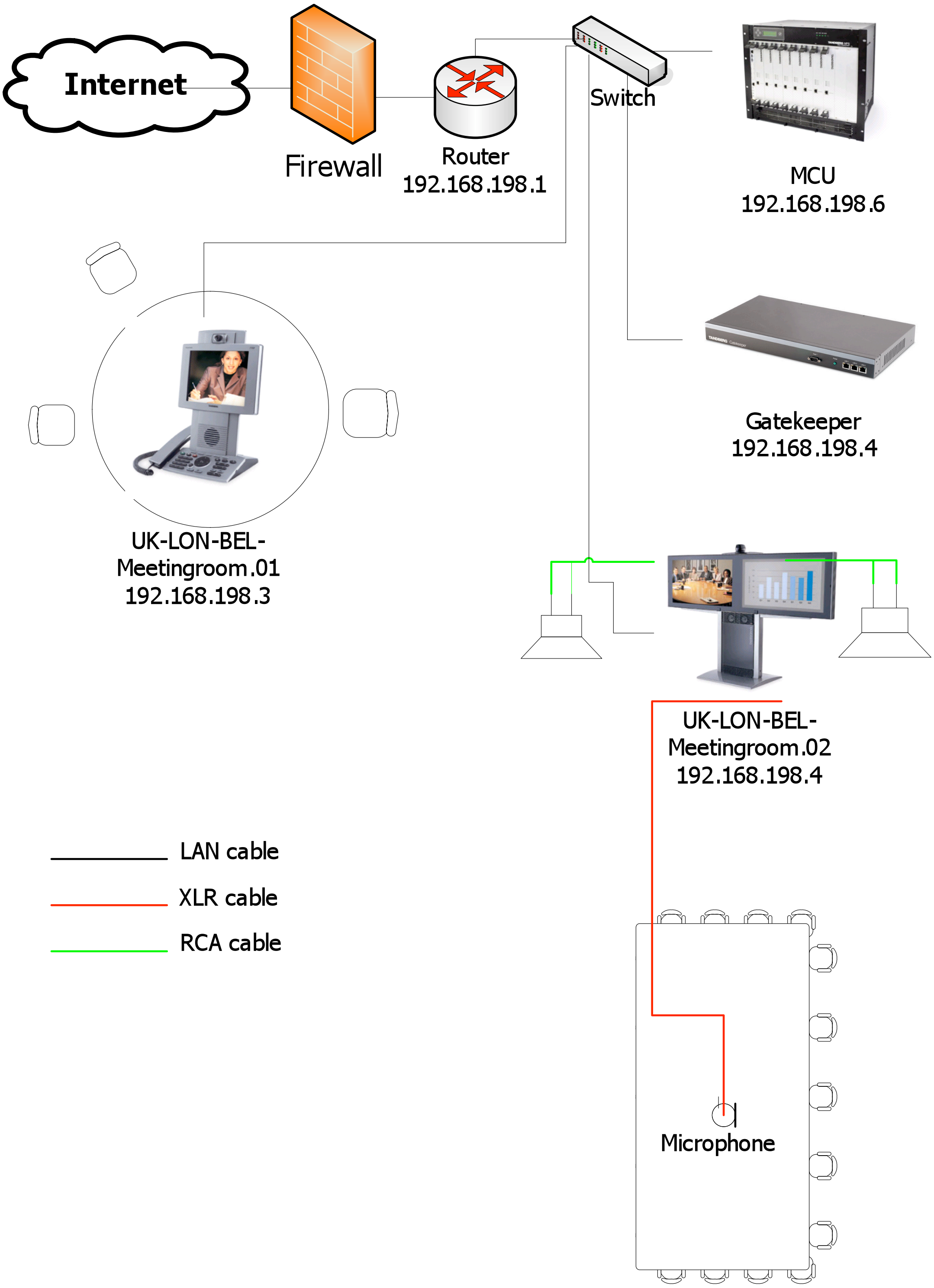


Figure 28: Example for a conference room setup

In the communications room all network infrastructure like switches, routers and servers are installed. The network devices in an office are interconnected over different switches. To simplify the schematics (Figure 28) the gatekeeper MCU and both endpoints are connected to one switch.

If endpoint UK-LON-BEL-Meetingroom.01 wants to call UK-LON-BEL-Meetingroom.02, the endpoint UK-LON-BEL-Meetingroom.01 sends an Admission Request (ARQ) to the gatekeeper. The ARQ contains the called units E.164 alias/ H.323 ID and the requested bandwidth. The result of an ARQ could be an Admission Confirm (ACF) from the gatekeeper. This way the connection is confirmed and the gatekeeper provides the remote IP address and the call's bandwidth to the calling endpoint. Due to bandwidth restrictions or other reason the gatekeeper can also deny the connection by sending an Admission Reject (ARJ).

If the call is confirmed by the gatekeeper, the capabilities exchange follows. Part of the connection process is to identify the capabilities of each device participating in the video conferencing session. The Capabilities Exchange (CapEx) is a process were information about the type and combination of protocols the VC units support (G.711 and H.263 etc.) are exchanged and the protocols used in the call are negotiated.

The final step is the opening of the media channels for audio and video transmission between all video conferencing endpoints in that call {Firestone, 2007 #8 p. 192}.

7.3 Software based video conferencing solutions at Google

Due to the acquisition of Marratech[35] (2007), a vendor of video conferencing software Google was able to offer its employees an alternative software based solution for voice and video collaboration. Marratech's product was a desktop video conferencing tool with Whiteboard and ApplicationSharing functionality. It was available for Windows, Mac and Linux and was interoperable with the existing Tandberg videoconferencing network due to its H.323 interoperability.

[35] http://www.pcworld.com/article/131048/google_buys_video_conferencing_software.html, accessed the 27.05.2009

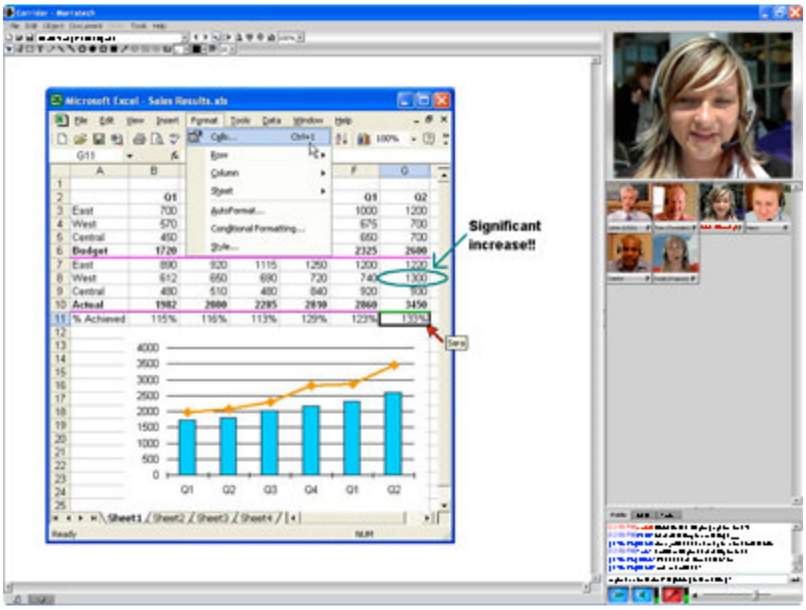


Figure 29: Marratech web conferencing software[36]

Although it met the criteria of Google's quite diverse IT environment it never became accepted from the Google staff. Very few people actually used the tool on a regular basis.

The main difficulties with the program resulted from:

5. Instability on Apple Mac computers – The programme crashed frequently and never worked flawlessly on a Mac OS X platform.[37]
6. Additional hardware required – The only computer that came with a built-in camera and microphone was the Apple Macbook laptops. All other computers needed to be upgraded.
7. Complexity – the tool was full of functionality. It was perceived by the users from Google little user friendly.
8. Picture quality was far behind modern video conferencing solutions. Marratech connected a H.323 call with the H.261 video codec.[38]

Marratech was a cost-effective alternative to a video conferencing desk system with cutbacks in usability and quality.

[36] http://1.bp.blogspot.com/_XYizHvjgLrA/RihS6t2sRpI/AAAAAAAAA8E/78XUV_h9X7k/s400/google+web+conference+software.jpg, accessed the 27.07.2009

[37] http://www.marratech.com/forum/index.php?showtopic=2129&hl=Mac, accessed the 25.05.2009

[38] http://www.marratech.com/forum/index.php?showtopic=1196, accessed the 25.05.2009

Google's reason to purchase Marratech was to engage the developers from Marratech in a Google project to enable video conferencing experience over Google's web based email client "Google Gmail"[39].

Google's major competitors like Skype®[40] and Microsoft® already offered video capability within their chat clients.

The long-term goal was to deprecate Marratech's video software and replace it with a new voice and video collaboration functionality within Gmail.

The product is called "Gmail Voice and Video"[41] and has already been released in November 2008[42]. Google employees are encouraged to use Gmail Voice and Video instead of using a video conferencing room. The current version of Gmail Voice and Video doesn't offer H.323 integration. Therefore desktop video conferencing cannot be used in conjunction with the Tandberg codec's in the meeting rooms.

39 www.Gmail.com, accessed the 25.05.2009

40 http://www.skype.com/intl/en-gb/allfeatures/videocall/#high-quality-video, accessed the 25.05.3009

41 www.mail.google.com/videochat, accessed the 25.05.2009

42 http://gmailblog.blogspot.com/2008/11/say-hello-to-gmail-voice-and-video-chat.html, accessed the 06.06.2009

8 Gmail Voice and Video

Google's free of charge webmail service is called Gmail, short for Google Mail. Gmail is designed as a unified collaboration solution with the possibility to email, instant message, and make voice and video communication.[43]

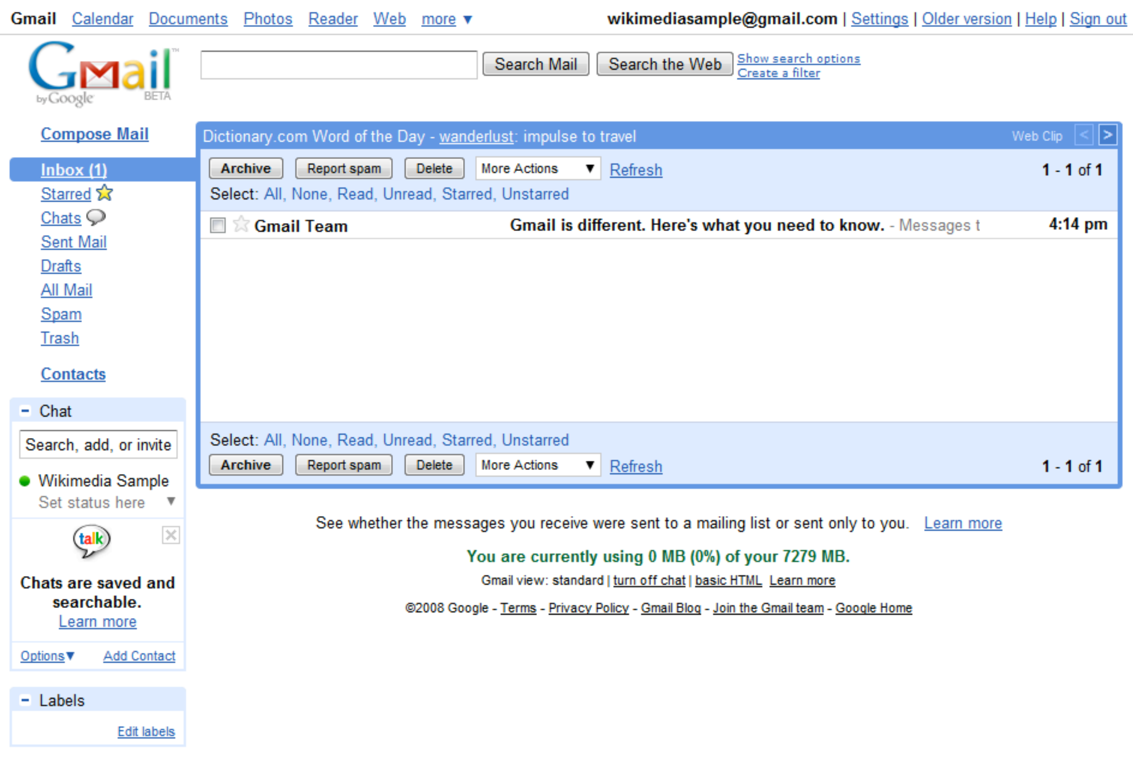


Figure 30: Gmail web interface[44]

Integrated in the Mail client is a chat functionality, which allows the Gmail users to chat to other users on the Gmail network. The chat offers emoticons[45] and the ability to chat with multiple people at the same time. Figure 31 shows the interface of Gmail chat window. All conversations can be saved in the mail client and reviewed by a later date. Chats can only be initiated to people that are in your contact list. Both sides have to agree to appear in each other's list to avoid being contacted by strangers.

[43] http://www.google.com/apps/intl/en/business/messaging.html#mail, accessed the 25.07.2009

[44] http://upload.wikimedia.org/wikipedia/en/6/6f/Gmail-01-01-09.png, accessed the 25.07.2009

[45] Emoticon is a textual portrayal of a writer's mood or facial expression.

Figure 31: Google Gmail chat function

To enable the video functionality it is necessary to install a plug-in on your computer which then enables Gmail to talk with your periphery (camera and microphone). The plug-in is available for all major platforms and browsers, except Linux (Last update: May 2009).

