EBU – TECH 3329

A Tutorial on Audio Contribution over IP

Source: N/ACIP

Geneva
May 2008

Page intentionally left blank. This document is paginated for recto-verso printing

Contents

A Tutorial on Audio Contribution over IP

EBU Committee	*First Issued*	*Revised*	*Re-issued*
NMC	2008		

Keywords: Audio Contribution, Internet Protocol, User Requirements, Interoperability.

Foreword

Audio over IP terminals are increasingly being used in radio operations for streaming of radio programmes over IP networks from remote sites or local offices into main studio centres. The IP networks used are invariably well-managed private networks with controlled quality of service. The Internet is increasingly also used for various cases of radio contribution, especially over longer distances. Radio correspondents will have the choice in their equipment to use either ISDN or the Internet via ADSL or other available IP networks to deliver their reports. ISDN services used by broadcasters will be closed down in some countries.

With very few exceptions, IP equipment from one manufacturer has until now not been compatible with another manufacturer's unit. Based on an initiative from German vendors and broadcasters, the EBU has worked to create a standard for interoperability in a project group, N/ACIP, Audio Contribution over IP. This standard, jointly developed by members of the EBU group and manufacturers, is published as EBU Doc Tech 3326-2007 and several manufacturers have already implemented the standard in parts. Some of the audio over IP units are still in a somewhat immature, more or less prototype stage but further development continues. A plugfest between nine manufacturers held in February 2008 proved that previously incompatible units can now connect according to the new standard.

The requirements for interoperability are based on the use of RTP over UDP for the audio session and SIP for signalling. The packet payload audio structure is defined by IETF RFC documents for commonly used audio formats in radio contribution. Four mandatory codec formats are specified: G.711, G.722, MPEG Layer II and linear PCM. Other audio formats can be used as well, within the EBU framework, if the encapsulation method is published as an IETF RFC document.

This document is based on the collective inputs from EBU members and manufacturers.

1. Audio over IP

1.1 Introduction

This document is intended to support EBU-members in the technology of audio over IP systems in audio contribution. It covers basic IP computer and network technologies and proposes suitable protocols for streaming audio over IP. Terminals for audio over IP are now common in radio operations for streaming radio programmes over IP networks. The units are used to create contribution circuits from remote sites or local offices into main studio centres.

The IP networks used are usually well-managed corporate networks with good Quality of Service

(QoS) and usually high bandwidth. Due to its availability, the Internet is also increasingly used for various cases of radio contribution, especially over longer distances. However, the use of high bit rates and reliable contribution transmissions over the Internet cannot be guaranteed. Correspondents have the choice in their equipment to use either ISDN or the Internet to deliver their reports. More than 20 manufacturers provide equipment for audio over IP applications.

1.2 Transfer types

Two different types of audio transfer can be used over IP networks:

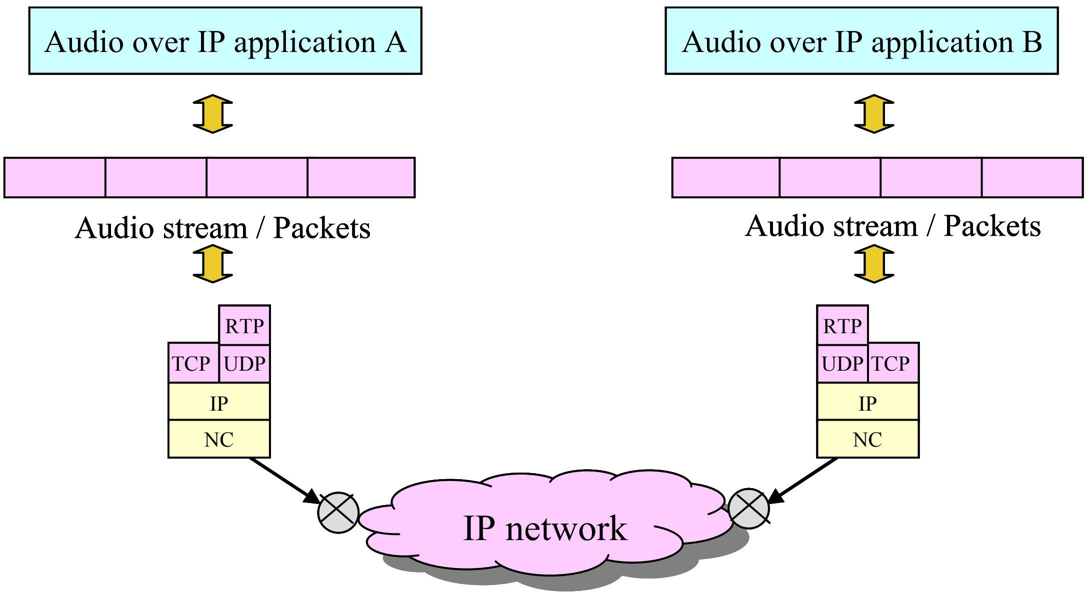
- File transfer
- Streaming audio

Audio over IP is used extensively for distribution, both for live web radio and for podcast, i.e. downloading of audio files to consumers. Private IP networks are now beginning to be commonly used in broadcasting for audio contribution, either with file transfers or as internal streaming.

1.2.1 File transfer

Security and good audio quality can be guaranteed through the use of FTP/TCP transfer of audio files. This method is widely used for file based operations in TV and radio stations. Audio File standards have been defined for use by the EBU. Broadcast Wave Format [2] and RF64 [3] are used worldwide for broadcasting and audio archiving. Limited network capacity may result in a slight delay of the delivery of a file, but gives no degradation of the audio quality in the original file. A small delay of seconds or minutes is often quite acceptable for pre-recorded material. File transfer is not treated in this document and has been studied by EBU group N/FT-AVC. An EBU guideline document: Tech 3319 is available.

1.2.2 Streaming



For live transmissions streaming must be used. Delivering live audio via packet switched networks, such as IP, can be challenging because excessive delays or loss of packets may cause interruptions in the audio. Streaming is the focus of this document.

