The MAYAH range –
Professional Audio & Video Encoding/Decoding, Recording,
Mixing and Transmission Products

- CENTAUR®II
  Multichannel Audio Gateway Codec
- Sporty
  Portable Reporter Codec
- MERK II
  Portable Audio Codec Mixer
- IO [io]
  Audio Video Encoder/Decoder
- ganymed
  IP Audio Encoder/Decoder
- Flashman II
  Portable Audio Recorder/Codec
Abstract:
Audio transmission over IP networks is becoming increasingly important for the radio industry since availability and bandwidths are rising, while costs are going down almost continuously. The supply of dedicated and ISDN lines is declining worldwide, so sooner or later we may face a lack of a particular service from the Telcos. Already some ISDN transmissions are being carried over IP Backbones.

This compendium covers exactly these topics; from the basics through to todays applications. By no means complete, it does however provide many interesting aspects for both beginners as well as experienced Audio-via-IP users.

We start with the chapter “IP Audio Compact” which will provide easy and fast reference.
IP Audio Compact

UDP, RTP, SIP… – IP Protocols, what are they all for?
The chosen protocol determines mainly the characteristics of the audio transmission in a network. Due to the fact that there are high requirements regarding the signal quality, the appropriate protocol needs to be selected.

RTP/UDP is typical for Corporate Networks and VPNs with dedicated communication partners, while SIP is the protocol used on the public Internet; it is particularly popular for VoIP applications.

SIP Session Initiation Protocol
Session Initiation Protocol (SIP) is a text-based protocol, for negotiation of connections based on the Internet Protocol (IP). SIP is used merely to handle the signalling between individual negotiation parties. The transport of media data runs – similarly to ISDN – separately from the negotiation. The media data are often sent over a different route, conveyed by a different Transport Protocol. If TCP or UDP are available for the negotiation, RTP is mostly used for the media data transport.

SIP-connections as substitution for ISDN
The new SIP connections in the public Internet can be considered as equivalent to ISDN if there is a loss free transmission with low latency. This can be achieved with guaranteed bandwidth, e.g. using RSVP and FEC.

VPNs with guaranteed bandwidth as substitution for E1 and X.21 connections
Virtual Private Networks, so called VPNs can substitute todays E1 or X.21 connections, if a stable and guaranteed bandwidth is achieved.

Using IP in existing E1 or ATM networks
The IP-protocol can also be applied in existing E1 or ATM connections. This allows the simultaneous use of audio/video and data signals. The reservation of the corresponding bandwidth has to be done within the routers.

HE AACv2 etc - Audio Coding Formats, which one for which application?
The huge number of audio coding formats can be confusing sometimes. The selection needs to be defined based on criteria like compatibility, quality and latency. While MPEG 4 HE AACv2 is providing excellent quality at very low bit rates at e.g. 48kbps stereo, transparent AES/EBU transmission is used in networks with more bandwidth to be completely loss less and obtaining production quality with up to 3 MBit/s.

Migration from ISDN to IP and vice versa
For the migration process of existing ISDN to IP infrastructure, a certain time will pass. Therefore technical solutions are required which allow transcoding of formats and bridging of networks, so called Gateways.

FEC – Forward Error Correction
FEC permits error detection and/or correction by adding redundant data to the stream. Thereby avoiding re-transmission or corruption of data and the associated costs of higher bandwidth needs and increased delay.

IP Overhead
At IP-transmissions the data stream consists of Payload (pure audio data e.g. one or more MPEG frame) and IP Overhead. In case of UDP protocol the absolute overhead is as large as 46 Bytes / packet and it consists of 8 Bytes UDP header, 20 Bytes IP header and 18 Bytes IEEE802.3 (Ethernet). RTP protocol has 58 Bytes overhead per packet, 12 Bytes - RTP, 8 Bytes UDP, 20 Bytes IP and 18 Bytes IEEE802.3. The relative IP Overhead is calculated in percents of the original payload.

MPEG TS via ASI, MPEG TS via IP
The MPEG Transport stream, named MPEG TS is used mainly in DVB and ipTV environments. It is a format containing one or more audio and/or audio/video streams which, typically, need to reach set-top boxes as final receivers.
Are standards available?
In order to result in common implementations, the EBU has established the working group N/ACIP. In September 2007 the first final recommendation shall be officially released.

RTP (Realtime Transport Protocol)
is a session protocol for IP-based networks which uses UDP (User Datagram Protocol) as transport protocol. In opposite to the pure UDP implementation RTP guarantees the right sequence of packets at the receiver side.

RTCP (RTP Control Protocol)
is working on top of the RTP. Not interfering with a real data stream it provides some control data exchange between streaming session participants. Receiving a feedback information as packets sent, lost packets, round trip delay, jitter, etc., the streaming can be accommodated on current network conditions.