The VC client offers the same layout as the normal chatting window, with the ability to chat while you are having a video conference (see Figure 32).Additionally the window contains a small video of the far-end above the chat box. Like most videoconferencing solutions Google Voice and Video plug-in also offers the possibility to mute the microphone or to alter the volume of the incoming audio signal.

Important during the development of the plug-in was the design of an easy-to-use interface, which should stimulate end-users to adopt the technology quickly.

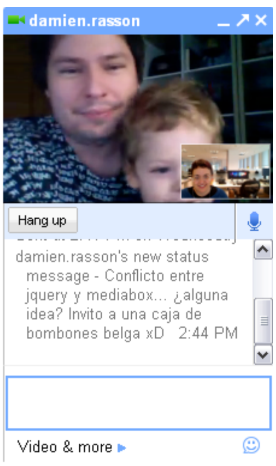


Figure 32: Google Gmail Voice and Video Chat screen[46]

[46] http://imagenes.sftcdn.net/es/scrn/75000/75586/3_6.jpg, accessed the16.07.2009

Before the product was launched it was tested by Google employees. The most appreciated feature of Gmail Voice and Video was the consolidated approach of the Gmail application. For your email, your chat and your video conference it was sufficient to have only one web browser window open.

8.1 Functionality of the Gmail Voice and Video plug-in

To enable your computer to do voice and video, special software has to be installed on your computer. For each browser that is running on your computer a plug-in is installed in the local settings of your operating system. The browser[47] is then able to run the Google Voice and Video plug-in when launched. The plug-in contains two pieces: a H.264 en- and decoder for the video compression and a voice (ISAC) Global IP Solutions (GIPS) codec[48] for the compression of the audio[49] .

8.1.1 H.264 en- and decoder

In chapter 3.2.3.6 the H.264 codec was in detail described. Google uses a proprietary version of the encoding and decoding sequence, different to the ITU standard[50]. With their own development Google wanted to tackle issues with regards to packet loss and jitter that are always problematic in real-time applications over the Internet.

8.1.2 Global IP Solution iSAC – audio codec

iSAC is an adaptive VoIP audio codec, which is designed to deliver wideband audio quality over IP networks (Global_IP_Solutions 2008).

Wideband meaning that the codec is able to encode frequencies in the range from 50 – 8000 Hz. The frequency range supported by normal public switched telephone network (PSTN) is around 200 - 3400 Hz (Linden 2008 p. 3). The sampling frequency of the iSAC audio codec is 16000 Hz.

The average human ear can detect sounds in a frequency range of approximately 20 Hz to 20,000 Hz. This would need a sampling frequency more than twice as much around 40 kHz[51]. Modern CD recordings use a sampling frequency of 44.1 kHz. With VoIP technology it is theoretically possible to transmit in CD quality, covering the whole spectrum. Due to unpredictable network conditions, like jitter and congestions and the

[47] In total nine different browsers are supported.

[48] http://www.gipscorp.com/products/voiceengine.php, accessed the 16.07.2009

[49] http://code.google.com/apis/talk/open_communications.html, accessed the16.07.2009

[50] http://www.itu.int/osg/spu/newslog/ct.ashx?id=fac75170-6f46-4ff3-b576-d432563cacd&url=http%3a%2f%2fwww.itu.int%2fitudoc%2fgs%2fpromo%2ftsb%2f87066.pdf accessed the 25.06.2009

need for higher bandwidth to transfer CD quality in real-time, the industry has chosen 16 kHz sampling frequency as best trade-off between bit rate and speech quality for wideband audio.

The codec is proprietary and was developed by Global IP Solutions. Due to its design it can deliver better than PSTN quality at dial-up modem data rates and adjust transmission rates from a low of 10 kbps to a maximum of 32 kbps. The codec matches the bit rate automatically to network conditions.

The fundamental role of a codec is to reduce the bit rate required for conveying the speech data. Most speech coding standards designed for traditional circuit-switched systems are fixed rate codec's, where the bit rate is the same for each coded speech block. As discussed in chapter 2.2, circuit switched networks offer a constant bandwidth for the time of transmission (Global_IP_Solutions 2008 p. 3).

Variable rate codec's are designed for packet-switched networks, where the available bandwidth constantly changes.

iSAC is designed for real-time applications like distance learning, multi-user gaming and conferencing over networks with constrained bandwidth and changing network conditions (packet-loss, high jitter) (Linden 2008 p. 4).

iSAC is used in other chat clients with voice functionality like AIM®[52] , Yahoo®! Messenger[53].

Apart from the H.264 codec and the GIPS audio codec the Google Voice and Video plug-in contains a piece of software that Google internally named browser channel. The next chapter will be introducing the browser channel in Gmail Voice and Video.

8.1.3 The browser channel

The browser channel is an integral part of Google's Gmail. It displays your list of active connections and displays their status. Furthermore does it allow you to engage in different forms of conversation. You can either initiate an instant messaging session, a voice call or a video conferencing session.

The browser channel is programmed in: C++, JavaScript, C and COCOA. Google goes a different way due to the fact that they are not using Adobe Flash, like most other applications on the Internet.

---

[51] Nyquist-Shannon sampling theorem

[52] http://dashboard.aim.com/aim, accessed the 15.07.2009

[53] http://uk.messenger.yahoo.com, accessed the 15.07.2009

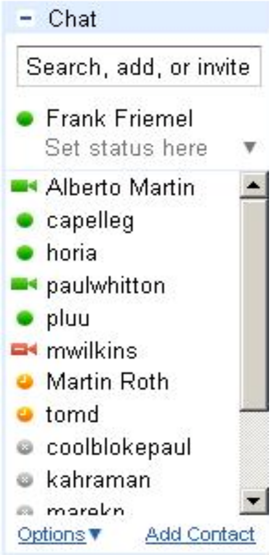


Figure 33: Browser channel

The browser channel is in continuous connection to a XMPP server on the public Internet and exchanges information about the status of you and your connections. A working connection between the user client and the server is called ‘presence’ and in the status message of your contacts you get a dot in a colour other than grey. As soon as the server loses connection to your client it makes you appear offline for all your contacts and your contact appear offline for you.

The browser channel[54] uses the open standard XMPP protocol for authentication, presence, and messaging. Thereby Google makes sure that the messaging client is able to interoperate with other instant messaging clients. Any client that supports XMPP can connect to Google Gmail.

XMPP[55] (Extensible Messaging and Presence Protocol) is the infrastructure behind the instant messaging systems. It is an open technology for real-time communication, using the Extensible Markup Language[56] (XML) (Tanenbaum 2003 p. 639), as the base format for exchanging information.

XMPP provides a way to send small pieces of XML from one entity to another in close to real time. XMPP offers a wide range of services for different kind of applications.

[54] http://code.google.com/apis/talk/open_communications.html

[55] http://xmpp.org/

[56] XML provides a basic syntax that can be used to share information between different kinds of computers, different applications, and different organizations, to structure information and separate content from formatting.

XMPP doesn’t handle the transmission of audio and video data over packet-switched networks directly. XMPP Extension Protocols (XEP) have been developed for that matter. The XEP-0167 (Jingle) describes the exchange of media such as voice and video. Like in H.323 videoconferencing it uses the Real-time Transmission Protocol (RTP) as a container protocol for the media (audio and video).

A Jingle RTP session is described by a content type that contains one application format and one transport method. Each <content/> element defines a single RTP session. A Jingle negotiation may result in the establishment of more than one RTP sessions (e.g., one for audio and one for video).

Table 3: Example of a session-initiate for a video session over Jingle[57]

```
<iq from='romeo@montague.lit/orchard'
    id='ij6s4198'
    to='juliet@capulet.lit/balcony'
    type='set'>
  <jingle xmlns='urn:xmpp:jingle:1'
          action='content-add'
          initiator='romeo@montague.lit/orchard'
          sid='a73sjjvkla37jfea'>
    <content creator='initiator' name='webcam'>
      <description xmlns='urn:xmpp:jingle:apps:rtp:1' media='video'>
        <payload-type id='98' name='theora' clockrate='90000'>
          <parameter name='height' value='600'/>
          <parameter name='width' value='800'/>
          <parameter name='delivery-method' value='inline'/>
          <parameter name='configuration' value='somebase16string'/>
          <parameter name='sampling' value='YCbCr-4:2:2'/>
        </payload-type>
        <payload-type id='28' name='nv' clockrate='90000'/>
        <payload-type id='25' name='CelB' clockrate='90000'/>
        <payload-type id='32' name='MPV' clockrate='90000'/>
        <bandwidth type='AS'>128</bandwidth>
      </description>
```

[57] http://xmpp.org/extensions/xep-0167.html, accessed the 25.07.2009

```
      <transport xmlns='urn:xmpp:jingle:transports:ice-udp:0'/>
    </content>
  </jingle>
</iq>
```

8.1.4 Initiation of a conversation via Gmail Voice and Video

A user selects one of his contacts and clicks on the button voice and video chat. Before a voice and/or video connection is established the XMPP server checks and transmits the IP addresses to the other involved client. The calling client sends a session-initiate to the other clients IP address. In the session-initiate request it is defined what media is wanted (audio, video or both) and the ports for the RTP packets, as well as the supported codec’s. When the second client accepts the call request it sends back a session-accept[58]. As soon as the connectivity is established, the parties begin to exchange the media. Two sessions are setup – One session for the audio packets and one session for the video packets. This has the advantage that during a running session the participants can take away or add new sessions. If the communication started out with audio only, it is possible to add a video session and make it a video call, without interrupting the communication.

After it has successfully established the communication the XMPP server only monitors the connection between the two nodes. Gmail Voice and Video chat tries to establish a direct connection for the media traffic between the two computers calling each other. If this fails, it will fall back on using a proxy server hosted at Google. The plug-in only works correctly if no Firewalls are stopping the traffic and making real-time communication impossible.

Gmail Voice and Video has not yet been implemented for the LINUX platform, as there are no existing drivers to support the external periphery. Furthermore do Linux users only account for 1% of the market targeted by Google[59].

[58] http://xmpp.org/extensions/xep-0167.html, accessed the 20.07.2009

[59] Source: comscore.com, accessed the 15.05.3008

9 Comparison of TANDBERG desktop systems and Gmail Voice and Video

Tandberg product portfolio and the Google Gmail Voice and Video solution are targeting a different audience. Tandberg is aiming to attract professional business user and large international operating companies that need conference solutions to minimize travel expenses and to bring teams together that are scattered over the world. It offers a wide range of video conferencing products for different needs and requirements.

Google Gmail Voice and Video is a product that was developed to offer video conferencing ability together with Google's successful mail client Gmail. Offering a mail client and additional chat functionality as well as voice and video features, Google hopes to attract new users to its growing base of private and corporate client base.

In the following chapter a Tandberg 1700 MXP desk system is compared with the Google Voice and Video plug-in. Both systems are designed for one-person use.

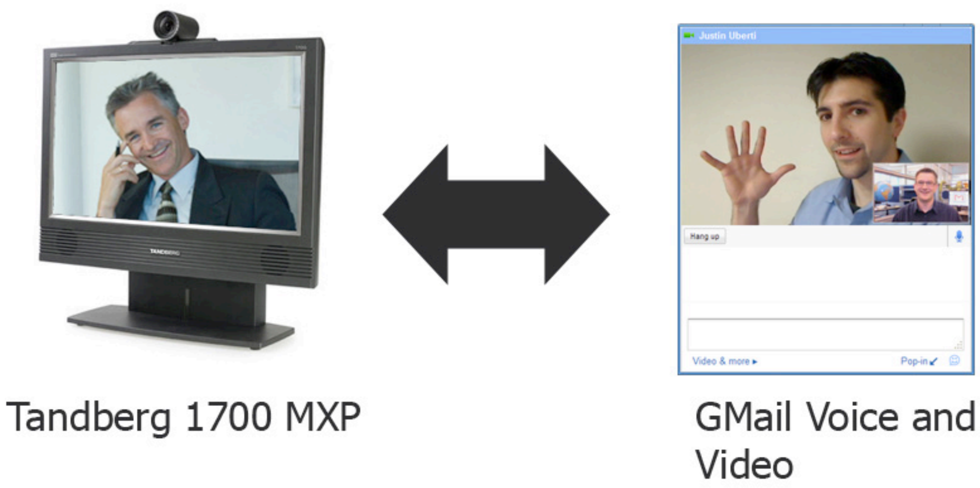


Figure 34: Comparison Tandberg 1700 MXP and Gmail Voice and Video

9.1 Functionality

In terms of functionality both solutions offer a wide range of different features.

Both use the latest video compression standards H.264. With the Tandberg system you have furthermore the possibility to switch to older compression standards like H.261 and H.263. However H.264 will deliver the best results and is most of the times the preferred compression method. The Gmail Voice and Video plug-in automatically uses H.264 for the best image quality at the lowest bandwidth possible.

Although both solutions are intended for video conferencing, they are also capable of doing pure voice calls. Google's solution also offers instant messaging capability.

The approach both systems go in terms of addressing is quite different. This is mainly because Google Gmail Voice and Video plug-in is not H.323 compliant and is designed for a use in the public Internet. Tandberg uses the two schemes introduced in chapter 5 (E.164 addressing and H.323 ID's) to identify each individual system on a LAN based network.

Google Gmail Voice and Video works with user profiles. Meaning before you can contact a specific person by chat or audio and video, this individual needs to have created a user account with Google Gmail and installed the necessary hardware (camera and microphone). Furthermore it is necessary to invite that person to chat[60]. This way both parties agree that they are willing share information. In theory it is possible to invite all Gmail users, which would be more than 91.6 million users (date: April 2008)[61].