1.3 EBU Initiatives

Up to now, solutions from manufacturers for audio contribution over IP have been incompatible with each other. Based on an initiative coming from German vendors and broadcasters, the European Broadcasting Union started a project group: N/ACIP, Audio Contribution over IP. One of its tasks was to suggest a framework for achieving interoperability. EBU members and manufacturers worked in cooperation to issue a standard for interoperability. This was published in 2007 [1]. More than 10 manufacturers have already begun to implement it, starting with the minimum interoperability option.

The interoperability specification is based on protocols used for Voice over IP with RTP over UDP for the transport of audio and SIP for signalling. The packet payload format already existed in IETF specifications (RFC) for commonly used audio formats in radio contribution, such as G.722, MPEG Layer II and linear PCM.

Another task from N/ACIP project group is to give guidelines to broadcasters for the introduction of Audio contribution over IP.

Further information on the EBU N/ACIP can be obtained at http://www.ebu-acip.org.

2. Why IP?

2.1 Change of technology

The Internet Protocol (IP) is very widely spread in all computers today. It is used worldwide within the Internet and in IP-based corporate or private networks used by broadcasters.

IP is independent from the underlying data transmission technology. Many IP adaptations exist for the physical layers, such as Ethernet, ATM and SDH over copper, fibre or radio link.

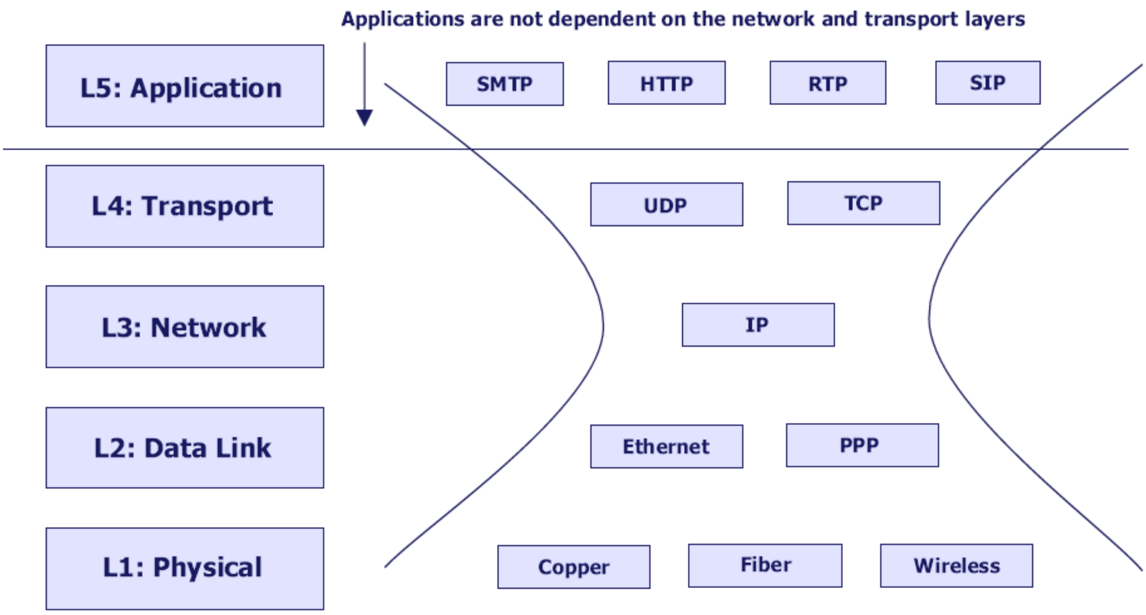


Figure 1

2.2 User benefits

Applications can communicate in a standardized way over interconnected IP networks. Host computers can be reached almost instantly wherever they are located. Fixed or temporary connections for contribution of audio can share similar types of applications. By using standardized protocols (SIP, RTP, UDP) used also for Voice over IP to transfer the streaming audio, radio programmes can be sent directly to another location just by dialling a number or an e-mail-like name. Many types of audio coding formats can be used at various bit rates. Higher bit rates will allow stereo audio or multi channel with linear PCM coding. The permitted maximum audio bit rate is dependent on the network and Quality of Service.

The Internet and also private IP networks are constantly being improved and are moving towards higher bandwidths, which will allow for the transfer of High Definition TV pictures and high quality audio to all consumers. In this rapid evolution, it will be possible to use both corporate IP-computer networks and the Internet for audio-only contribution and distribution. Costs for the networks will decrease and the Quality of Service will be improved. NGN (Next Generation Networks) is the broad term name for the improved networks to the public, with higher bandwidth. Triple play offers data, telephone and TV.

Mobile IP networks will be available with good population coverage in many countries. 3G/UMTS and LTE systems (4G) will soon offer higher bit rates upstream. WiMAX and WLAN hot spots are other suitable evolutions that may offer sufficient bit rates for radio contribution. However, quality of service is not guaranteed on these systems because of interference and shared access. The added value of these new networks will be near instant access for OB or news reporters in all urban areas. For example, the benefits of the use of SIP (Session Initiation Protocol), will be that the reporter can be easily reached almost anywhere in the world with one identity, irrespective of the equipment being used.

It is expected that audio over IP systems will be used more and more for live IP streaming. This will happen gradually within the next few years, when terminals will have reached maturity, standards will have been implemented and when the network transport capacity possibly has become cheaper. Dedicated hardware or software systems running in PCs will compete on the market. Cost trends are not easy to predict, but a more mature market with more units sold in a competitive market will probably lead to lower prices.

2.3 Quality of Service

The Quality of Service is usually very high within existing synchronous systems used for contribution, at fairly high cost. Generally, traditional systems based on SDH/PDH over fibre or microwave links with non-IP transfer systems offer only minutes of downtime per year. It will probably take some time to come to a similar level of uninterrupted service with the various IP network solutions currently in use. For example, physical redundancy (by using two independent terminals at each end and two network paths) must be considered in the most critical broadcast contribution circuits using well-managed private IP networks.