SDP (Session Description Protocol).
Descriptions of multimedia streaming sessions with SDP include important initialization parameters information. They are being used to invite participants and to announce the start of the session or to initialize the connection on some other way.

SAP (Session Announcement Protocol)
is used to announce the information provided by SDP on some well known multicast address to give the potential clients an overview of the contents available on the current (sub) network. MAYAH supports in addition a non traditional way for SAP unicast.

RTSP (Real Time Streaming Protocol)
is an application protocol for use in streaming media systems. It lets client to know about the contents available for streaming and to choose the appropriate stream as well as to control a streaming media server with a typical commands such as “play” or “pause”.

Contents

1 Audio and the introduction of IP
   1.1 IP Basics
   1.2 Packet Switching versus digital point-to-point connection

2 Audio coding for IP: Latency vs Bitrate
   2.1 G.711
   2.2 G.722
   2.3 MPEG Layer 2
   2.4 MPEG Layer 3
   2.5 MPEG 4 HE AAC
   2.6 apt-X and Enhanced apt-X
   2.7 Linear Audio and AES/EBU transparent

3 IP as a platform for modern media and telecommunication networks
   3.1 Basic requirements for media transport

4 Switched Communication in the Public Internet
   4.1 ISDN – SIP a comparison
   4.2 SIP – and how it works
   4.3 Operational security
   4.4 SIP for high quality media links
   4.5 SIP -> Future

5 Practical considerations for SIP and audio-via-IP
   5.1 Requirements for SIP based connections via the public Internet
   5.2 Requirements for other IP based connections via the public Internet – Unicast vs. Multicast

6 Typical IP or mixed IP/ISDN applications

7 References
   7.2 G. Stoll, F. Kozamernik, “EBU subjective listening tests on low-bitrate audio codecs”, EBU, 2003
   7.3 “Too many audio codecs”, Detlef Wiese, Tonmeistertagung, 2002
   7.4 White Paper “Discussion on coding algorithms”, Detlef Wiese, 2002
   7.5 White Paper “NAT Traversal for Multimedia over IP”, Newport Networks, 2005
   7.6 White Paper „Understanding SIP“, D. Sisale, J. Kutan, GMD Fokus, 2000
   7.7 CENTAURI II Manual, Werner Ludwig, 2006
1 Audio and the introduction of IP

1.1 IP Basics

Following the OSI layer model there was a TCP/IP reference model for the Internet Protocol family defined. It describes a structure and interaction of the different network protocols in four layers: Network Layer, Internet Layer, Transport Layer and Application Layer, which includes protocols related to certain tasks, like Telnet or FTP (Figure 1).

<table>
<thead>
<tr>
<th>Layer 7</th>
<th>Application Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer 6</td>
<td>Presentation Layer</td>
</tr>
<tr>
<td>Layer 5</td>
<td>Session Layer</td>
</tr>
<tr>
<td>Layer 4</td>
<td>Transport Layer</td>
</tr>
<tr>
<td>Layer 3</td>
<td>Network Layer</td>
</tr>
<tr>
<td>Layer 2</td>
<td>Datalink Layer</td>
</tr>
<tr>
<td>Layer 1</td>
<td>Physical Layer</td>
</tr>
</tbody>
</table>

Defining application protocols (e.g. FTP, TELNET, mail...)

Describing the data presentation (e.g. ASCII, Unicode)

Connecting and disconnecting interlinks, login, logout, handshake for non duplex connects...

DATA transport between communication partners or applications (e.g. TCP, UDP)

Interconnecting network stations or segments (IP)

Transfer protocol for data framing and secure transfer (e.g. HDLC, SLIP, PPP)

Specifying the bit transportation on the link media (e.g. modulation, pin assignment of the connector...)

---

ISO/OSI reference model to TCP/IP reference model

<table>
<thead>
<tr>
<th>Application Layer</th>
<th>Application Layer</th>
<th>FTP, HTTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation Layer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session Layer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transport Layer</td>
<td>Transport Layer</td>
<td>FTP / UDP for Media Transport</td>
</tr>
<tr>
<td>Network Layer</td>
<td>Internet Layer</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>Datalink Layer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Layer</td>
<td>Network Layer</td>
<td>IEEE 802.3 or Ethernet or IEEE 802.11 on wiLe</td>
</tr>
</tbody>
</table>
In order to exchange information between applications a minimum of three addresses is required. The individual partners within a physical network segment (Ethernet IEEE802.3) are identified by a 6 byte MAC address (Media Access Control).