The bandwidth both solutions provide differs substantially. Google's product bandwidth depends on network conditions, but is generally quite low. Google has tested bandwidth with the plug-in as low as 128 kbps[62]. The user cannot choose a connection speed. The bandwidth is determined from a negotiation between the two plug-ins involved, taking into account the total bandwidth available as well as jitter and latency. This is because it uses the Internet to transmit real-time information. In the Internet no Quality of Service is in place and bandwidth is shared between different users.

The bandwidth of Tandberg's 1700 MXP video conferencing codec can be modified. The maximum bandwidth is 1920 kbps (Tandberg 2008). Most of the companies use speeds between 384 kbps and 768 kbps. The minimum bandwidth is 64 Kbps, which is virtually not used for video conferencing.

Additionally to the transmission of video and audio Tandberg supports the transmission of a second video channel to share presentations or the image of a second camera. This way you are not only able to see your colleagues, but also to present documents and slides to them over a second channel. This is described under the ITU-T recommendation for H.239 (ITU-T 2005).

---

[60] http://mail.google.com/support/bin/answer.py?hl=en&answer=33508 , acessed the 04.06.2009

[61] http://www.email-marketing-reports.com/metrics/email-statistics.htm, acessed the 04.06.2009

[62] http://groups.google.com/group/Gmail-Help-Chats-and-Contacts-en/browse_thread/thread/e34c6524cda99b2/13499e3de3f3707c?lnk=gst&q=, accessed the 15.07.2009

Another feature that Tandberg 1700 offers is MultiSite® capability. The codec has an built-in MCU and can bridge call with up to four video participants and three audio participants (Tandberg 2008). Gmail Voice and Video can only connect two sites at the same time.

9.2 Security

In terms of security offers Tandberg the option to encrypt calls with modern encryption standards like the Advanced Encryption Standard (AES). AES-128 is considered to be very secure. It uses a 128-bit key. AES is a symmetric encryption scheme which uses a fixed-length key to encrypt and decrypt the data (Firestone, Ramalingam et al. 2007 p. 299). Gmail Voice and Video audio and video packets remain unencrypted while transmission.

9.3 Handling

Each Tandberg endpoint comes with a remote control with an intuitive layout to control calls (cf. Figure 35). The remote includes buttons to increase and decrease the volume, muting functions and camera operations (tilting and zooming in and out). Furthermore a keypad is included which can be used to type in numbers for E.164 numbers or letters, as used in the H.323 ID.

The codec is equipped with a 20" widescreen LCD display (Tandberg 2008). One person or a small group of people can comfortably have a video conference with the Tandberg codec.

The Gmail Voice and Video plug-in is designed for one-person use only as you need to sit directly in front of the computer. Everything is controlled with the peripherals of the computer.

9.4 Price

Gmail Voice and Video is free. You only need to provide the necessary hardware and peripherals (mouse, keyboard, camera and microphone) as well as the latest software (operating system). A broadband connection to the Internet is necessary for sufficient bandwidth.

At the moment a Tandberg 1700 MXP codec as basic model would cost around £6,090[63]. In order to use Tandberg products successfully you would need at least two video conferencing units, which would incur further costs.

[63] BT Conferencing internal pricing for Tandberg for Q1/2009.

9.5 Connectivity

In an IP based network TANDBERG endpoints can talk to endpoints from different manufacturers, provided that they are H.323 compliant. The H.323 standard is designed for video conferencing over Local Area Networks (LAN). The Tandberg 1700 MXP does not provide sockets for ISDN BRI's and offers only an Ethernet port.

If you have the necessary soft- and hardware provided it is theoretically possible to initiate a video chat with Gmail Voice and Video from everywhere in the world with a broadband Internet connection.

10 Development of a kiosk for the Gmail Voice and Video conferencing solution

The initial request was to develop a conference platform for the telephone booths in the Google London office, to enable Google employees to make video conferencing calls in a private (noise insulated) environment. It was mandatory to use Google's own Gmail Voice and Video plug-in. The plug-in should be installed on an off-the-shelf generic computer platform. At the time of the development of my project, Google Gmail Voice and Video plug-in was only available for Windows. Therefore I was restricted in using a PC to implements the solution running Windows XP.

Figure 35: Phone booth in the Google London office

The primary objects were:

Security: It was a very important aspect of the whole enterprise to keep security tight around the Google intranet. The Google office is frequented by visitors on a daily basis.

In all circumstances access to the Google intranet from the phone booths was to be prevented, as this would have been considered a major security breach.

Accessibility: The phone booths should be accessible for all Google employees 24/7/365 without pre-booking.

Usability: The conference should be easy initiated with a few clicks and be intuitive in handling.

Cost: The video conferencing experience needed to be accomplished with the minimum possible cost in equipment and maintenance.

Scalability: The solution for a single user system should be scalable to larger systems for larger audiences without increasing the cost extensively and keeping the basic user interface.

With those guidelines Mr. James Bond, Head of Operations at Google in EMEA, assigned me the project.

10.1 Basic concept

My idea was to take a computer, install only the necessary software and driver and restrict access only to the software needed for video conferencing. Every time a new user comes in a phone booth to use the system he is prompted for the login credentials of his Gmail account. After logging in, he can access his email or doing a video conference with the Gmail Voice and Video plug-in.

The computer is locked down. The user can only access his or her Gmail account. Any access to other programs or computer settings is not possible. This way it is ensured that the system remains all times in a workable condition.

Additionally a kiosk is created by applying extensive security mechanism with the Microsoft program Microsoft Windows SteadyState®[64]. The product provides security for multi-user systems to prevent errors due to misconfiguration of the systems.

[64] http://www.microsoft.com/Windows/products/winfamily/sharedaccess/, accessed the 16.07.2009

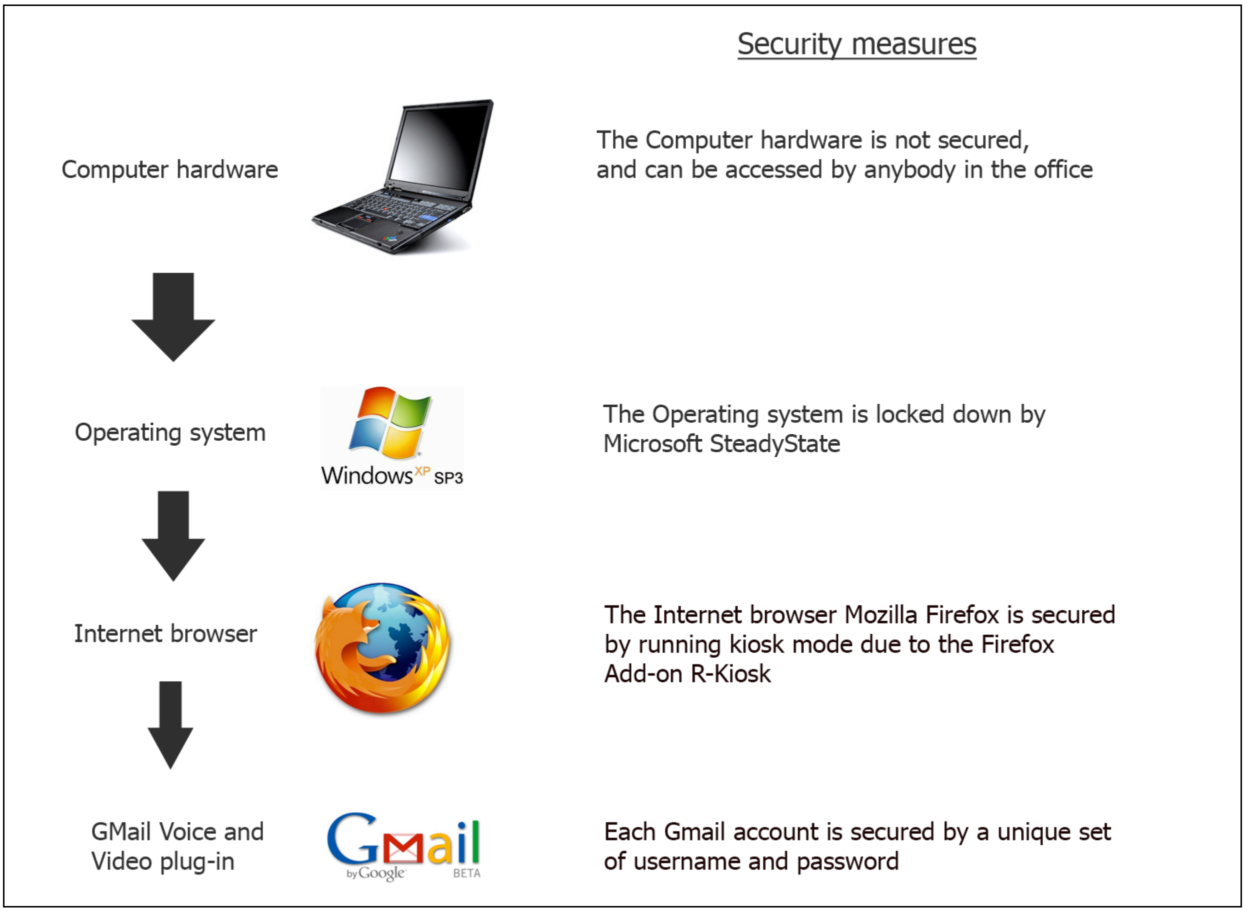


Figure 36: Security measures for the video conferencing solution

10.2 Hardware

In order to fit the computer in the confined space of a phone booth, the choice of computer hardware was a laptop. Google offered its employees laptops with Windows XP preinstalled. To keep the costs down I use an IBM Thinkpad T42, which has been deprecated. The machine was preconfigured with Windows XP and as browser Mozilla Firefox 2.0 was installed.

Figure 37: IBM Thinkpad T42[65]

10.3 Creation of a local user account

A Windows user's profile allows the user to have a personalized desktop environment. The desktop environment includes the content and arrangement of Start Menu groups, screen colours, desktop shortcuts, network and printer connections, and peripherals settings. When a new user logs onto a computer, a profile is created automatically. While a user is logged on, changes made to the desktop environment are saved to their user profile.

For the phone booth kiosk project only one profile is necessary. All users are using this one profile in order to log on to their Gmail account.

Creating a new Windows XP User Account

1. Click Start, click Run and type Control Panel and then click OK

[65] http://www.thelaptopcentre.co.uk/user/products/ibm-thinkpad-t42.jpg, accessed the 25.07.2009