3. Protocols

3.1 Transport Protocols

The EBU has issued an interoperability standard [1]. It is based on the mandatory use of SIP for signalling and RTP over UDP for transfer of audio. TCP is optional. RTP (over UDP) has been designed particularly for live media transfer and offers an efficient method. The mechanism for segmenting the original audio format into IP packets is based on existing RFC's from the IETF. Interoperability between audio over IP terminals and IP telephones will be made possible via the SIP protocol for signalling.

3.1.1 UDP

UDP, User Datagram Protocol, is very simple and consists of sending information "as is" in packets, without congestion control, session establishment or acknowledgements. UDP is often known as "fire and forget". All functionality and error control mechanisms must be implemented in the upper application layers. There are some issues when passing through firewalls doing NAT (Network Address Translation). Special methods, such as STUN (Simple Traversal of UDP through NATs) can handle NAT, which is used to map multiple private IP-addresses to one public address.

3.1.2 TCP

TCP, Transmission Control Protocol, is a reliable protocol using acknowledgements/retransmissions and including a congestion avoidance mechanism. It needs bi-directional transmission. This protocol is not well suited to streaming because the retransmission mechanism leads to longer delays and the efficiency is lower. Congestion control may decrease the sending rate under the media codec bit rate, which may lead to sending buffer overrun and receiving buffer under-run. This may cause a connection break. On the other hand, TCP has the advantage of reliability and can pass more easily through firewalls than UDP, as it is connection oriented.

3.1.3 RTP

RTP (Real-time Transport Protocol) has been designed for the transport of multimedia streams over IP networks. It works on top of UDP. IP networks are asynchronous and may lose or reorder packets. Some fields in RTP are defined to address these problems:

PT (Payload Type) tags the content of the packet. Some standardized values exist, but for other codecs, the type must be described in a service description protocol, SDP (Session Description Protocol).

Sequence number is primarily used to identify and detect lost packets and secondly to reconstruct the order in which packets where sent, which may make loss detection easier.

The **timestamp** is the sampling instant for the first octet of media data in a packet. It can be used to help recover the clock frequency at the receiving side, if it is not given by other means.

RTP does not include a loss recovery mechanism in case of packet losses.

3.1.4 RTCP

RTCP (RTP Control Protocol) is used to send receiver reports back to the sender. Based on information on jitter, bandwidth or packet loss the sender can resend packets or adapt the transmission according to these messages, such as synchronisation of media streams, session membership and source description information.

3.1.5 UDP or TCP?

The EBU-standard suggests using RTP over UDP rather than TCP. A one-way RTP stream with a small header (low overhead) is more suitable for audio transfers. Moreover, RTP over UDP sometimes has higher priority than TCP in routers.

3.2 Signalling protocols

Signalling is necessary for "call" setup, termination and negotiation of parameters between terminals. SIP (Session Initiation Protocol), the popular protocol used in Voice over IP telephone systems, has also been chosen for the EBU specification to connect broadband audio.

This opens interoperability between IP telephones or mobile handheld devices using SIP, and broadcast audio over IP systems. During call establishment, SIP and SDP offer the possibility of automatic negotiation between terminals to find a common audio coding system.

3.2.1 SDP

SDP (Session Description Protocol - according to RFC 4566) is used to define the type of audio coding used in the RTP media stream. It is a text-based protocol, which can easily be parsed and monitored.

Example of SDP:

Description of a PCM 16 bit 48 kHz stereo session to host.anywhere.com:

```
v=0
o=alice 2890844526 2890844526 IN IP4 host.anywhere.com
s=
c=IN IP4 host.anywhere.com
t=0 0
m=audio 49232 RTP/AVP 98
a=rtpmap:98 L16/48000/2
```

The “m=” field indicates that it is an RTP stream with payload type 98. The “a=rtpmap” field gives audio coding format and parameters with this payload type.

SDP is merely a text format and no transport is specified. RTSP (Real-Time Streaming Protocol), SIP (Session Initiation Protocol) or any other transport mechanism can be used to transport SDP text.

3.2.2 SIP

SIP (Session Initiation Protocol) is a signalling protocol that is increasingly used in IP telephone systems, instead of H.323. It has many useful features that can also be used in radio contribution systems with higher audio quality. SIP carries SDP messages.

Port 5060 is the port used by SIP. Terminals listen to this port for incoming requests. SIP is active only for the setup, termination and control of a call. The other end is found through the use of DNS-servers or ENUM-databases. After the SIP session is initiated, SIP stays idle and the separate streaming audio transfer media session is started.

A useful feature with SIP in combination with SDP is the negotiation of audio codec to be used for transmission during call establishment. Every terminal advertises the audio coding it understands and a common format is then chosen.

SIP sessions can happen directly between terminals or via intermediate systems like SIP gateways or proxies. The types of SIP elements within networks are

- User Agent (UA) , a codec or a telephone
- Proxy server (often called either SIP server or SIP proxy)
- Location server
- Registrar server
- Redirect server

Different elements may use common hardware.

A key SIP feature is '*presence*', where an terminal can register to a location server which holds information about IP addresses and other parameters. All calls can then be made via this server without knowing the IP address from the correspondent. This feature, in combination with fixed or handheld SIP devices, will allow the newsroom people to quickly find their reporters.