This address needs to be unique within a segment. MAC addresses are assigned typically by manufacturers, identifying him by the first three bytes. In order to recognize participants within the Internet, four byte long (IPv4) or 16 (IPv6) byte long Internet addresses are used. The relation between physical MAC and logical IP address is achieved with the address resolution protocol (ARP). The IP addresses of sender and receiver are part of the IP Protocol Header.

The third address is the so called port- or socket number. This one identifies a certain application of the participant. The port number (16 Bit) is located within the transport layer. Certain port numbers are already assigned to certain applications, e.g. FTP is always using port 20/TCP for data and 21/TCP for control information. Telnet is using port 23 and for SMTP always port 25 is used. The complete list of port definitions can be found at www.iana.org/assignments/port-numbers. IANA (Internet Assigned Numbers Authority) is also responsible for the allocation of registered port numbers.

1.2 Packet Switching versus digital point-to-point connection

In broadcasting mainly digital point-to-point connections are used for the long distance transport of media data. Normally those lines are of high availability, constant delay and negligible jitter without Sequence Errors or Packet Losses.

For File Transfer, but also increasingly for the Live Transmissions in the field of Distribution and Contribution the use of Internet-Services is growing. Already in 2001, ORF (Austrian public radio & television) has begun with an implementation of their L-Net. Although at that time and during the following years a focus was mainly on the so called Corporate Networks, today the public Internet is becoming more important.

Although not planned originally for the Internet, the descending availability of ISDN in broadcasting requires more and more packet transmission in this media. The inventors of the Internet have defined the Internet Protocol in 1981 and thus, offered some possibilities to use it for realtime purposes. At the moment standard connectivity does not support this. The IP Header includes a flag, type-of-service, which is used to mark the priority of the data contained. In the public net, such are either ignored by the routers or even reset. The reason is obvious: as soon as the type-of-service flag is really applied, e.g. the „Internet Gamer Community“ might use this immediately for games in order to have them prioritised.

For reliable transmission of media data via a packet-based network particular requirements have to be fulfilled: Delay, Jitter and Availability of the Network must remain in a known range, suitable for a 24/7 broadcasting application. The necessary bandwidth for the signal transmission must be provided at any time. It also must be known if any Sequence errors occur in the order of the arriving datagrams.

For a stable service there are appropriate means and tools to be applied. Maximum jitter defines the size of a compensation buffer at the receiver side. Possible packets losses can be recovered using the Forward Error Correction (FEC). Sequence errors can be avoided using an appropriate protocol at the Transport Layer (e.g. RTP, Realtime Transport Protocol).

The influence of certain means, such as protocols and FEC has a quite measurable impact on the latency of the transmission route. A further difference to conventional digital dedicated lines is the overhead caused by the protocols.

Each Ethernet datagram occupies 18 Byte for the IEEE802.3 header and 8 Byte for its prefix. The IP header occupies 20 Byte, the UDP header further 8 Byte and an optional used RTP header another 12 Byte. Totally, 66 Byte header data per RTP datagram are transmitted. If in one RTP datagram exactly one MPEG 2
Layer 3 audio frame with 16 kHz sampling rate and a bit rate of 24 kbps needs to be transported (48 Byte), a total bandwidth of 57 kbps will be needed.

If MPEG4 AAC is used instead of MPEG 2 Layer 3 an exact calculation of the required bandwidth is not possible anymore because the length of the AAC audio frame is changing from frame to frame and thus, the relation between audio data and header, the overhead, can not be calculated. In order to minimize the impact of the protocol overhead, more than one audio frame can be combined in one datagram. This is increasing the latency, because the necessary number of audio frames has to be collected first.

Applying this to the MPEG 2 Layer 3 signal as mentioned above, it means that with optimum payload totally 30 complete audio frames need to be transported in one datagram. Each audio frame contains 576 samples and, at a sampling rate of 16 kHz exactly 36 ms of audio. Thus, an additional latency of 1044 ms, (figure 3) is created.

Another option to reduce the protocol overhead is to remove IP, meaning the UDP and RTP header, which is a technique of systems like Ethersound and Cobranet. Media data are transported as payload of IEEE802.3 packets. In that case, IP addresses are missing and only participants of the same network segment can be identified via MAC addresses. A transmission outside the segment, e.g. via a gateway, is not possible. Even in case the MAC address of an outside participant is know, the router would not find the path, because the logical address of the outside network segment is still missing.
Audio coding for IP: Latency vs Bitrate

During the last twenty five years various coding algorithms were established in the broadcasting industry, e.g. J.41, J.57, MPEG Layer 2 and 3, as well as non-standardized algorithms such as apt-X or ADPCM4SB (Micda). Because of a variety of demands it’s not easy for an operator to choose the proper bitrate, operation mode and sample rate. It becomes more difficult with the flow of the new algorithms developed in the last years.