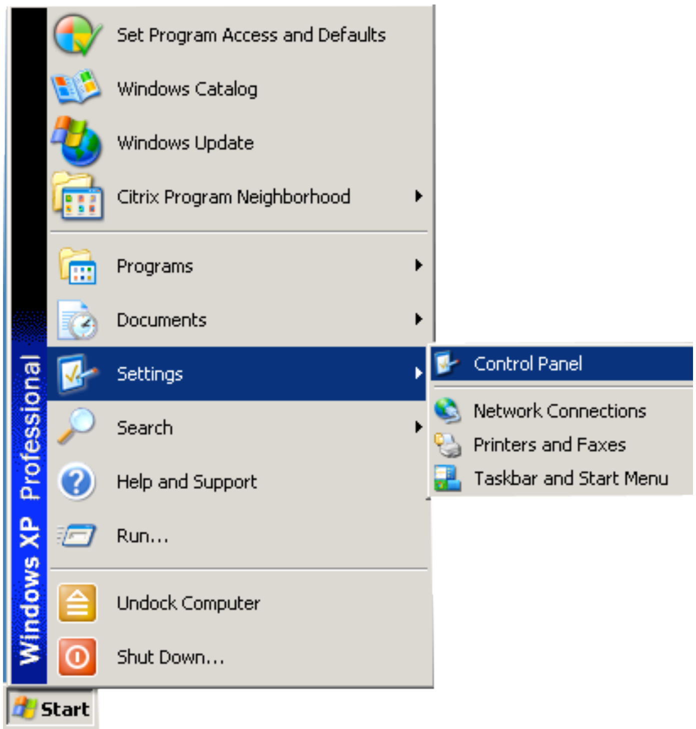


Figure 38: Launch of the Control Panel

2. The Control Panel Window appears. Click on the User Account icon.

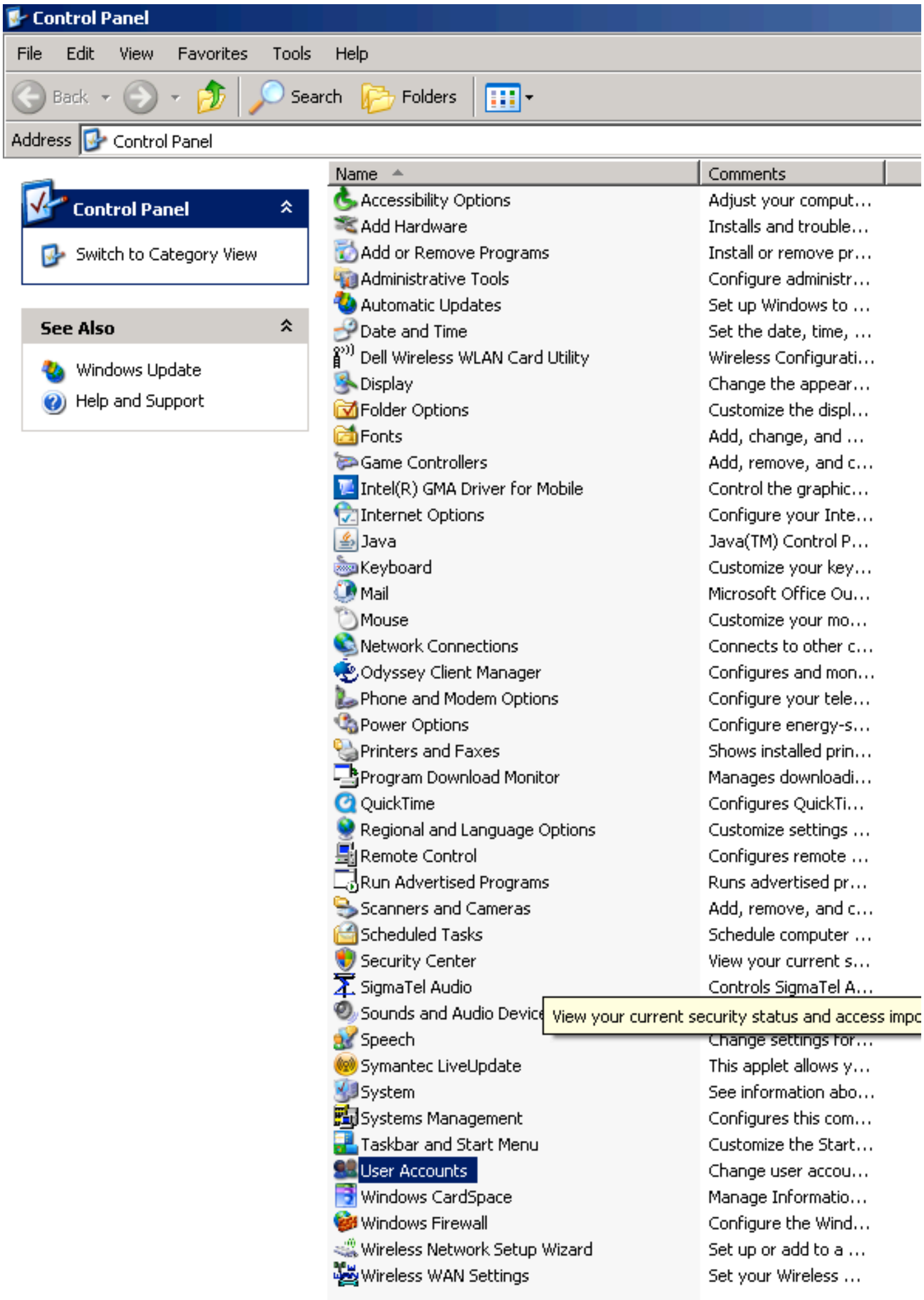


Figure 39: Control Panel in Windows XP

3. In the User Account Properties Window click on “Create a new account.” A textbox comes up where you have to enter the accounts new name.

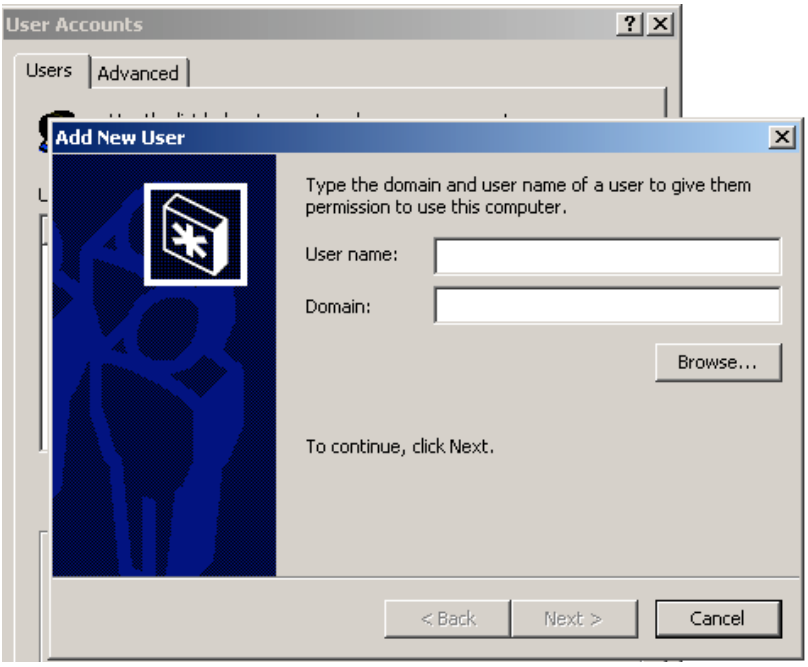


Figure 40: „Add new User“ textbox in the user accounts dialog

9. In the next window you can specify what type of access rights the new profile gets. You can choose between full access (“Computer administrator”) or restricted access rights (“Limited”). The new local account will be created immediately and you can use that account the next time you logon to a new Windows session. Then click on “Create account.”

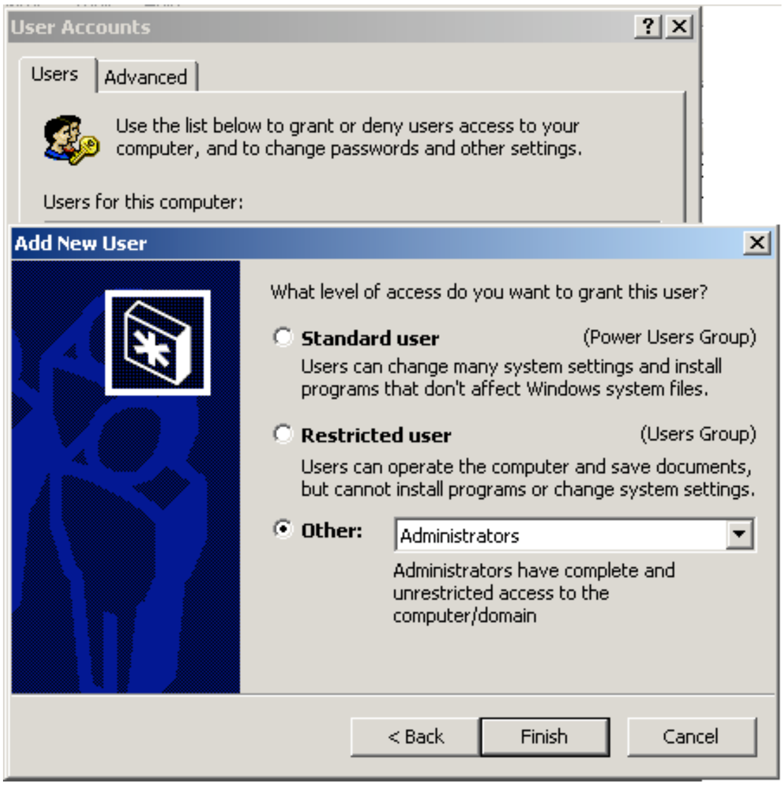


Figure 41: Access rights dialog for the new user

In our case it doesn't matter which account type we choose. Both types don't offer enough security measurements for our purpose and therefore need to be reinforced by additional protection measures.

Each time the computer is rebooted Windows prompts the user for his or her username and password. In our set-up we want any user to have open access to the system. A way to accomplish that is to enable auto login on the client.

10.4 Enabling auto login

Microsoft offers the possibility to enable auto login[66] for its operating system Windows XP.

The username and password are then stored in the Registry[67] database of Windows.

The Registry is a database containing all the settings for Windows XP, as well as the applications installed on your system (Karp 2003 p. 234). All your file types are stored in the Registry, as well as all network, hardware, and software settings for Windows XP

[66] http://support.microsoft.com/kb/315231

[67] The Windows Registry is a database which stores settings and options for Microsoft Windows operating systems. It contains information and settings for hardware, operating system software, most non-operating system software

and all the particular configuration options for most of the software installed on the system.

Many advanced settings in Windows XP can only be changed by manipulating data in the Registry.

After enabling auto logon the machine it logs itself into Windows after booting. This way the user is not prompted for any username or password.

With auto logon enabled, anyone with physical access to the computer can use it and in theory modify content. Therefore it is of utmost importance that other elaborated mechanisms are in place to secure the content of the configured device.

Security measures like kiosk mode and the browser kiosk prevent the user from doing malicious or non-intended configuration changes on the computer.

Auto Login can be easily achieved by altering the registry entries:

ENABELING AUTO LOGON:

1. Click Start, click Run, type regedit and then click OK

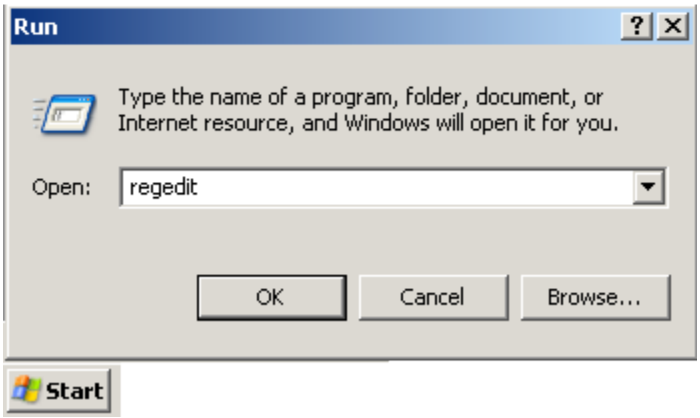


Figure 42: Dialog to open the Windows registry

2. Locate the following registry key: HKEY_LOCAL_MACHINE\SOFTWARE\MICROSOFT\WINDOWSNT\Current Version\Winlogon

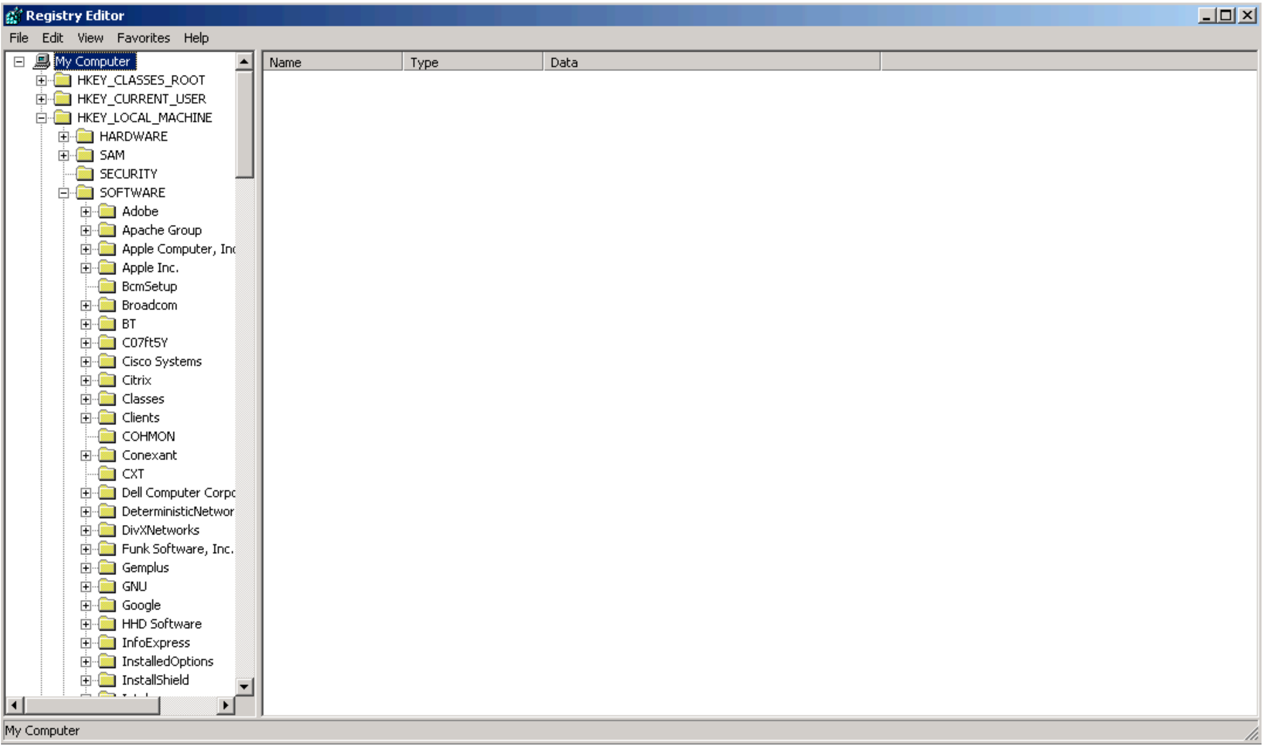


Figure 43: Windows registry

3. Using the account name and password, double click the DefaultUserName entry and then type your user name click OK.

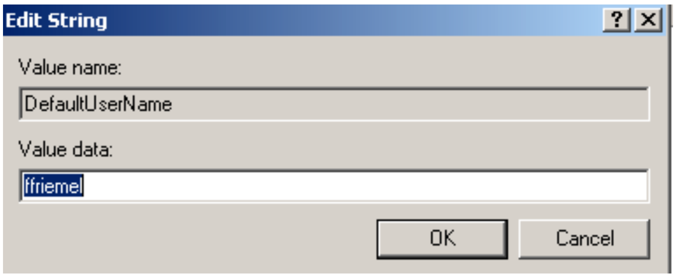


Figure 44: Setting of the default user name

4. Double-click the DefaultPassword entry, type your password under the value data box and then click OK.

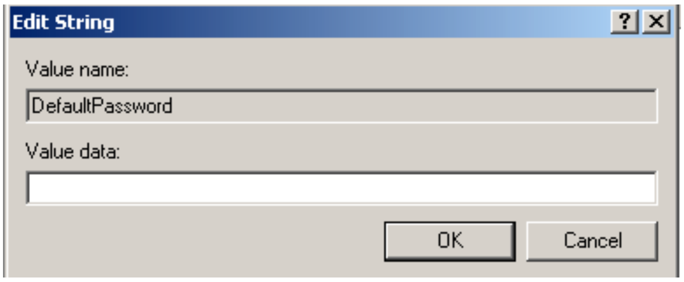


Figure 45: Setting of the default password.