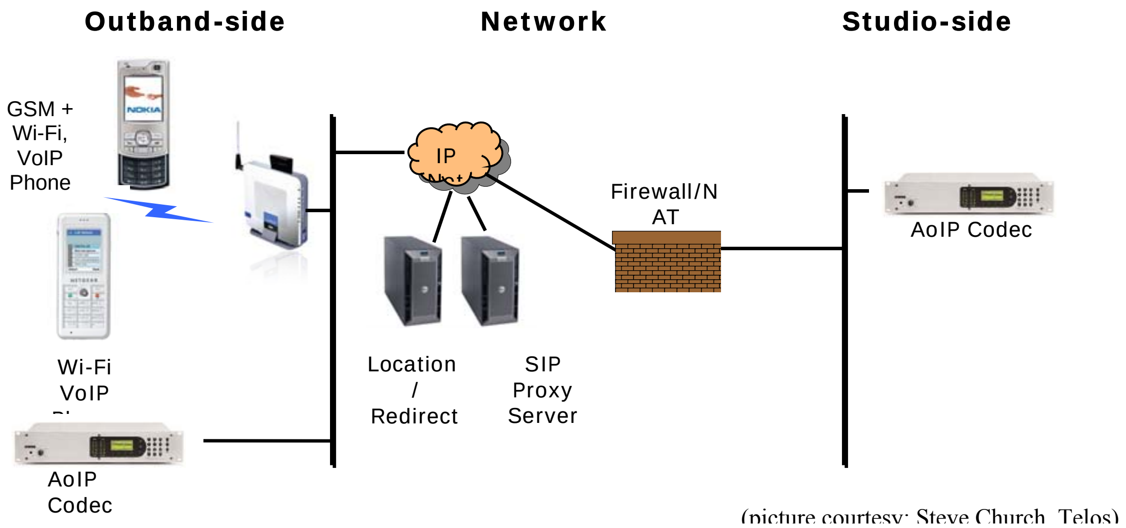


Figure 2: Typical infrastructure using SIP

3.2.3 SAP

SAP (Session Announcement Protocol) is a simple unidirectional protocol for the transport of SDP messages. Messages are regularly broadcast. It is useful in the case of an absence of return channel (example: satellite link) or with multicast sessions.

3.2.4 RTSP

RTSP (Real Time Streaming Protocol) is used for establishing and controlling a single or several media streams. It provides a remote control style functionality of the stream and session. It is used for instance in some IPTV systems.

It is not used in the N/ACIP specification for interoperability, as SIP has been chosen.

4. Audio coding algorithms

4.1 Mandatory audio codecs

Four mandatory audio coding formats have been suggested for interoperability when using contribution over IP by the EBU. These formats must be present in devices that are compliant with the specification. The aim is to have a minimum common base. These four coding formats have deliberately been chosen to be cheap and not complex to implement.

- ITU G.711
- ITU G.722
- MPEG Layer II
- PCM (optional for portable units)

4.1.1 ITU G.711

G.711 can be found in almost all voice over IP systems. The audio bit rate is 64 kbit/s. Two compression profiles exist: µ-law, used in the U.S. and Japan, and A-law for Europe and rest of the world. 20ms of audio per RTP packet should be used as the default, for improved compatibility with Voice-over-IP systems, although other values may be used.

4.1.2 ITU G.722

G.722 is very common and easy to implement despite a limited audio bandwidth. The audio bit rate is 64 kbit/s. 20ms of audio per RTP packet should be used as the default, although other values may be used.

4.1.3 ISO MPEG-1/2 Layer II

The MPEG-1/2 Layer II audio format is low complexity and has low patent costs. It provides high audio quality at medium/high bit rates and has the option to transmit ancillary data. The range of possible bit rates and sampling frequencies is large. EBU N/ACIP decided to specify mandatory/recommended bit rates to be found in audio over IP equipment. The following table gives mandatory bit rates in bold and others as recommended. Manufacturers can optionally implement other bit rates.

The RTP payload format used is the one for MPEG elementary or transport streams. It is very common and is used for some IPTV systems.

The MPEG parameters are also signalled in the SDP text so intermediate systems can get all the information about the stream.

Bitrate [kbit/s]	Sampling rate			
	16 kHz*	**24 kHz***	**32 kHz**	**48 kHz**
32	M			
40	M			
48	M			
56	M	M	M	M
64	M	M	M	M
80		M	M	M
96		M	M	M
112		M	M, JS, S	M, JS, S
128		M	M, JS, S	M, JS, S
160			M, JS, S	M, JS, S
192†			M, JS, S	M, JS, S
224			S	S
256†			S	S
320			frame too large	S
384†			frame too large	S

*MPEG-2 †OPTIONAL for portable equipment

Legend: M = Mono, JS = Joint-Stereo, S = Stereo

4.1.4 PCM

PCM (linear audio) has no patent costs, is of low complexity and is not subject to degradation in the case of audio coding cascading. Quantisation of 16, 20 and 24 bits are specified. It should be noted that this is raw transport. User bits and other information found in the AES/EBU format are not transported. No RTP payload format exists for AES/EBU audio frames, although this could be defined in the future. Implementation of linear audio format is optional for portable units.

4.2 Recommended audio codecs

In addition to the mandatory codecs, the interoperability standard EBU Doc Tech 3326 can be used as a common framework for other audio coding formats. Some, recommended by EBU N/ACIP, which are commonly used, are listed below.

Manufacturers are free to implement them or not, even if they are recommended. If they do implement them, they should follow the specification below in order to stay compliant.

4.2.1 ISO MPEG-1/2 Layer III

(ISO 11172-3 MPEG-1 Layer III and ISO/IEC 13818-3 MPEG-2 Layer III).

In order to reduce the number of possibilities the following table shows the recommended implementation. Bit rates and sampling rates in bold are mandatory.

Bitrate [kbit/s]	**Sampling rate**			
	16 kHz*	**24 kHz***	**32 kHz**	**48 kHz**
32	M	M		
40	M	M		
48	M	M	M	M
56	M	M	M	M
64	M	M	M	M
80		M	M	M
96		M	M	M
112		M	M, JS, S	M, JS, S
128		M	M, JS, S	M, JS, S
160			M, JS, S	M, JS, S
192†			M, JS, S	M, JS, S
224			S	S
256†			S	S
320			Large frame	S

*MPEG-2 †OPTIONAL for portable equipment

Legend: M = Mono, JS = Joint-Stereo, S = Stereo

Two payload formats can be used: either the generic MPEG format (RFC2250) or a format which has specifically been designed for robust MPEG Layer III transport (RFC3119). The reason is that MPEG Layer III frames are not independent, so in the case of a single packet loss, there is an increased degradation effect. The robust transport mode is intended to avoid this situation by gathering dependant frames into a single or common group packet.

These transport modes are signalled differently in SDP.

4.2.2 MPEG-4 AAC, MPEG-4 AAC-LD

(ISO/IEC 14496-3 MPEG-4 AAC Low Complexity Profile, MPEG-4 AAC-LD).