Besides those known formats, others, e.g. from telephony, such as G.711 and G.722 have been applied in broadcasting as well as the more modern and successful AAC variations, such as MPEG 2 and 4 AAC, HE AAC (formerly aacPlus), HE AACv2, AAC ELD*, as well as linear audio or AES/EBU transparent transmission. All those formats with modes, like mono, stereo, in all different sampling- and bit rates, some even in multichannel (5.1/7.1) technique.

The development of coding algorithm has been given a focus to optimization of such parameters as bitrate, quality – also after multiple en/decoding, latency and compatibility. Algorithms described here can be categorized in such a way. HE AACv2 was developed solely for reduction of the bitrate while reaching a very good sound quality. On the other hand, apt-X always focused on the delay parameter.

Due to the growing bandwidth available there are linear transmission methods, such as 16, 20 or 24 bit linear audio or even the transparent transmission of 3,072 MBit/s AES/EBU come today more and more to consideration. Here are quality and delay playing the main roles at the cost of bandwidth.

There is no detailed research on market share of various coding algorithms in the broadcasting world. However it is widely assumed that MPEG Layer 2 and G.722 are dominate, while such algorithms as linear audio, ADPCM, apt-X and Enhanced apt-X, as well as MPEG 4 Standard HE AAC are being increasingly used for corresponding applications.

A detailed discussion on choosing a proper method must include consideration on quality, flexibility, bandwidth, latency, compatibility, standardization, market share and expectations. In fact an optimal solution can be found for almost any system.

*work item of MPEG
2.1 G.711
One of the most important standards of ITU-T is G.711 which digitizes mono audio at a sampling rate of 8 kHz. This covers the frequency range between 300 and 3400 Hz. In Europe 64 kbps and in North America 56 kbps, are used for classical telephony with G.711. If a conventional telephone needs to be reached via IP, G.711 is used in VoIP. The latency of G.711 is not noticeable.

2.2 G.722
Another ITU-T standard is G.722, which offers a higher audio quality due to its 16 kHz sampling rate. The transmission bit rate is also 64 kbps (e.g. one ISDN B channel). Similar to G.711, latency is very low.

There are two possibilities to synchronize audio codecs with G.722:
- **G.722 with H.221 Inband Signalling (G.722/H.221):** In G.722/H.221 a bitrate of 1.6 kbps is occupied for Inband-information which is used for synchronisation.
- **G.722 with statistical synchronisation (G.722/SRT):** With G.722/SRT (SRT = Statistical Recovery Timing), a statistical analysis of the audio signal results in the recognition of the beginning of one Byte

2.3 MPEG Layer 2
Dating from the 1990’s, Layer 2 is still very popular and widely used in broadcasting applications. The main reasons for its popularity are its cascadability (good for a lossy codec), its high audio quality at high bit rates and availability in lots of first generation hardware & software products. MPEG Layer 2 supports bit rates of 8 to 384 kbps; target bit rate for stereo is 256 kbps.

2.4 MPEG Layer 3
Better known as mp3, this format is also used in broadcasting, e.g. when lower bit rates than Layer 2 are required. MPEG Layer 3 supports bit rates between 8 and 320 kbps with a target bit rate of 128 to 192 kbps for stereo.

2.5 MPEG 4 HE AAC
HE AAC is a further development of AAC using the so called SBR, Spectral Band Replication from Coding Technologies (www.codingtechnologies.com). AAC is the audio coding algorithm in MPEG 2 & 4 with the highest audio quality and a target bit rate of 128 kbps stereo. Because many applications need even lower bit rates, the Swedish-German co-operation in Coding Technologies invented the SBR technology which allows AAC with bit rates of even 32 or 48kbps stereo.

More than 90% of the bit rate is still used by the conventional AAC encoding, but a small part (<4kBit/s) is used for the SBR-information. The conventionally encoded AAC part is sampled with half the sampling frequency, 16, 22.05 or 24KHZ which results in a higher coding efficiency.

The combination of SBR and AAC is a high quality format. Although no claims for real transparency can be made, since the frequencies above 7/8 kHz are synthesized, it is still possible to reach an extremely good, CD like quality at such low bitrates. In connection with HE AAC, CD-like really means high quality, which is proven by the many applications and systems using it. Combined with so called “parametric stereo” coding, the algorithm is known as HE AACv2 and provides astonishing sound quality even at 16, 20 or 24 Kbit/s.