If there is no password value specified, you have to create a new value:

a) In the Registry editor, click Edit, click New and then click String Value.

b) Type DefaultPassword as the value name and then press ENTER.

c) Double-click the newly created key and then type your password in the Value Data

5. Double-click the AutoAdminLogon entry and type 1 in the Value Data box and then click OK

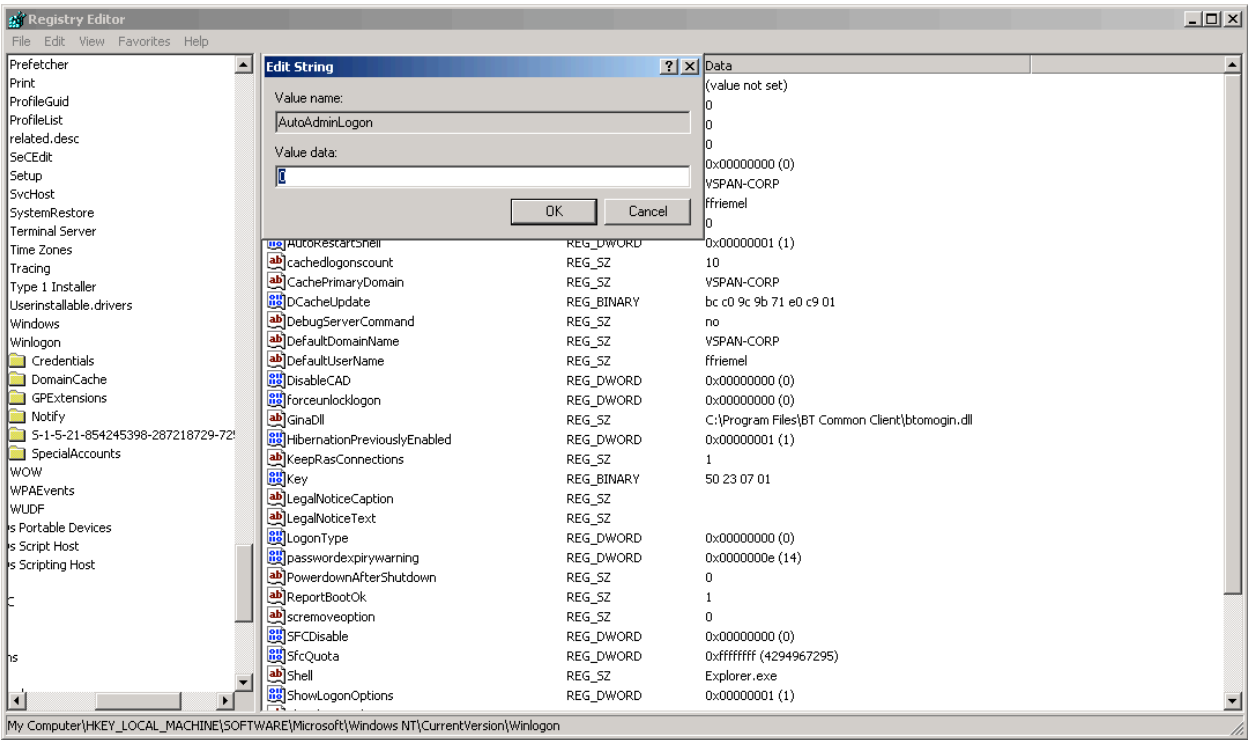


Figure 46: Setting of the AutoAdminLogon

6. Quit Registry Editor
7. Click Start, click Restart and then click OK.

After a restart Windows XP logs itself automatically on to the newly created user account.

10.5 Establishment of an interactive kiosk

An interactive kiosk is a computer terminal that provides information access via electronic methods. A kiosk serves usually one purpose and has a user interface specifically designed for doing one specific task. Kiosks provide secure environments and enable the clients to use do self services like Internet use or paying bills.

Figure 47: Ludicus QuickPhoto 55 is a compact kiosk for ordering photo prints[68]

Kiosks are usually normal operating system environments were only particular interactions with the customer interface are allowed. For the reason of system stability and security certain parts of the operating system are disabled.

In most cases Internet kiosks allow only one browser session. Access to the control panel is disabled. With kiosks it is possible to restrict access to the applications installed and also hide components like hard drives and periphery from the user.

[68] http://www.alibaba.com/product/ludicussp-10947481-10691899/QuickPhoto_55_Digital_Photo_Kiosk.html, accessed the 16.07.2009

A shared access computing environment usually offers consistency to their users, meaning each time they log on to a new session they find the desktop configuration the same as the last time.

10.5.1 Microsoft Windows SteadyState®

For my project I decided myself to use Windows SteadyState®[69], a free kiosk software available from Microsoft for their operating systems Windows XP and Vista. Windows SteadyState® works flawlessly with Microsoft Windows and is easy to configure. Additionally it offers the functionality you need to completely secure your system in a shared computing environment like libraries or schools.

Figure 48: Microsoft SteadyState®[70]

Restrictions applied with Microsoft SteadyState®[71]

- No access to any drives on the system

---

[69] Windows Steady State registered Trademark of Microsoft Cooperation.

[70] http://www.Windowsreference.com/free-utilities/Windows-steady-state-powerful-tool-for-shared-computer-protection/, accessed the 16.07.2009

[71] http://www.microsoft.com/Windows/products/winfamily/sharedaccess/whatis/userrestrictions.mspx, accessed the 16.07.2009

- No access to any other program other than Mozilla Firefox
- No access to Registry, Control Panel, Command Prompt or Explorer or any other configuration environments
- A Session timer logs the current user off after six minutes of inactivity
- No rights to save data on the hard drive.

Microsoft SteadyState® takes care of the necessary security in place in order to prevent damage to the system and comply with Google security policy.

10.6 Browser kiosk

We have achieved so far that a user is automatically logged on to Windows. Then it is possible to launch Mozilla Firefox and surf the web or in our case Google's corporate intranet. Google's security policy forbids access to the intranet from open-access systems.

Therefore we need to apply additional security measures

1. The home page of Firefox should be set to the Google Gmail logon page.
2. The Navigation Toolbar in Firefox must be hidden in order to prevent the user from surfing other web content.

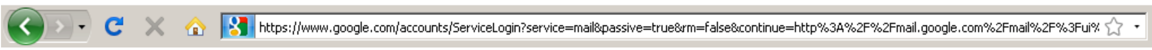


Figure 49: Firefox Navigation Toolbar

3. Right click menu in Firefox should be deactivated to prevent configuration changes of the browser.
4. Hiding the Microsoft Close button

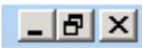


Figure 50: Microsoft Windows close button X

Apart from kiosk modes for operating systems there exist also kiosks for browsers. If your kiosk serves for a simple web interface application it is sometimes enough to secure the browser and leave the operating system as it is. The user will see a full screen browser version with no option to change applications. Browser kiosks are useful if only one website is of interest and the users are not using the computer for anything else than accessing this particular website.

10.6.1 R-Kiosk

For Firefox several kiosk add-ons are available. R-kiosk is a free Firefox add-on and proved after testing several different kiosks to be the most reliable solution.

Real Kiosk is a Firefox 2.0 - 3.0 extension that defaults to full screen, disables all menus, toolbars, key commands and right button menus.

R-kiosk fits perfectly in the concept as it is designed to give access to one web page only. It opened the login in one full screen window with no possibility to get access to other pages on the Google network.

Unfortunately you can close the browser kiosk by pressing Alt + F4 and so returning back to the Windows interface.

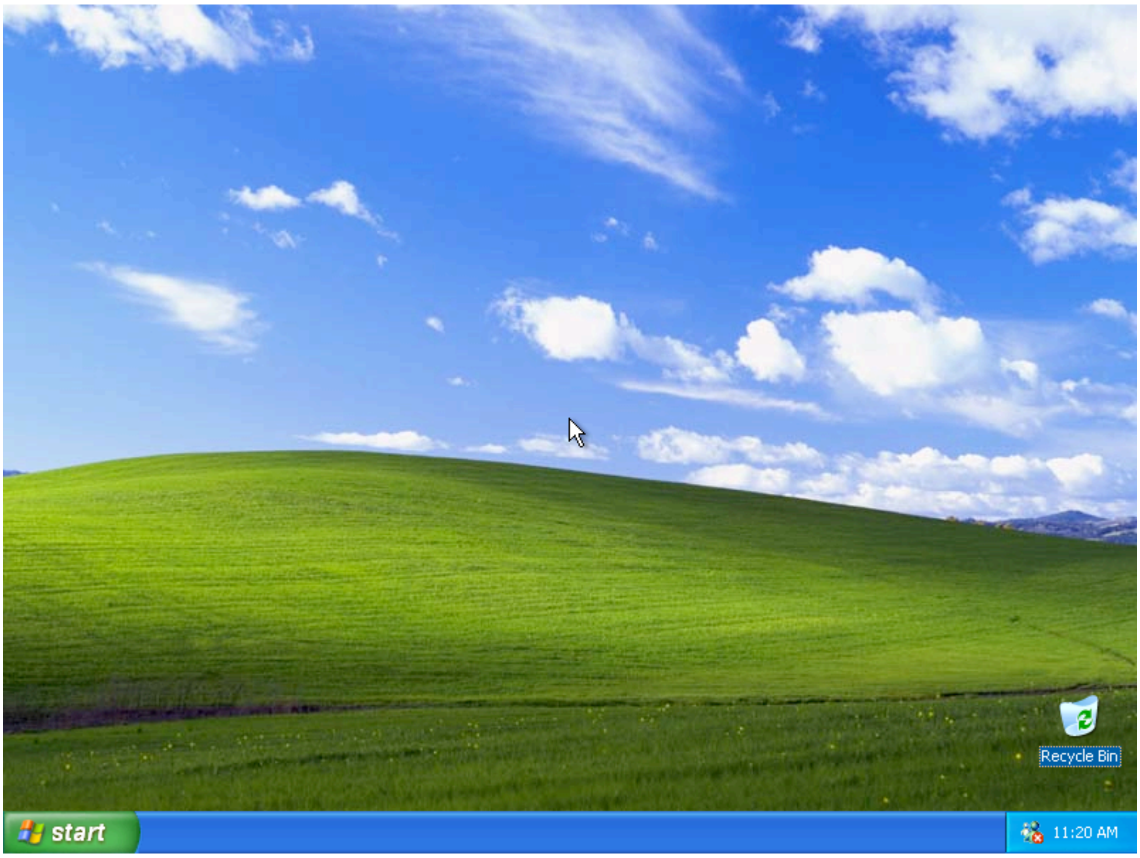


Figure 51: Windows shell after installation of Windows XP[72]

In a non-modified system the user then would be able to manipulate the configuration of the computer.

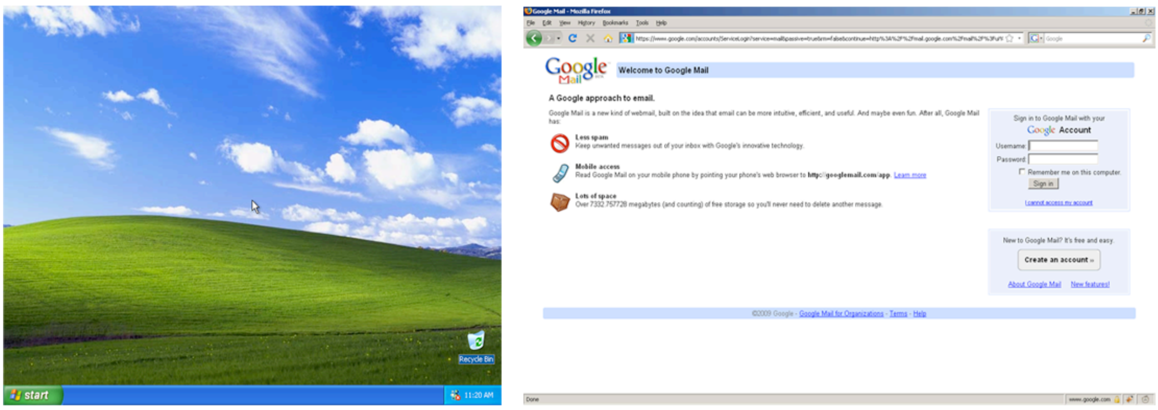
To prevent that two measures are applied:

[72] http://www.guidebookgallery.org/pics/gui/desktop/empty/winxppro.png, accessed the 17.07.2009

1. the OS[73] starts in kiosk mode and a modification of the initial configuration is not possible.
2. Instead of loading the Windows shell Firefox is loaded.

10.7 Changing the Windows shell

In computing the Windows Shell is the most visible aspect of the Microsoft Windows line of operating systems. The shell is the container inside of which the entire user interface is presented, including the Task bar, the Desktop, Windows Explorer, as well as many of the dialog boxes and interface controls. Commonly the Windows shell is referred to as the graphical user interface (GUI) of Windows. Figure 51 shows the shell of Microsoft Windows XP.



Windows shell: explorer.exe   Windows shell: Firefox.exe

Figure 52: Traditional Windows shell and replacement shell Firefox.exe

The Windows shell is loaded when the operating system is started and is for the user the starting basis of all operations within Windows.

For our purpose it is useful to take the Windows shell away from the user and so to prevent him accessing anything other than Firefox and the website provided.

Instead of loading explorer.exe, which is the standard Windows shell, Firefox.exe is loaded.

In chapter 10.4 Auto Logon was enabled to make Windows automatically logon.