An RTP payload format for MPEG-4 elementary streams exists (RFC3640). It covers all MPEG-4 AAC flavours. (AAC-LC, AAC-LD,…). In order to decrease the overhead no frame signalling is used (ADTS or LATM). This means that all parameters must be signalled in the SDP in order for the decoder to get the necessary parameters.

4.3 Optional audio codecs

Further specifications are made for other audio coding formats. The rule in the EBU specification is that audio formats must have an RTP payload format defined and registered at IETF. Every codec manufacturer is welcome to submit a payload format to IETF AVT group that will then be published as an RFC. This is important for interoperability.

4.3.1 Enhanced APT-X

This is an ADPCM format from APT corporation (www.aptx.com)

An RTP payload format is in the process of being standardised at IETF.

This format has the advantage of low coding delay. It has been used for 5.1 multi channel audio transfers. MIME Subtype names: 'aptX', 'EaptX 16', 'EaptX 24', 'aptXLive'

4.3.2 MPEG-4 HE-AACv2

RTP encapsulation according to RFC 3640 should be used (same as for other MPEG-4 AAC formats).

4.3.3 Dolby AC-3

RTP encapsulation according to RFC 4184 should be used.

4.3.4 AMR-WB+

Extended Adaptive Multi-Rate Wideband is defined in 3GPP TS 26.290. 3GPP originally developed the AMR-WB+ audio codec for streaming and messaging services in GSM and 3G cellular systems. The codec is designed as an audio extension of the AMR-WB (G.722.2) speech codec. The implementation of these codecs will open possibilities for future interoperability with mobile communication networks and higher audio quality to mobile phones.

If AMR-WB+ is implemented, RTP encapsulation according to RFC 4352 should be used

If AMR-WB (G722.2) is implemented, RTP encapsulation according to RFC 3267 should be used.

5. IP Networks

It is important to make the distinction between managed IP networks and the Internet. The Internet is a public unmanaged IP network and has no guaranteed Quality of Service (QoS).

Telecom operators offer an increasing number of IP-based services. Traditional services based on ATM, PSTN, ISDN, SDH and PDH will gradually be phased out or will become expensive niche products. They tend to be replaced by all-over-IP services over copper, fibre or microwave links. These services concern not only data communication, but also voice telephony and TV, based on IP networks with managed Quality of Service.

5.1 Last mile access to the end user

The following last mile access methods will probably be used for Audio over IP:

- Fibre optics: High quality, but still a bit expensive.
- Copper cables with either ADSL or SDSL: Good performances at reasonable prices, widespread.
- Mobile (3G/UMTS, WiMAX): Widespread but unfortunately no solution with managed quality of service so far. Long delays.
- Satellite: Some solutions exist, with managed quality of service emerging, but they are still expensive. Long delays.
- Wireless (WiFi /WLAN): Frequencies are shared, so interference problems may happen.

5.2 Network impairments

5.2.1 Packet loss

This can be caused by network congestion (buffer overruns in routers), but also by transmission line errors. UDP and TCP transport protocols use a checksum to verify the integrity of each packet. In the case of a single bit error, the entire packet is discarded. Available network bit rates will vary on shared, consecutive IP networks and congestion may occur. This can lead to bandwidth reduction and bursts of lost packets.

5.2.2 Packet delay

This is the latency due to the propagation time and buffering. Sudden changes in route lengths, caused by dynamic routing, if used, will cause rapid changes in delay values.

5.2.3 Packet delay variation / jitter

The original stream of audio in broadcaster's systems is synchronous and a continuous stream is sent generally with equal spacing between the audio frames. IP networks are asynchronous with packets arriving with an uneven spacing in time. This is known as packet delay variation or jitter in the time domain. This artefact must be compensated for with a buffer to store packets at the receiving end, which is then read out from the buffer with correct timing at play out. If the amount of jitter is smaller than the buffer size, the varying arrival time of packets will be smoothed out by the receiver. A synchronous stream of audio can thus be regenerated. If the jitter is too high, bits or packets will be lost, resulting in audible distortion.

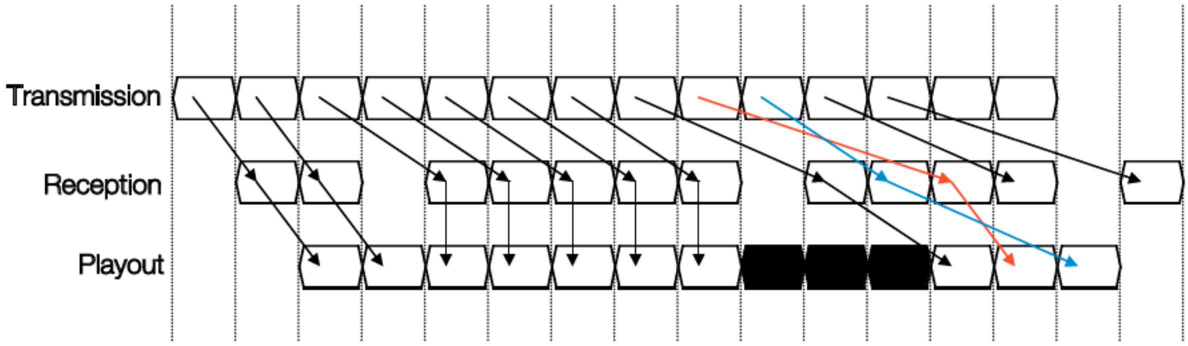


Figure 3 (courtesy: Greg Massey, APT)

5.2.4 Packet disorder

Packet disorder happens in dynamic routing networks. When a route change occurs, some packets may arrive at the receiving end in a different order than that in which they were sent because of different path lengths. The arrows crossing each other in the Figure above illustrate a shift in packet order, which has been corrected at play out.

The impact of these impairments depends on how well the networks are managed regarding routing, transmission lines and sharing. The Internet is an unmanaged public network, consisting of an interconnection of different networks types or sections, with dynamic routing, using many different transmission lines.

Some of the impairments are negligible on private IP networks and point to point IP links, particularly when audio over IP packets are not interleaved with other data packets.