2.6 apt-X and Enhanced apt-X
Apt-X is a low latency coding format using ADPCM (Adaptive Differential Pulse Code Modulation). A typical data rate is 192 kbps for mono and sound quality is retained even after multiple en/decoding cycles; so called cascading. Theoretically the minimum latency is 3 ms at a sample rate of 48 kHz. The algorithm can be used at various other sample rates and recent improvements, especially to the dynamic range, resulted in Enhanced apt-X, which uses word lengths up to 24 bit. Apt-X is one of the most widely used systems for audio transmission with a short delay. Its key features are:
- 4:1:4 data compression
- Mono / stereo audio encoder / decoder
- Flexible sample rate up to 96kHz
- Ancillary data up to 12 Kbit/s
2.1 Linear audio and AES/EBU transparent

The increasing availability of higher bandwidths means the use of linear audio and “AES/EBU Transparent” is a realistic alternative. Linear audio is understood as a PCM signal with a specified word length of 16, 20 or 24 bit and a fixed sample rate. In the broadcasting world this is usually 48 or 96 kHz, resulting in a bitrate of 1.5 to 4.5 Mbit/s for a stereo signal.

An AES/EBU Transparent transmission can include Dolby E or DTS etc within the AES/EBU Signal without any sample rate conversion (the encoded data would otherwise be irreversibly corrupted). The data rate of AES/EBU signals is 3.072 Mbit/s with multichannel signals being a multiple of this.

2.2 Multichannel

Encoding of 5.1 or 7.1 multichannel signals is supported by a variety of formats. In particular HE AAC offers very efficient coding with bit rates below 128 kbps. Enhanced apt-X provides up to 8 channels with bit rates between 1-2 Mbit/s while linear formats move the bit rate up to 18 MBit/s.

Audio Encoding Algorithms: Quality and Delay

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>&lt; 4 kHz</th>
<th>&lt; 8 kHz</th>
<th>&lt; 16 kHz</th>
<th>&gt; 16 kHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.711</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>MPEG 1 Layer 2</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>MPEG 2 Layer 2</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>MPEG 1 Layer 3</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>MPEG 2 Layer 3</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>MPEG 2.5 Layer 3</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>mp3PRO</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>MPEG 2/4 AAC</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>MPEG 2/4 AAC LD</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>MPEG 4 HE AACv2</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>J.41</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>ADPCM45/BICM60</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>EapX / apt-X</td>
<td></td>
<td></td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>J.57</td>
<td></td>
<td></td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>Linear Audio</td>
<td></td>
<td></td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>AES/EBU Transparent</td>
<td></td>
<td></td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>5.1/7.1 AAC und HE AACv2</td>
<td></td>
<td></td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>8 Channel linear Audio</td>
<td></td>
<td></td>
<td>variable</td>
<td>variable</td>
</tr>
<tr>
<td>8 Channel apt-X / EapX</td>
<td></td>
<td></td>
<td>variable</td>
<td>variable</td>
</tr>
</tbody>
</table>

*proprietary enhancement by Fraunhofer

Audio Coding Algorithms: Sample Rates in kHz

- G.711
- G.722
- MPEG 1 Layer 2
- MPEG 2 Layer 2
- MPEG 1 Layer 3
- MPEG 2 Layer 3
- MPEG 2.5 Layer 3
- mp3PRO
- MPEG 2/4 AAC
- MPEG 2/4 AAC LD
- MPEG 4 HE AACv2
- J.41
- ADPCM45/BICM60
- EapX / apt-X
- J.57
- Linear Audio
- AES/EBU Transparent
- 5.1/7.1 AAC und HE AACv2
- 8 Channel linear Audio
- 8 Channel apt-X / EapX

Audio Coding Algorithms: Bit rates in kbit/s

- G.711
- G.722
- MPEG 1 Layer 2
- MPEG 2 Layer 2
- MPEG 1 Layer 3
- MPEG 2 Layer 3
- MPEG 2.5 Layer 3
- mp3PRO
- MPEG 2/4 AAC
- MPEG 2/4 AAC LD
- MPEG 4 HE AACv2
- J.41
- ADPCM45/BICM60
- EapX / apt-X
- J.57
- Linear Audio
- AES/EBU Transparent
- 5.1/7.1 AAC und HE AACv2
- 8 Channel linear Audio
- 8 Channel apt-X / EapX

*proprietary enhancement by Fraunhofer
3 IP as a platform for modern media and telecommunication networks

3.1 Basic requirements for media transport

From an economical point of view, classical broadcasting services offer a coverage to as many destinations as possible for low cost. Streaming media services, typically get more expensive with each destination, because the data have to be send individually. In the network environment, multicast is known very well. It allows to send data to an unlimited number of destinations at the lowest network load possible. Multicast is not used very often, because it is not supported by all routers in the network.

Instead, streaming services to a mass audience (e.g. sport events or big shows) are transported via so called overlay networks. They provide the same data at many locations simultaneously. High quality streaming requires the transport protocol RTP/RTCP which is the basis for realtime transmissions. For control of streaming content the protocols SIP, SAP, RTSP and HTTP are applied. Such protocols are controlled via different ports and offer control functionality. They all have RTP as the basic protocol for information exchange.