In order to do that the Windows registry of Windows has to be modified.

1. Open the Windows registry, click Start, click Run, type regedit and then click OK.

[73] OS: short for operating system

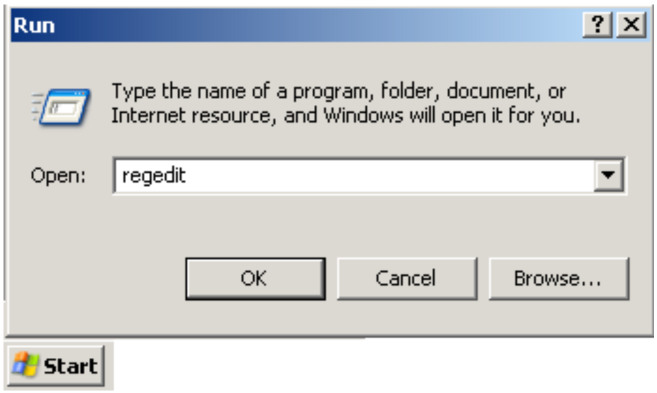


Figure 53: Dialog to start the Windows registry

2. Locate the following registry key: HKEY_CURRENT_USER\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Winlogon

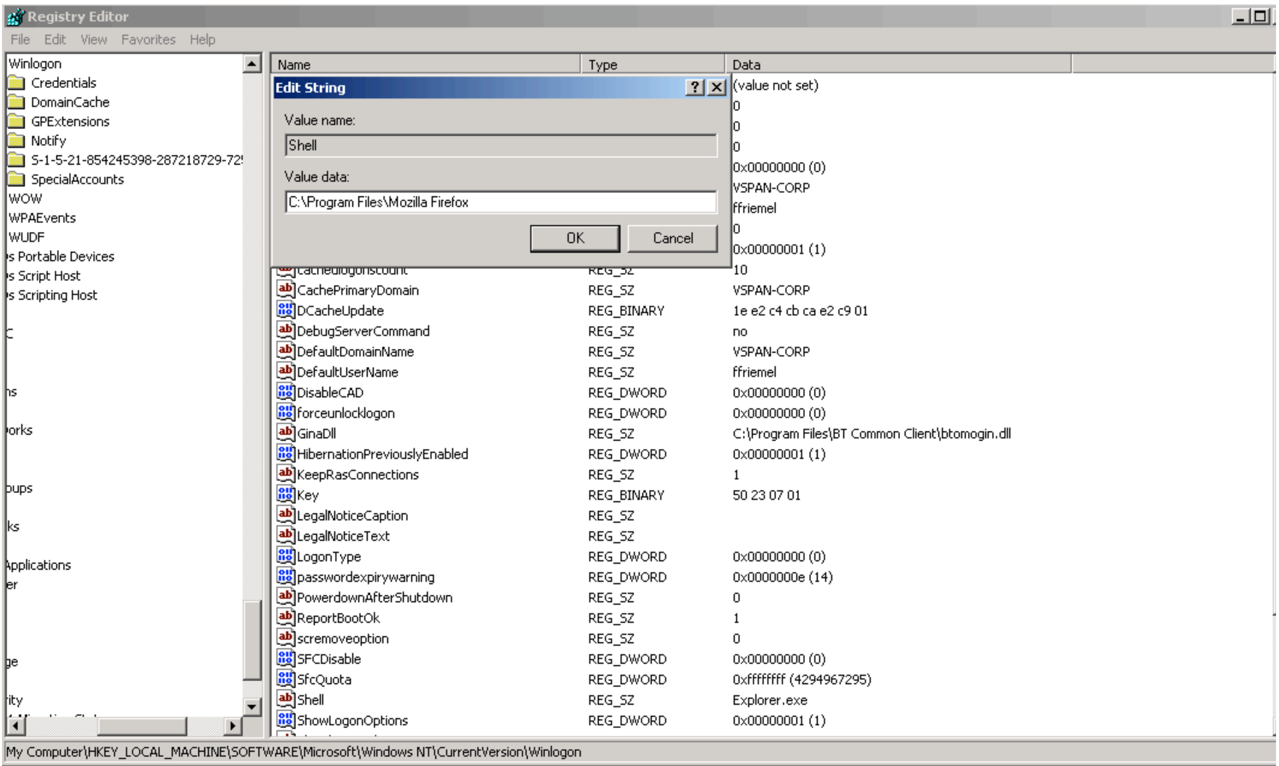


Figure 54: Windows registry entry to change the Windows shell

3. Add a new string value: Click Edit, click New, click String Value. The string value should be called shell and the value needs to be the path of the new shell (e.g C:\Program Files\Mozilla\Firefox.exe)

4. Log out of Windows and log back in.

Each time Windows is booted Firefox automatically starts.

10.8 Changing the homepage of Firefox to the Gmail logon page

The last step to accomplish is to make the logon page for the Google Gmail the homepage of Firefox.

1. Open Firefox.

2. Open the Menu bar Tools and then click on Options.

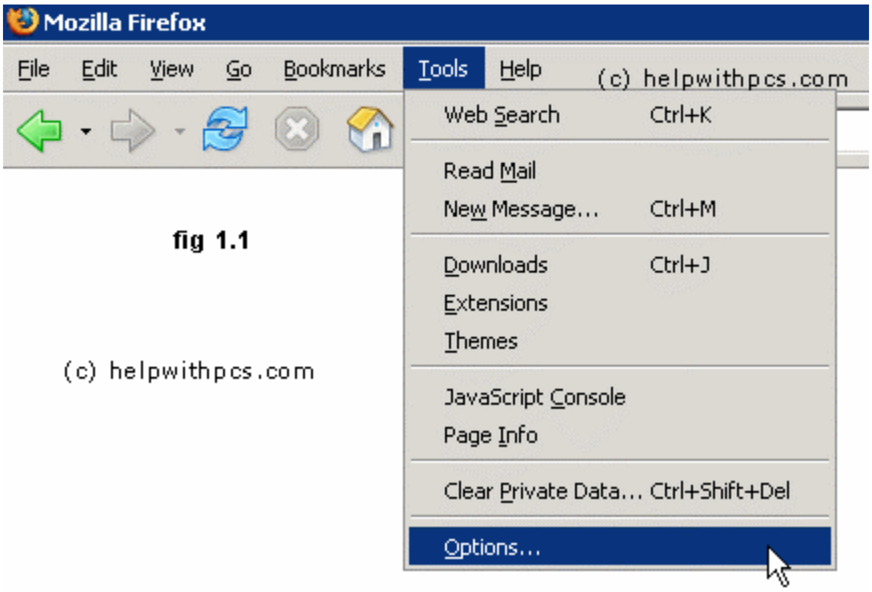


Figure 55: Menu bar in Mozilla Firefox

3. Click on the "Main" tab.

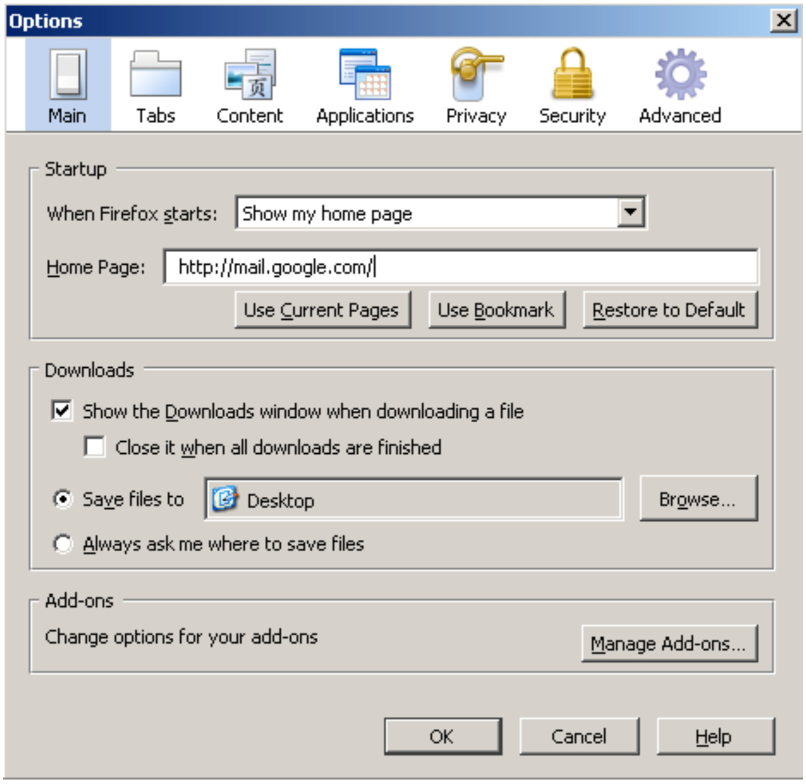


Figure 56: Settings in Mozilla Firefox „Main“ tab

Therefore when Firefox is open, we click on Firefox in the taskbar and then click on Preferences and set on the main tab the home page to http://mail.google.com

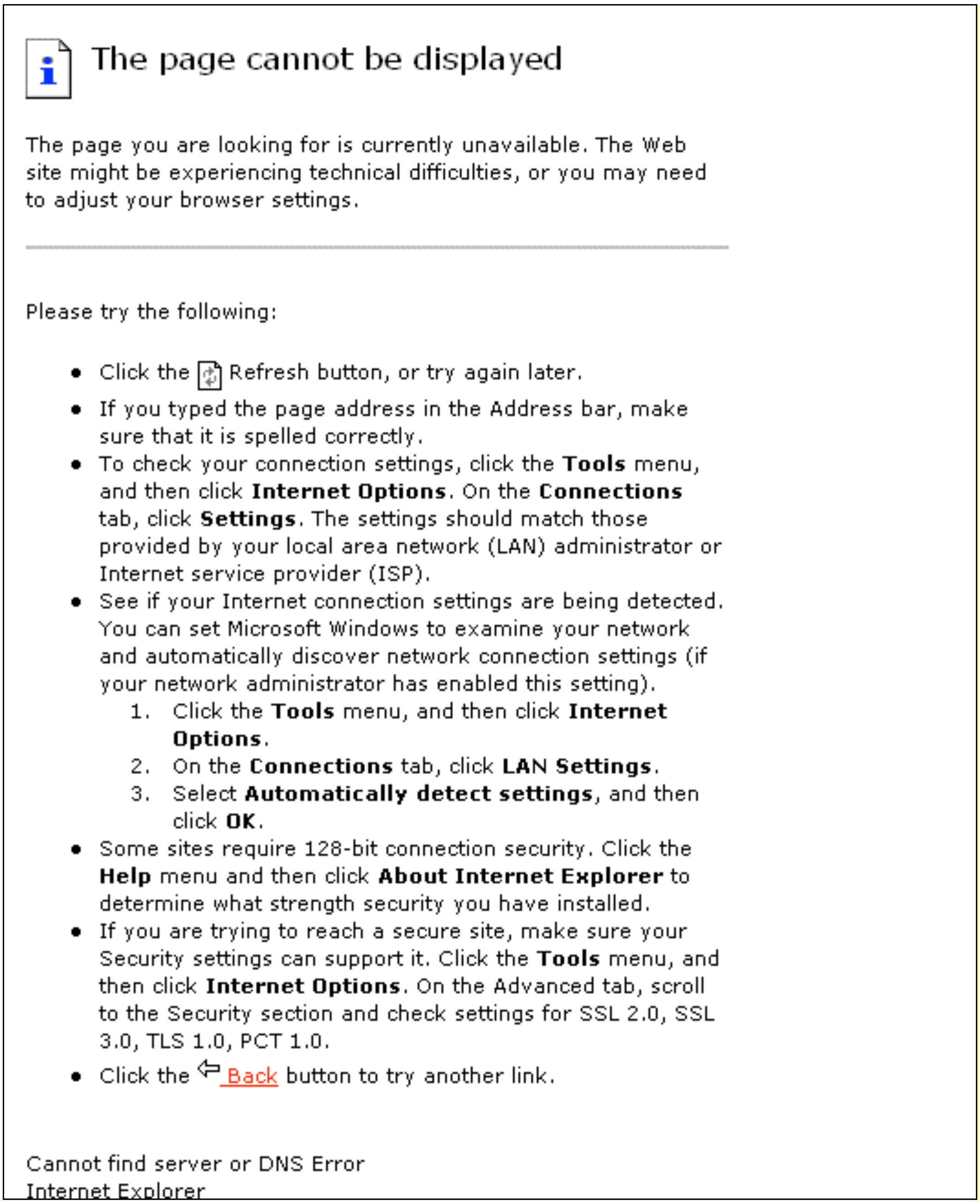


Figure 57: HTTP error code 404 – File not found

When the computer is booted the user is automatically logged in Windows XP and Firefox opens. Unfortunately the Gmail Logon page does not come up , but an error message is displayed (cf. Figure 57).

The error is due to the fact that at the point of time when Firefox is started, the network drivers aren’t loaded yet and therefore the connection to Internet is not successfully.

In a further step a script needs to make sure the Internet connection is established, before it tries to connect to the Gmail page http://mail.google.com.

10.9 Network connection check

10.9.1 Localhost

An indication of all network drivers being successfully loaded is the issue of the localhost address to the machine. In a network environment localhost means "This computer" and is the standard hostname given to the address of the loopback network interface. The localhost always translates to the IP address 127.0.0.1.

If a web server is running on a computer, than the input of http://localhost in the browser of the computer would display the homepage of the local website.

The following script checks for the existence of the local loopback address, which confirms that the network drivers are loaded and a connection to the Internet is established. The script is written in HTML with some JavaScript to check for the network connection.