5.2.5 Packet fragmentation

Frames that exceed the maximum transmission unit of about 1500 bytes must be fragmented into several RTP packets before transmission. Fragmentation can be made by the application at the RTP layer, or by the network using IP fragmentation. Fragmentation should be avoided because it increases network stress and single packet loss results in discarding the content from all other fragments, amplifying the loss effect.

5.3 Network management

5.3.1 Profiling according to ITU-T Rec Y.1541

IP network impairments can be partially controlled using Quality of Service mechanisms (QoS). Recently the ITU has issued a recommendation on network profiles defining classes with bounds on packet loss probability, delay and jitter. This recommendation (Y.1541) [5] should help operators to agree on the class for different services and manufacturers to offer equipment scaled to these defined classes. See also Y.1221 'Traffic control and congestion control in IP based networks'

5.3.2 Quality of Service

Some methods can be used in order to control corporate IP network impairments. Two main approaches for the transfer of Quality of Service parameters (QoS) exist:

- "IntServ" (Integrated Services) is considered a fine-grained flow-based mechanism. The principle is to reserve network resources by external signalling to all machines (routers) capable of supporting QoS. This is the approach of RSVP (Resource ReSerVation Protocol) defined in RFC2205.
- "DiffServ" (Differentiated Services) is considered a course-grained class-based mechanism and uses tagging of IP packets. Different priority policies in routers are used depending on the tag. The per-hop behaviour is indicated by a 6-bit value (or tag) – called the DSCP (Differentiated Services Code Point) encoded into the 8-bit DS (Differentiated Services), also known as the TOS (Type of Service) field of the IP header.

The distinction between well managed IP networks and the Internet is important. For the Internet, no mechanisms to achieve a good QoS yet exist. The Internet is a "best effort" network with no guaranteed Quality of Service at all. Over a ten-year period the packet loss, delay and jitter over the Internet have slowly improved, but the network performance still poses a major problem to the developers of audio over IP units.

5.3.3 Unicast and Multicast

Two main schemes for media delivery exist within IP:

- **Unicast**: A one to one delivery between end points. It is the most used delivery mode and is simple. Unicast may become expensive where there is a need to deliver the same content at the same time to a large number of receivers, as the number of streams is equal to the number of clients.
- **Multicast**: For one-to-many or many-to-many delivery between end points. Multicast is easy to implement within the bounds of a dedicated corporate network. It requires special signalling mechanisms and routing algorithms in order to control the delivery. Additionally, it requires special support and configuration of the routers. For multicast, the following protocols can be used:
    - IGMP (Internet Group Membership Protocol) - enables hosts to dynamically join/leave multicast groups; membership information is communicated to the nearest router (which must support multicast).
    - Multicast Routing Protocols, such as DVMRP (Distance Vector Multicast Routing Protocol), MOSPF (Multicast extension to OSPF) & PIM (Protocol Independent Multicast), - enable routers to build a delivery tree between senders and receivers of a multicast group for unicast delivery; routing can be fixed or dynamic (choosing the best route or changing the route in case of link disruption).

Multicast is generally well developed technically and introduced in isolated islands of routers and private managed IP networks. However, no good business models yet exist for using multicast over the Internet. Internet Service Providers in general avoid the peering transfer of multicast protocols between them, since it is not considered to be profitable. General use of IP-TV may change this situation.

6. Synchronisation

Variations in the delivery time of packets occur, mainly due to varying delays in routers and the sharing of available capacity with other data traffic. Buffering is required to compensate this variation. There is no clock transported at IP level and so it must be reconstructed at the receiving end. Many different clock recovery algorithms exist. The difficulty is to estimate the slow clock drift correctly and separate it from the short-time network jitter. Streaming audio with high quality is dependant on a guaranteed and stable clock rate at both the sending and the receiving end. Another possibility is to use an external clock source, such as GPS or from a common non-IP network clock.

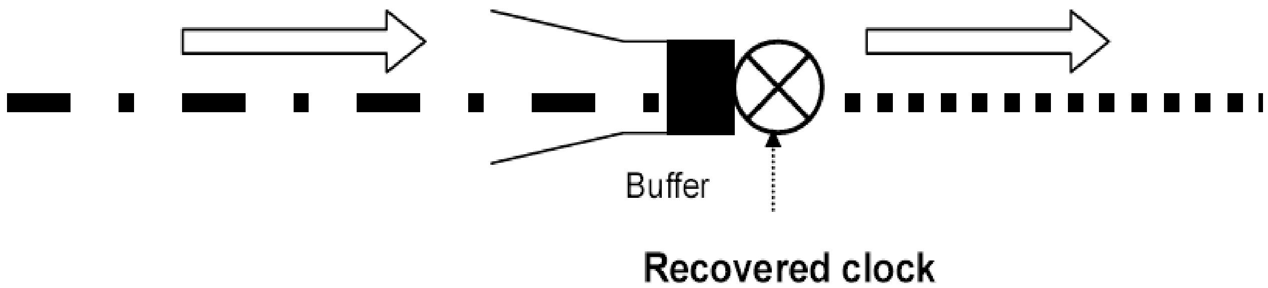


Figure 4

7. Operational practices

7.1 Operational classes

Audio over IP contribution can be divided into the following classes:

- Outside broadcasts (concerts, sports, news)
- Interviews (two persons talking on two separate locations, duplex)
- Discussions, Talk shows, Roundtable (many persons, multiple locations)
- Multicast, Point-to-Multipoint for internal distribution of programmes.
- Non-live / non-real-time broadcasts (using file transfer instead of streaming audio)

Different types of audio contribution modes in broadcasting can be identified:

- Unidirectional with no return channel (example: contribution by satellite).
- Bidirectional, where the return audio is voice quality non-broadcast narrowband and for the purposes of cueing the contribution (examples: concert, football commentary). Latency is not a major problem.
- Bidirectional, with bidirectional broadcast quality audio, e.g. interview, discussions, etc. Latency is critical and has to be considered. Clean Feed/ "N-1" type of operations may solve some of this problem.

The maximum allowed delay is dependent on the application. It must be lower for bidirectional conversation programmes than for a unidirectional transmission, which only has to be cued to start.