![RTP/TCP: Protocol Architecture](image)
Switched Communication in the Public Internet

4.1 ISDN – SIP a comparison

Compared to ISDN the data packets of an IP network are always supplied with the source and destination addresses by the terminals. So they are being sent with no prior connection establishment. To allow the switched connection within such non-connection-oriented networks the Session Initiation Protocol (SIP) has been developed by the Internet Engineering Task Force (IETF) group (RFC 3261).

During the establishment of the connection via ISDN, the initiating party dials to the communication partner. This request is presented to the network which accepts and – if there is enough capacity or bandwidth – send back a receipt confirming the reservation of the channel or bandwidth. Such a network typically consist of multiple network segments and such request for reservation is exchanged and confirmed between each of them. That is how the request can move from a local to a regional and then to a national/international network. As soon as the receiver is found and is in the position to accept the call, both communication partners are connected directly via such reserved channels.

In order to manage this communication efficiently, the telecommunication network is separated in two networks: the signalling and the user data channel network. While the signalling network send all necessary information for the connection itself, the user data channel network is a free switchable channel, which allows 64 kbps within the ISDN.

Due to the high coverage of packet orientated networks with higher bandwidth, more service can now move to such networks. Although it started originally as a pure data and information exchange platform, e.g. with Email, now it allows also services like TV, radio and telephony. „Voice over IP“ is the successor for the current telephone network.

Other than ISDN, data packets are assigned to sender and receiver address and connected immediate without dialing process. In order to allow in such a network also dialing, the Session Initiation Protocol has been developed. Although the details of SIP remind very much the details of the D-channel protocol of ISDN, it was not the telecommunication bureau, but the Internet Engineering Task Force
One of the advantages of SIP is the possibility to negotiate about the settings and properties of the media data connection, so they do not have to be defined by the calling party. By all means it is possible, instead of one individual audio codec, to exchange some priority lists, which can be modified or replaced by the remote side. Since the messages exchange during the connection set-up consists of minimum of three messages, the calling party always has the last word, and can actually define the audio codec for the later usage.

Another advantage of SIP reveals itself by closer look on the connection establishment using an SIP-Server.

Addresses of SIP-clients are marked with a prefix “sip:” and appear mostly in a following form: “sip:user@host”, “sip:user@domain” or “sip:user@ip_address”. In the simplest case as a direct connection between two end devices there are “host”, “domain” and the “ip_address” of the end device meant. That means that the IP address of receiver must be known or the host or domain must be resolvable by the common mechanisms like Domain Name Service (DNS). In practice it is rather rare, so there is a possibility to detect the terminal with the help of SIP-Servers.

4.2 SIP – and how it works

Session Initiation Protocol (SIP) is a text-based protocol, for negotiation of connections based on the Internet Protocol (IP). SIP is used merely to handle the signalling between individual negotiation parties. The transport of media data runs – similarly like by ISDN – separately from the negotiation. The media data are often sent over a different route, conveyed by some other Transport Protocol. So if there are TCP or UDP for the negotiation available, RTP is mostly used for the media data transport.

In the simplest case (Peer-to-Peer) individual components are connected directly to each other with no any central unit., s. Figure 3. Within one session the entities of one Peer are called User-Agents (UAs). One UA can play two roles:

- **User-Agent-Client (UAC)** initiates the session, and
- **User-Agent-Server (UAS)** answers to the requests of the client.

One SIP-device can do both. To start such negotiation there is an INVITE-message sent from the calling party to the SIP-address of the receiver. The keyword INVITE is one of the so called Methods, which any User Agent must handle. Other Methods are for example ACK, BYE or REGISTER. Should the receiving party accept the call, it can answer directly with an OK-message, which is then acknowledged by the calling party with the ACK message. These three messages are enough in the simplest case to establish the media data connection. However the three mentioned messages mostly have a specific additional content in the so called Body. The Body contains a description of the media data connection being established. Here it is specified which audio and video codecs must be used, which parameters they must use and which IP-addresses and Ports have the user terminals. To describe the media data connection there is a Session Description Protocol (SDP) used. A protocol which is successfully used for other negotiation and transmission mechanisms too and which has established itself a long time ago.

(IETF) which finalized the documentation which is now Request For Comments (RFC 3261).