10.9.2 HTML

HTML is an acronym for Hypertext Mark-up Language. It is a publishing language for the world-wide web. HTML was designed for Web browsers to present text and graphics.[74] The language is an international standard (ISO/IEC 15445:2000), but is officially maintained by the World Wide Web Consortium (W3C)[75].

Elements are the basic structure for HTML mark-up. An element has the form <elementname>. Typical elements are <fonts>, <head> or <p>.

Elements have two basic properties: attributes and content. Each attribute and each element's content has certain restrictions that must be followed for an HTML document to be considered valid. An element usually has a start tag (e.g. `<body>`) and an end tag (e.g. `</body>`). The element's attributes are contained in the start tag and content is located between the tags (e.g. <body="document">Start page</body>).

The following script is written following the HTML 4.01 standard[76].

The different parts of the scripts are colour-coded:

All HTML code is coloured in red. All JavaScript is coded in green and the CSS is colour-coded in blue.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/loose.dtd">
```

[74] http://www.w3.org/MarkUp/Guide/ accessed the 13.06.2009

[75] www.w3.org, accessed the 13.06.2009

[76] http://www.w3.org/TR/html401/, accessed the 13.06.2009

```
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<title>Gmail Voice and Video</title>



function checkForConnection() {
  if (hasNetworkConnection() == true) {
    window.location = "http://mail.google.com";
  } else {
    setTimeout(checkForConnection, 5000);
  }
}

function hasNetworkConnection() {
  var ipStr = getIpAddress();
  if (ipStr == "127.0.0.1") {
    return false;
  } else {
    return true;
  }
}

function getIpAddress() {
 var ip = java.net.InetAddress.getLocalHost();
 var ipStr = new java.lang.String(ip.getHostAddress());
 return ipStr;
}
</script>
<style type="text/css">
<!--
p {
   font-family: Georgia, "Times New Roman", Times, serif;
   font-size: 60px;
```

```
    text-decoration: underline;
    border: 1px solid #000000;
    position: absolute;
    height: auto;
    width: 897px;
    left: 127px;
    top: 253px;
    right: auto;
    bottom: auto;
}
-->
</style>
</head>
<body onLoad="checkForConnection();">
<p>Please wait while system boots up</p>
</body>
</html>
```

Code explained:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<script type="text/javascript">
```

We are using the HTML version 4.01.[77]

```
<html xmlns="http://www.w3.org/1999/xhtml">
```

The third line officially starts the HTML document with opening the tag <html>. HTML documents can be displayed by web browsers and offer a way to display static and dynamic web content[78].

```
<head>
```

The fourth line opens the head tag. The head tag in a HTML document includes information about the title of a document, which is then displayed in the head of the browser.

[77] http://www.w3.org/TR/html401/struct/global.html

The head tag also includes meta-information about the document. This is not intended for the reader of the actual web page but gives the displaying browser additional information on how to handle the page.

```
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
```

In line five the meta tag explains the browser that the whole content of the HTML document consists only of plain HTML text. A HTML page can include plug-ins like Adobe Flash and Java, embedded movies and sound clips. The meta tag will inform the browser which type of content the web page entails.

```
<title>Gmail Voice and Video</title>
```

The Title of the document in the title tag is the only information actually visible in the head section of an HTML document. All the other information is meta-information explaining the HTML document in more detail.

```
<script type="text/javascript">
```

In the eight line, a piece of JavaScript code is included in the HTML document.

```
</head>
```

After some JavaScript and CSS information, which will be explained in the following chapters, the head tag is closed. All meta-information like style and scripting have been included. The body tag on the other hand deals with the content that is actually displayed on the webpage, like text and images.

```
<body onLoad="checkForConnection();">

<p>Please wait while system boots up</p>

</body>
```

The body tag in this script comprises of only three lines. First the body open tag. In the open tag is a JavaScript command which tells the browser to execute the function checkForConnection() when the HTML body is loaded.

The second line is a <p> tag, which defines a paragraph. The text inside a paragraph will be displayed on a web page. Extra space is created around the text. If a document consists of several paragraphs the extra space helps to visually distinguish one from another.

After the normal HTML procedure the open body tag has to be closed again which is done on line 50.

```
</html>
```

[78] http://www.w3.org/TR/html401/, accessed the 13.06.2009

In the line 51, the last line, the html closing tag is written down, which tells the browser that the html document is finished for that webpage and that he cannot expect HTML code below the html closing tag.

The HTML file is saved in a local folder on the computer.

10.9.3 JavaScript

JavaScript is a client-side scripting language mainly used for web development (Goodman and Morrison 2004 p. 3). Client-side means that the JavaScript code is executed in the browser of the client machine and not on a web server. It is mainly used to build interactions between page content, the state of the browser, and the actions of the reader. For example in a webpage, you use JavaScript to create animated textboxes or to validate user input on forms.

Figure 58: JavaScript mouse rollover effect

The following code is the JavaScript part of the HTML document. Three functions are defined. A function is a reusable code-block that will be executed by an event, or when the function is called. Instead of writing down a specific piece of code over and over again you can put it in a function. Every time in a script you need the function to be executed, you call the function.

The first function is the function getIpAddress(). This function defines two variables. The first variable **ip** is defined by a Javaclass[79] method which returns the IP of the local host in the form localhost/127.0.0.1[80]. If no IP address could be found it throws back UnknownHostException.

---

[79] Although JavaScript does not provide direct support for obtaining the user's host name and IP address, it does give you access to the standard Java classes. And within those classes is one called "java.net.InetAddress," which contains methods to get the user's host name and address.

[80] 127.0.0.1 is the standard IP address used for a loopback network connection.

The second variable **ipstr** takes the result of the variable IP and forms it into a raw IP address in a string format and returns the IP address in a textual presentation. Basically it converts localhost/127.0.0.1 into 127.0.0.1, another string format.

```
function getIpAddress() {
 var ip = java.net.InetAddress.getLocalHost();
 var ipStr = java.lang.String(ip.getHostAddress());
 return ipStr;
}
</script>
```

The second function hasNetworkConnection() simply checks if the variable ipStr which is returned from the function getIpaddress() equals the value “127.0.0.1” of the variable. If the variable ipStr contains the string 127.0.0.1 the function returns True, else it returns. False.

```
function hasNetworkConnection() {
  var ipStr = getIpAddress();
  if (ipStr == "127.0.0.1") {
    return false;
  } else {
    return true;
  }
}
```

The last function checkForConnection() takes the result of the function hasNetworkConnection(). If hasNetworkConnection() returned True the user is redirected to another page called mail http://mail.google.com. If hasNetworkConnection() returns False after a timeout of 5000 ms the function getIpAddress() is called again and the procedure begins once again.

```
function checkForConnection() {
  if (hasNetworkConnection() == true) {
    window.location = "http://mail.google.com";
  } else {
    setTimeout(checkForConnection, 5000);
  }
}
```

10.9.4 CSS

Cascading Style Sheets is a style sheet language used to describe the presentation, that is, the look and formatting, of a document written in a mark-up language. Its most common application is to style web pages written in HTML.

CSS is primarily designed to separate the content from the formatting of the document. HTML is only responsible for the content of the web document, whereas CSS is designed to build the layout of web content. CSS provides flexibility to web content. The advantages of Style Sheet formatting are (Niderst 1999 p. 389) :

- Greater layout control: style sheets can specify traditional layout elements like font size, line spacing and letter spacing and offers additionally methods to alter indents, margins and element positioning.
- Separation of the structure of a document from its layout.
- Decreasing file size: In order to apply layout information with HTML, layout tags have to be applied each time. CSS styles can be applied globally for the whole document and so keep document sizes small.
- Simplified site maintenance: One CSS document can link to several HTML document. The layout of several web pages can be maintained with one CSS document.

In the script that verifies the IP address, CSS provides mainly the formatting and positioning of the text. In this case the font family, font size and position are described with CSS.

The CSS specifications are maintained by the World Wide Web Consortium (W3C)[81].

```
<style type="text/css">
<!--
p {
  font-family: Georgia, "Times New Roman", Times, serif;
  font-size: 60px;
  text-decoration: underline;
  border: 1px solid #000000;
  position: absolute;
  height: auto;
```

[81] http://www.w3.org/TR/CSS21/, accessed the 13.06.2009

```
  width: 897px;
  left: 127px;
  top: 253px;
  right: auto;
  bottom: auto;
}
-->
</style>
```

The style tag is used to define style information for an HTML document. A style defines how the document is rendered in the browser.

The last step to accomplish is to change the start page of Firefox to the script locally stored on the hard drive of the machine.

10.10 Changing the homepage of Firefox to the script locally stored on the hard drive of the kiosk computer

1. Open Firefox
2. Open the Menu bar Tools and then click on Options

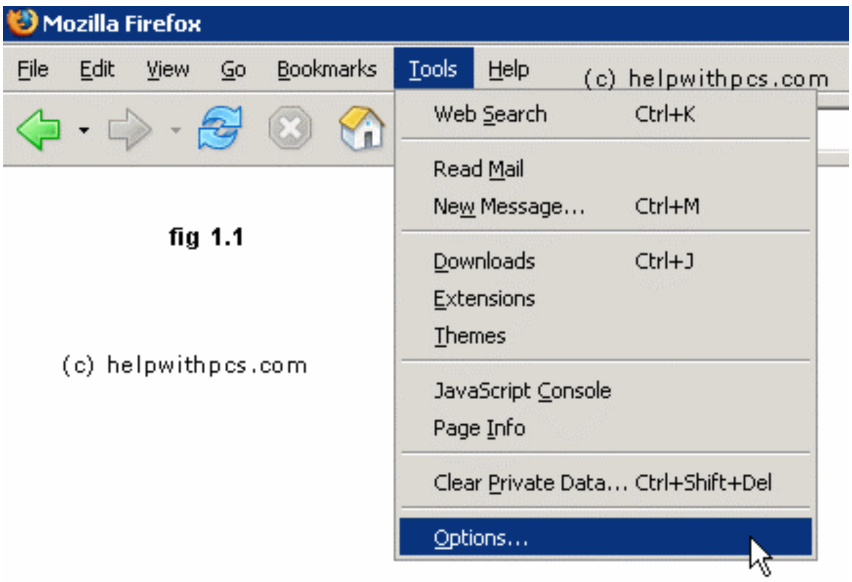


Figure 59: Menu bar in Mozilla Firefox[82]

3. Click on the "Main" tab

[82] www.helpwithpc.com, accessed the 20.07.2009

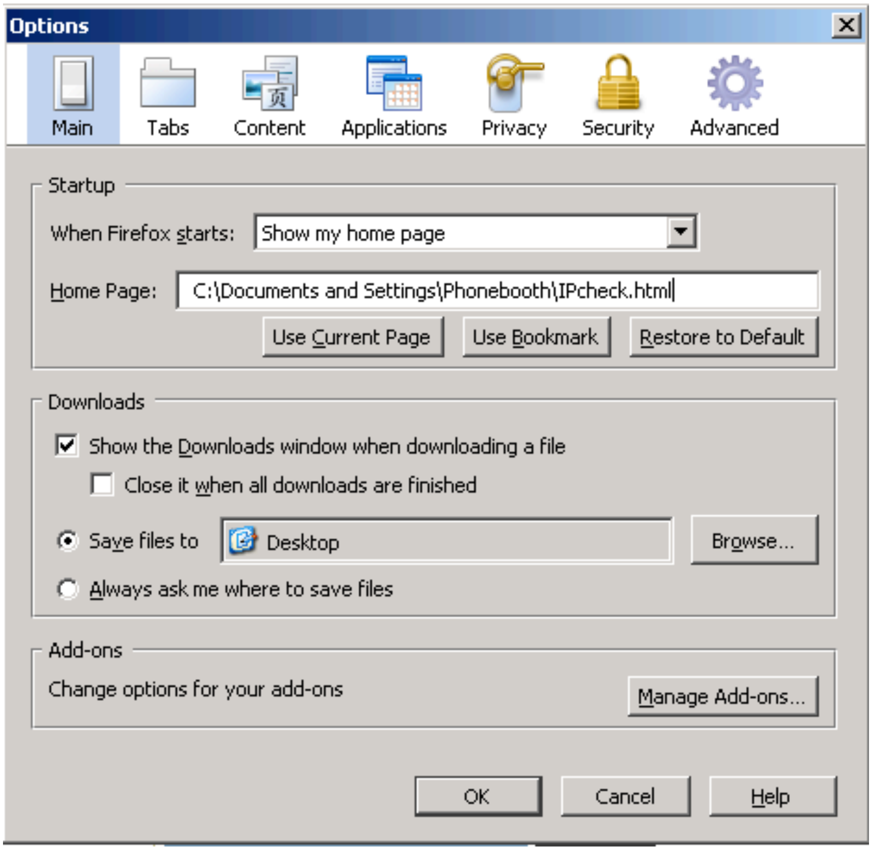


Figure 60: Settings in Mozilla Firefox „Main" tab

Therefore when Firefox is open, we click on Firefox in the taskbar and then click on Preferences and set on the main tab the home page to http://mail.google.com.

In Firefox we change the start-up page and point it to the locally stored HTML file. This way Firefox loads a locally stored file on start-up which doesn't require a network connection. After the network connection is established the JavaScript in the HTML file redirects the user to the login page of Google Gmail.

After the implementation is working on the software side the input devices for the audio and video signals are chosen and configured.

10.11 Hardware configuration

The following hardware is standard kit, which is available in any Google office to maintain interoperability between different devices. The camera is a Logitech QuickCam® Pro 9000 with built-in microphone to record the audio and video signal, which is then transferred via USB to the client machine. Every time the camera is active a red light indicates the recording.