7.2 Connection Types

Two types of connections can be used:

- Permanent connections, which are generally based on managed private networks with constant and well-known bandwidth and Quality of Service. For permanent connections, audio codec types are usually known in advance.
- Temporary connections, which may be based on previously unknown networks with shared and unknown bandwidth over the Internet or over temporary leased private networks. Codecs and endpoint may be unknown. Audio codec type can be found through negotiation using the SIP and SDP protocols according to the EBU interoperability specification. Available bandwidth can be measured prior to the transmission in order to set parameters.

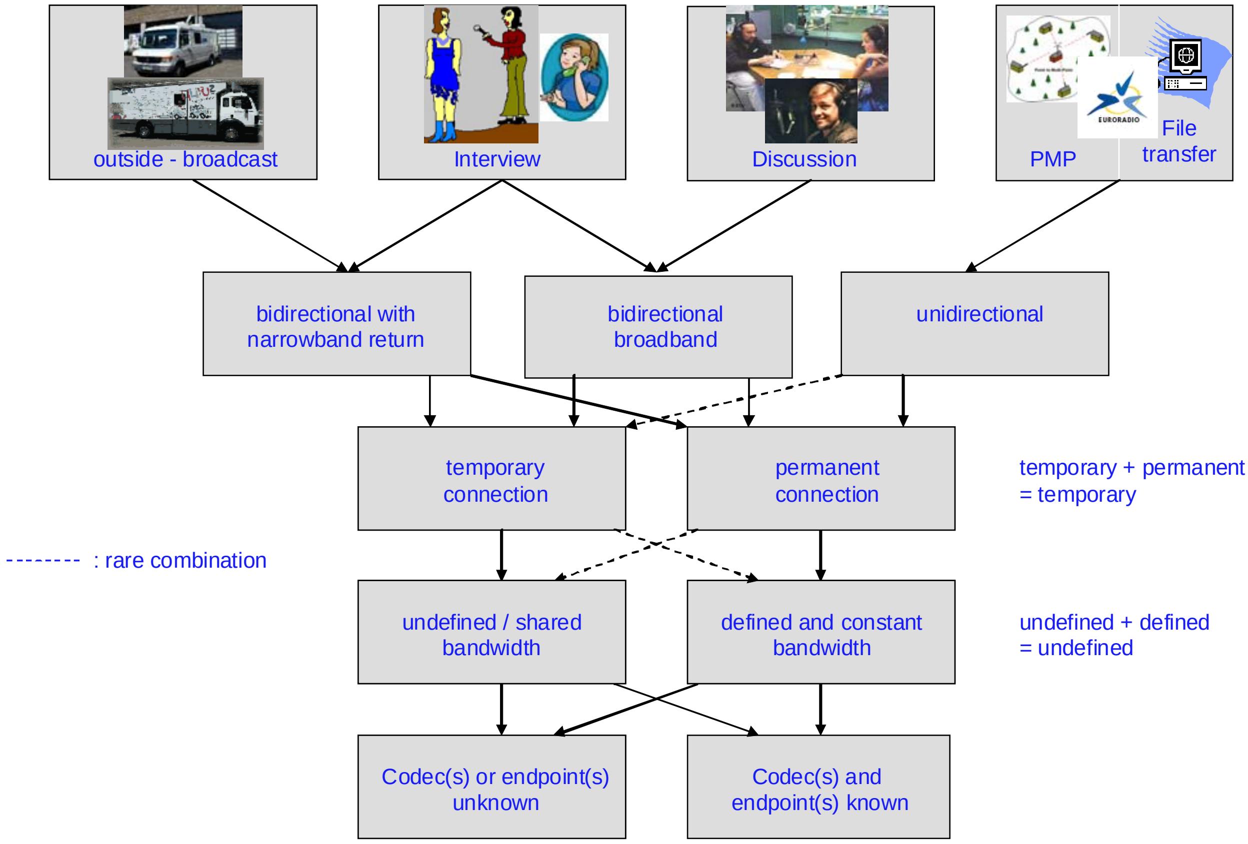


Figure 5: illustration of some of the different types of operations for audio over IP

7.3 Delay

Buffers at the receiving end can introduce a considerable amount of delay. The delay buffer size is a trade-off between an acceptable delay and a reliable transmission. In addition, the IP network itself has a delay, from a few tenths of milliseconds in well managed networks up to 500ms or more on very long distances over the Internet or satellite links. The audio encoding itself introduces delays from milliseconds for PCM to more than hundreds of milliseconds for some bit-rate reduced coding formats. In the case of a two-way conversation, the total round trip delay should be kept as low as possible, because otherwise a conversation becomes difficult, especially when non-experienced reporters or the general public are interviewed. When using audio over IP in combination with video contribution, lip sync will be an issue.

Elements that influence the end to end latency can be summarised thus:

- Audio encoding/decoding delay (~ 1- 500ms)
- Audio encoding algorithm + implementation
- Packetisation delay (~ 1 – 100ms or more)
- Trade-off between delay and packet size/overhead/packet rate (network stress)
- Network latency (~ 1 – 500ms)
- Due to distance, buffering, packet routing, (tunnelling)
- Receive buffer delay (~ 10 – 500ms or more)

For these reasons, delay is generally longer with Audio contribution over IP than with traditional contribution based on synchronous networks. It is possible to reduce delay to minimum by using high quality IP networks with very low jitter, low delay audio coding algorithms and short packets.

7.4 Network availability and quality

Computer systems in general may have other demands for immediate downtime than an audio over IP link with millions of listeners. The computer and IP network industry will probably gradually adapt to this higher demand for high reliability and low delay. For audio file transfers, a retransmission mechanism usually takes care of all lost packets in IP networks. This is not the case for lost packets in streaming sessions of audio. Hence, systems for Forward Error Correction and/or Error Concealment, with a minimum of added delay, must be developed further for use with streaming audio over IP.

7.5 Security

Methods to protect systems from unauthorized intrusion are needed. Encryption of the content may be used, but this leads to additional packet delay, which is not desirable and so it is not usually used.