The working group did base their work on already existing blocks and thus, it is not a surprise that many elements have similarities to HTTP or SMTP.
Let’s examine a following case as an example. Terminal 1 is wishing to establish connection to the mobile Terminal 2. IP-addresses of the remote unit are unknown to each of the parties. It may be some mobile devices connected to the network over UMTS or ADSL. To establish connection nevertheless, both terminals log on at some SIP-Server with a special SIP-message. To do so the user identifies itself with a previously assigned SIP-address in form of “user@service-provider” and mostly with some password. The SIP-Server stores the information about the current IP address of the user at the Location-Server. So during the connection request of the Terminal 1 there is a registered address of the Terminal 2, for example “device2@service-provider”, is used. The INVITE-message is then sent to the SIP-Server, which on his part asks the Location-Server for the IP-address of the Terminal being called and transmits the answer in the slightly changed form to the Terminal 1. Terminal 2 answers with the SIP-messages the SIP-Server, which forwards the messages again to the Terminal 1. Messages “100/Trying” shown in the Figure 4 are optional and serve for the acknowledgment of the request. “180/Ringing” messages are optional too. They solely signalling to the calling party about the ring tone at the receiver side.

As shown in the figure, the devices never communicate directly except using „ACK“ information. Communication is running only via the SIP-Proxy. From the device point of view the negotiation is handled like a direct negotiation without proxy – i.e. the SIP-proxy is transparent. The example does not show the connectivity of participants at different SIP-servers, which is typical and possible without any problems, because connection requests are simply passed from one SIP proxy to another. Between two devices there can be just one or even a chain of proxy servers.

Also, proxy server do not necessarily need to be with just one provider. Thanks to open source solutions like Asterisk or routers with integrated SIP server functionality building a network structure is quite simple. A new protocol is bringing new potential problems, obviously. Like in any peer-to-peer protocol, firewalls can stop a requested communication also during SIP connections. All outside packets (mainly from the Internet) are defined as non-requested and are rejected by the firewall. Assuming that both devices of a SIP connection are behind the firewall, none of them can initiate a connection because all requests will be rejected.

With the appropriate settings of the firewall, e.g. port-forwarding, this can be avoided. Another problem is the IP address of the destination device, which is necessary for the user data connection and part of the SDP information. If Network Address Translation (NAT) is applied, as it is common in non-public networks, the IP address can not be reached from another network and thus, is wrong. Non-public networks are using typically IP addresses being part of a special address area which are only allowed to be used in private networks. Examples are addresses with the structure 192.168.X.X or 10.X.X.X. A packet of this type in the public Internet will be rejected immediately.

NAT takes care of the IP packets transition from the private to the public network and overwrites the non-public address of the sender with a public one. By the incoming packets there are public addresses of the destination being converted to the private addresses. Please note, that only IP addresses in the header are changing. Contents of the packet (e.g. description of the media data connection in the SDP part by SIP) remains untouched. There are various possibilities to translate the addresses. Those different methods can lead to the different addresses or ports written as a sender ID. Depending on type of the NAT seem the translated packets (the packets reaching the public network) to come from different addresses or ports and this can be unpredictable. So it’s sometimes difficult to put the proper address and port number in the content (Body) of a SIP message.
data, such content is worth to protect. Such a protection mechanism is not part of the SIP standard. Other, already known and introduced mechanisms can be used, examples are:

- **VPN**: Participants of a VPN can exchange data like within a LAN. The individual participants do not need to be connected directly. The connection via the public Internet is encrypted.
- **ATM or MPLS**: services of the transport layer, based on address mechanisms
- **IPSec**: tunnel based on the Internet protocol

One of the most critical security risk is an attack from outside. Such attacks target either to disturb the service or even block it completely (Denial of Service / DoS). It is conceivable, that the RTP-user data stream is tried to be re-routed to a new destination and thus, access might be obtained to the content. On the other hand, an attack could also replace RTP user data streams by other ones in order to distribute wrong information which is not requested. Such attacks can not be defended with the means as described above, but need network structures and applications for security, such as firewalls and application layer gateways (ALG). Such structures and applications are without impact if not continuously controlled and supervised.

### 4.3 Operational security

Basically, there are two aspects of security for this type of connections:

- **The security of the transported information**
- **The security against attacks from outside**

For a SIP-based connection it is necessary to exchange information via two different channels. Registration and signalling information on one channel and user data on the other (both via RTP). Both connections could contain secret data. In order to protect the SIP signalling, the SIP standard (RFC 3261) defines the SIPS protocol which can use keywords to show the need of encryption. Such a protection is applied with the TLS procedure, also known from HTTPS. In order to guarantee a transparency of the information, only the body of such information – normally the information about the used audio and video code, etc. – is protected. When transmitting media

---

**SIP - NAT/Firewalls**

The “Firewall Problem”

![Diagram of SIP - NAT/Firewalls](image)

As a solution there is a STUN (Simple Traversal of UDP over NATs) protocol available. It lets the client find out what kind of NAT is used in the network and which addresses and port numbers must be used. To reach this there is an exchange of some test packets with some STUN-Server (e.g. stunserver.org) must take place.