Figure 61: Logitech QuickCam® Pro 9000[83]

As headset we are using the Sweex HM450 Headphones with integrated volume controller and a 3.5mm stereo plug. The 3.5mm plug will be connected to the audio line-in of the computer.

Figure 62: Sweex HM450 headphones[84]

As the client machine an IBM - Thinkpad T42.

Figure 63: IBM T42 laptop computer[85]

[83] http://www.logitech.com/index.cfm/webcam_communications/webcams/devices/3056&cl=US,EN, accessed the 16.06.2009

The keyboard of the computer will only be used to input the username and password.

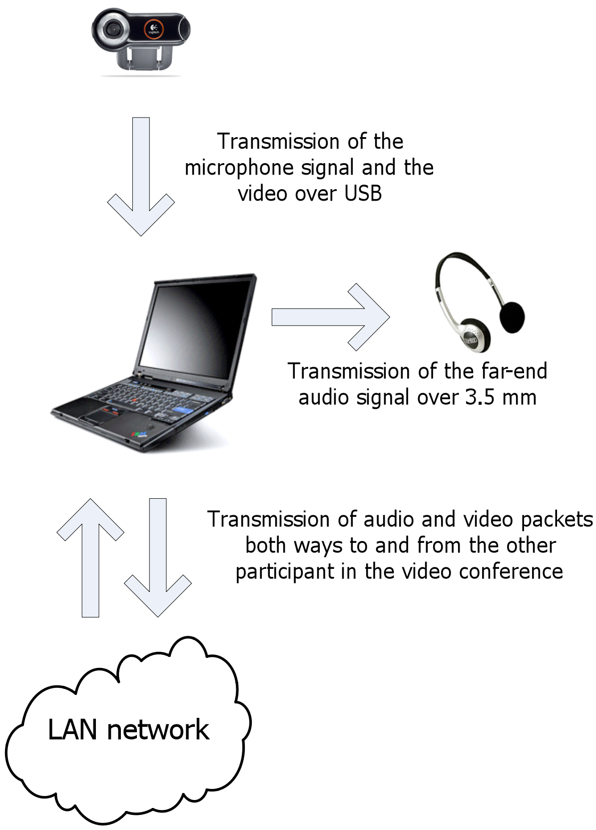


Figure 64: transmission of audio and video streams

10.12 Summary the hardware and software configuration steps

1. In order to fit the computer in the confined space of a telephone booth, an IBM T42 laptop computer is being configured as the kiosk.
2. On the laptop Windows XP is installed, as at the time of the development the Gmail Voice and Video plug-in was only working in a Microsoft Windows environment.
3. A local user account was created in Windows, to allow the users in the phone booth to login the operating system.
4. Microsoft SteadyState® was installed and configured. SteadyState® creates a kiosk environment primarily to restrict access rights of end-users.

[84] https://www.sweex.com/producten.php?lang=4&%20sectie=&subsectie=input &item=86&artikel=780&detail=a, accessed the 14.06.2009

5. Auto login was enabled. In the case of system power loss or shut down the system the machine logs itself into a new Windows session after rebooting.
6. The Windows shell was changed from explorer.exe to firefox.exe. This is another security measurement to prevent users to gain access to vital parts of the system settings of Windows operating system.
7. Installation of the R-kiosk add-on of Firefox: This web browser kiosk prevents Firefox user to surf other web content or change setting within Firefox. Firefox starts now full screen and can only be closed by using the key combination Alt + F4.
8. Firefox launches before the network drivers are fully loaded. That means that Firefox attempts to connect to the website http://mail.google.com fails. The solution is to load a script upon launch of the browser which checks for the local loop back address of the kiosk computer. A Java method in the script checks if the machine has successfully obtained an IP address. When it does the scripts redirects the user to the login page of Google Gmail.
9. Hardware devices as the Logitech QuickCam® Pro 9000 camera and the Sweex HM450 headphones are attached to the T42 laptop computer and the whole system is being placed into the phone booth.

[85] http://www.thelaptopcentre.co.uk/user/products/ibm-thinkpad-t42.jpg, accessed the 25.07.2009

11 Summary

The H.323 standard for video conferencing over packet-switched networks

In my thesis I have presented an overview over the H.323 standard and described components and protocols as well as audio and video compression in detail. In the first chapter I have given a detailed insight about the workings of packet-switched and circuit-switched networks. Subsequently the components of a H.323 infrastructure were at length presented following their function in a video conferencing network. An endpoint, MCU, gatekeeper and gateway are devices that have administrative and processing functions within a H.323 video conferencing network.

- Endpoints are the clients on the LAN that provide real-time bidirectional multimedia communications. An H.323 terminal can either be a personal computer (PC) or a stand-alone device, running an H.323 stack and the multimedia applications. It supports audio communications and can optionally support video or data communications.
- An MCU maintains the exchange of media in conferences with more than two participants. An MCU handles call control and the media exchange (voice and video) between the different participants during a conference.
- A gatekeeper is responsible for addressing, bandwidth control and security of a video conferencing network. It provides addressing schemes: (E.164 numbers) and alphanumeric names (H323 ID’s).
- A gateway is any device that allows calls to be established between networks, whether of the same or different types. It may also provide protocol conversion between H.323 endpoints and endpoints that do not support H.323. For example: An ISDN gateway does the protocol translation between a H.323 protocol stack to a H.320 protocol stack and vice versa.

Gmail Voice and Video plug-in

The Google Voice and Video plug-in offers private users as well as professionals the ability to perform video conferences over the Internet. It is implemented in the Gmail online webmail service of Google. The Google Voice and Video conferencing client makes video conferences available from any standard PC with the operating systems Windows or Mac OS X. The plug-in entails a H.264 encoder and GIPS iSAC audio codec. It uses the open standard XMPP and its extension protocols (XEP-0167 (Jingle)) to initiate and maintain video conferences with other Gmail Voice and Video user. XMPP uses the Extensible Markup Language (XML) as a base format for exchanging information.

The plug-in is free-of-charge and the requirements for it to run successfully is a broadband Internet connection a computer equipped with either Windows or Mac OS X and a camera, a microphone as well as headphones, or respectively speaker.

In comparison to a professional video conferencing system from Tandberg (Tandberg 1700 MXP) the Gmail Voice and Video plug-in offers the ability to collaborate via Voice, Video and instant messaging, to virtually any user in the world that have the necessary prerequisites.

The disadvantages of the plug-in are: The traffic is not encrypted. Bandwidth and quality of service cannot be guaranteed because the bandwidth of the Internet is shared between its users. Google Gmail Voice and Video plug-in is a 1:1 solution for two participants and is not H.323 compliant. Video conferences to H.323 enabled endpoints are not possible.

Development of a video conferencing kiosk for the Google London office phone booths

The project task was to develop a video conferencing solution to be implemented into the Google London office phone booths. It was a requirement to use the Gmail Voice and Video plug-in. The video conferencing solution should enable Google employees to perform video conferencing from the London office phone booths. The solution I have developed offers the following advantages:

- Security: It offers secure access to the video conferencing functionality of the Google Gmail voice and video plug-in, without giving the users exposure to the Google intranet. The access to any other pages of the Internet as well as access to other parts of the computer is restricted due to security requirements.
- Easy-to-use: the user is only exposed to the Gmail interface. This is the interface the user is familiar with and there is little reason for inhibition of adoption.
- Managing costs: The hardware components used are adapted for the confined space of a phone booth and are at the same time cost effective. The whole setup was established with off-the-shelf computer equipment.

The disadvantages of the Gmail Voice and Video plug- in are:

- Lack of H.323 integration: The plug-in is not interoperable with the H.323 standard.
- Lack of multi-site conferencing: The plug-in allows only 1:1 conversations and is therefore not deployable for team meetings or presentations.
- Network security: all voice and video traffic transmitted over the network is not being encrypted.
- No guaranteed bandwidth and connection quality.

Conclusion

The video conferencing kiosk in the phone booths in the London office will stimulate people to using the newly developed plug-in. However, it will not replace Tandberg video conferencing solutions that are already in place throughout the offices of Google. The comparison of the Tandberg 1700 MXP codec and the Gmail Voice and Video plug-in made very obvious that the plug-in is capable of handling consumer needs for private video conferencing sessions, but fails to serve the needs of modern business communication.

In corporate environments where reliability of a technical installation is number one priority the plug-in is not suitable. In our fast-paced environment no second look is wasted. Either something works right the first time or it is marked as unusable. Nobody plans in time for setting up a video conferencing session. The participants walk in the room and the technology needs to work instantaneously. This is the biggest advantage of the Tandberg kit at Google: reliability of service.

The Google Voice and Video plug-in and professional video conferencing equipment targeting different audiences and it will take some time that Internet based video conferencing applications offer the same reliability, security and connection quality as their H.323 equivalents.

Making the technology more powerful, more reliable and cheaper will be the enterprise for the second decade of the 21st century.

Bibliography

Books

(VCC), M. d. K. f. V. (2008). Videokonferenz Handbuch, TU Dresden.

Charbel, A. (2008) Schnell und einfach zur Diplomarbeit, Bild und Wissen

Firestone, S., T. Ramalingam, et al. (2007). Voice and Video Conferencing Fundamentals, ciscopress.com.

Goodman, D. and M. Morrison (2004). Javascript bible. New York, Hungry Minds ; Chichester : John Wiley.

Henle, H. (2001). Das Tonstudio Handbuch: Praktische Einführung in die professionelle Aufnahmetechnik. Grundlagen der Akustik. Analoge und digitale Audiotechnik. ... Lautsprecher und Regieraum-Design, GC Carstensen Verlag.

Karp, D. A. (2003). Windows XP pocket reference. Sebastopol, Calif. ; Farnham, O'Reilly.

Karrenberg, U. (2005). Signale Prozesse Systeme. Berlin, Springer.

Kraut, R. E., R. E. Rice, et al. (1994). Life and death of new technology: task, utility and social influences on the use of a communication medium. Proceedings of the 1994 ACM conference on Computer supported cooperative work. Chapel Hill, North Carolina, United States, ACM.

Niderst, J. (1999). Webdesign In A Nutshell - A Desktop Quick Reference. Cambridge, O'Reilly & Associates

Rossig W., J. Praetsch (2008). Wissenschaftliche Arbeiten BerlinDruck, Achim

Schmidt, U. (2003). Professionelle Videotechnik, Springer

Tanenbaum, A. S. (2003). Computer Networks, Pearson Education International.

Wainfan, L. and P. K. Davis (2004). Challenges in Virtual Collaboration, National Defense Research Institute

Wilcox, J. R. and D. K. Gibson (2005). Video Communications - The Whole Picture, CMP Books

Winder S., M. Tooley (2002). Data Communciations, Newnes

Journals

Friebel, M., J. Loenhoff, et al. (2003). "Siehst Du mich?" – „Hörst Du mich?" Videokonferenzen als Gegenstand kommunikationswissenschaftlicher Forschung." kommunikation@gesellschaft(Jg. 4, 2003, Beitrag 1)

Global_IP_Solutions (2008). How to select the best codec for HD Voice.

ITU-T (2005). ITU-T Recommendation H.239.

ITU-T (2007). Packet-based multimedia communications systems

Linden, D. J. (2008). iSAC Supporting the Wideband VoIP Evolution.

Tandberg (2008). Tandberg Product sheet 1700 MXP, Tandberg

Schmitz, H. W. (1999). "Videokonferenz als eigenständige Kommunikationsform Eine explorative Analyse. Unveröffentlichter Vortrag."

Solutions, G. I. (2008). How to select the best codec for HD Voice.

Internet sources

Email marketing reports

http://www.email-marketing-reports.com/metrics/email-statistics.htm

date: 03.06.2009

Google

http://www.google.com/apps/intl/en/business/messaging.html#mail

date: 03.06.2009

http://gmailblog.blogspot.com/2008/11/say-hello-to-gmail-voice-and-video chat.html

date: 02.06.2009

http://www.google.com/apps/intl/en/business/index.html

date: 02.06.2009

http://code.google.com/apis/talk/open_communications.html

date: 02.06.2009

http://groups.google.com/group/Gmail-Help-Chats-and-Contacts-en/browse_thread/thread/e34c6524cda99b2/13499e3de3f3707c?lnk=gst&q=

date: 15.07.2009

Internet Engineering Task Force

http://tools.ietf.org/html/rfc3550

date: 03.06.2009

Marratech

http://www.marratech.com/forum/index.php?showtopic=1196

date: 22.11.2008

http://www.marratech.com/forum/index.php?showtopic=2129&hl=Mac

date: 22.11.2008

Microsoft

http://www.microsoft.com/windows/products/winfamily/sharedaccess/

date: 16.07.2009

North Carolina State University

http://www.ncsu.edu/scivis/lessons/colormodels/color_models2.html

date: 14.06.2009

Skype

http://www.skype.com/

date: 04.06.2009

University of Southampton

http://www-mobile.ecs.soton.ac.uk/speech_codecs/standards/adpcm.html

date: 12.06.2009

World wide web consortium

http://www.w3.org/TR/html401/

date: 13.06.2009

http://www.w3.org/TR/CSS21/

date: 13.06.2009

XMPP protocol

http://xmpp.org/extensions/xep-0167.html

date: 02.06.2009

Personal contacts

Martin, Alberto (Technical Programm Manager at Google): description of the Google Voice and Video plug-in

Bond, James (Head of Information Technology at Google): Assignment of the project and allocation of the funds

Earnshaw, Rick (Field tech at Google) code review

Statement of Authorship

This work has not been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where reference is made in the thesis itself.

Ort, Datum

Signature