The audio traffic and control/management systems in audio over IP may use several different TCP/UDP-ports. When operated through a firewall, certain ports have to be opened, which creates an unwanted security hole. Therefore it is desirable that these different port numbers are limited in number. Firewalls and other security devices should be aware of the protocols used and act accordingly. Web interfaces must be well protected.

7.6 Controlling units via SNMP

Most of the audio over IP units can be supervised and controlled via SNMP (Simple Network Management Protocol). This feature is recommended for broadcasters, since basic systems for router and network control usually already exist in Master Control areas.

EBU Project Group, N/CNCS (Common Network Control Strategy) is working towards a common MIB (Management Information Base) either in whole or in part, which can be incorporated into manufacturer's products. It is being standardised in a parallel activity through the IEC (http://www.iec.ch).

Originally developed for audio over ATM in radio broadcasting, the control framework has been extended to encompass video and other time-critical media, other networking technologies and other applications in both professional and consumer environments.

There are currently six parts in the standard and they are:

1. General
2. Audio
3. Video
4. Data
5. Transmission over Networks
6. Packet Transfer Service

Part 1 specifies aspects which are common to all equipment and has been published as an International Standard [6] [7]

Parts 2 to 4 specify control of internal functions specific to equipment carrying particular types of media; in the case of Part 4 this would be time-critical media other than audio and video, e.g. RS232 and RS422 application control data. Part 4 does not refer to packet data such as the control messages themselves. At the time of writing, Part 2 (Audio) has been submitted for and is undergoing its standardisation process [8].

Part 5 specifies control of transmission of these media over each individual network technology, with a separate sub-part for each technology. It includes network specific management interfaces along with network specific control elements that integrate into the control framework. Part 5-1 covers aspects common to all network technologies. Parts 5-2 and onwards will cover other networking technologies such as IP (Internet Protocol), ATM (Asynchronous Transfer Mode) IEEE 1394 (FireWire), etc.

Part 6 specifies carriage of control and status messages and non-audiovisual data over transports that do not support audio and video, with (as with Part 5) a separate sub-part for each technology. Initially there is just one sub-part, 6-1, which covers RS232 serial links.

Originally proposed was a Part 7 specifying control of different types of broadcast transmission equipment, containing provisions common to all transmitters and another four covering DVB-T, DAB, FM radio, and DRM (Digital Radio Mondiale) respectively.

An associated document dealing with the control of transmitter is available from EBU Tech 3323

8. Conclusion

The continuous development of IP networks combined with more sophisticated audio over IP terminals will lead to more use of this technology in the future. EBU group N/ACIP has made a proposal for interoperability. Connections over the Internet with different types of telephony and professional units for broadcasting will improve telephone audio quality and worldwide access to reporters. Small handheld units and also software codecs in laptops or mobile phones will provide very efficient tools for reporters. SIP will provide a very powerful way of finding the other end and negotiating a suitable audio coding format. Fixed audio over IP units will begin to replace older synchronous point to point equipment for contribution of stereo or multichannel 5.1 audio.

9. ACKNOWLEDGEMENT

This document has been created with submissions from EBU-members and manufacturers working in EBU project group N/ACIP.

10. References

[1] EBU Doc Tech 3326: Audio contribution over IP Requirements for Interoperability

[2] EBU Doc Tech 3285-2001: Broadcast Wave Format

[3] EBU Doc Tech 3306-2007: RF64: An extended File Format for multichannel audio

[4] AES Convention Paper 118th Barcelona: New File Format and Methods for Multichannel Sound in Broadcasting /Jonsson, Holzinger

[5] ITU-T Rec Y.1541

[6] IEC62379-1: Common Control Interface for Networked Digital Audio and Video Products– Part 1

[7] IEC Project Team 62379 web site (http://www.iec62379.org/)

[8] IEC62379-2: Common Control Interface for Networked Digital Audio and Video Products – Part 2: Audio.

11. Glossary

3GPP	3rd Generation partnership project
AAC	Advanced Audio Coding **ADTS** (or **LATM**) are flavours of AAC signalling
ADSL	Asynchronous Digital Subscriber Line
AMR-WB	Adaptive Multi Rate - WideBand (G.722.2)
AoIP	Audio over IP (broadband audio)
ATM	Asynchronous Transfer Mode
BGP	Border Gateway Protocol
CSRC	Contribution Source (in RTP)
DRM	Digital radio Mondial
DSCP	Differentiated Services Code Point
DVMRP	Distance vector Multicast Routing Protocol
DVB-T	Digital Video Broadcasting - Terrestrial
FEC	Forward Error Correction
FTP	File Transfer Protocol
GPS	Global Positioning System
HTTP	HyperText Transfer Protocol
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IP	Internet Protocol
IMS	IP Multimedia Subsystem
ISDN	Integrated Services Digital Network
LTE	Long Term Evolution (4G mobile system)
MIB	Management Information Base
MIME	Multipurpose Internet Mail Extensions
MOSPF	Multicast extension to OSPF
MPEG	Motion Picture Experts Group
NAT	Network Address Translation
NGN	Next Generation Networks
OSPF	Open Shortest Path First
PCM	Pulse Coded Modulation
PDH	Plesiochronous Digital Hierarchy
PIM	Protocol Independent Multicast
PPP	Point-Point Protocol
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RFC	Request For Comments (IETF standard)
RSVP	Resource ReSerVation Protocol
RTP	Real time Transport Protocol
RTCP	Real time Control Protocol

RTSP	Real time Streaming Protocol
SAP	Session Announcement Protocol
SDH	Synchronous Data Link Hierarchy
SDP	Session Description Protocol
SDLC	Symmetric Digital Subscription Line
SIP	Session Initiation Protocol
SMTP	Simple Mail Transfer Protocol
SNMP	Simple Network Management Protocol
STUN	Simple Traversal of UDP through NATs
TCP	Transmission Control Protocol
TOS	Type of Service
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunication System
VoIP	Voice over IP (narrowband audio)
WLAN	Wireless LAN IEEE 802.11(a,b & g)
WiMAX	Worldwide Interoperability for Microwave Access IEEE 802.16