### 4.4 SIP for high quality media links

Until now, SIP is mainly used within telephone applications. In such VoIP applications it is very well positioned and planned to be the standard for the Next Generation Network (NGN). We can count on SIP as the VoIP solution being a prerequisite for smooth operations with the appropriate network structures and resources. As described earlier, SIP is also a good basis for media links with professional quality. The very well defined separation of hand shaking and successive user data transmission allows the use of any audio and video codec as long as they are supported by the participating devices. SDP, which is used for the audio and video codec description is already part of transmission mechanisms and offered by serious codec vendors. SIP is simply a new signalling method.
5 Practical considerations for SIP and audio-via-IP

5.1 Requirements for SIP based connections via the public Internet

The first part of a successful SIP testing / integration is the selection of an appropriate IP-connection. In Germany for example there are quite a lot of regional and big service providers that offer S-DSL links with useful capacities starting at 2 Mbit/s. When getting in contact with a service provider in your country please take the following points into consideration:

- Ask about prices for an S-DSL link
- Capacity: minimum of 2 Mbit/s
- Be sure to get a minimum of 3 public IP-addresses (recommended: 16 public IP-addresses)
- Ask about quality of service (QoS) in the providers network

All SIP-clients need to register at a specific SIP-registrar. This can be a device in the same network or at any IP-address that can be reached over the public internet. The SIP-registrar is the key to SIP connections as it may be regarded as a kind of phone book or database with all available clients.

If you are interested to do SIP testing at the moment, it is recommended to acquire your own SIP registrar. There are many different models available on the market, such as the LANCOM 1722 VoIP for example.

Then if you start working with SIP you only need to register as a “SIP user”. All other settings are directly configured at the client (e.g. CENTAURI II). Please note, it’s possible to use public SIP-registrars instead of using your own. However you always need to check if all the required features are supported by a public service.

5.2 Requirements for other IP based connections via the public Internet – Unicast vs. Multicast

IP point to point connections are called unicast. Multicast is used to establish point to multipoint IP transmissions. Unicast can use UDP and TCP as transport protocol. Using CENTAURI II and other MAYAH Audio codecs UDP unicast connections can be uni- and bidirectional. CENTAURI II TCP connections are...
Each specific Multicast data stream is defined by a Multicast group ID. If a host wants to join a certain Multicast group for sending or receiving specific data, it has to inform his immediately-neighbouring router by sending the specific Multicast ID via an IGMP-telegram. This ID will then be forwarded from router to router for to see if there are other Multicast group members (see Figure 9: Request for Multicast).

If a host wants to join a certain Multicast group for sending or receiving specific data, it has to inform his immediately-neighbouring router by sending the specific Multicast ID via an IGMP-telegram. This ID will then be forwarded from router to router for to see if there are other Multicast group members.

On IP-level there is a specific range of class-D addresses reserved for Multicast. The four significant bits of the class-D addresses are set to 1 1 1 0. The 28-bit number following these four bits is called „multicast group ID“ which spans from 224.0.0.0 to 239.255.255.255.
6 Typical IP or mixed IP/ISDN applications

Fig. 10 Transmission from the reporter to multiple studios with IP Multicast via UMTS/3G

Fig. 11 Transmission from the reporter to multiple studios with IP Multicast via a WLAN hotspot access

Fig. 12 Backup of an IP -via-Satellite-connection via ISDN with automatic recovery to IP

Fig. 13 Point-to-Multipoint Backup to up to 8 remote sites with one CENTAURI II and built-in 4 BRI

Fig. 14 8 audio channels are transported via AES/EBU Transparent Mode via IP with a bit rate of 12
7 References


7.2 G. Stoll, F. Kozamernik, “EBU subjective listening tests on low-bitrate audio codecs”, EBU, 2003

7.3 “Too many audio codecs”, Detlef Wiese, Tonmeistertagung, 2002

7.4 White Paper “Discussion on coding algorithms” Detlef Wiese, 2002

7.5 White Paper “NAT Traversal for Multimedia over IP”, Newport Networks, 2005

7.6 White Paper „Understanding SIP“, D. Sisale, J. Kutan, GMD Fokus, 2000

7.7 CENTAURI II Manual, Werner Ludwig, 2006
The MAYAH range –
Professional Audio & Video Encoding/Decoding, Recording,
Mixing and Transmission Products

CENTAURI®II
Multichannel Audio Gateway Codec

Sporty
Portable Reporter Codec

MERK II
Portable Audio Codec Mixer

IQ [iei]
Audio Video Encoder/Decoder

ganymed
IP Audio Encoder/Decoder

Flashman II
Portable Audio Recorder/Codec