EBU

OPERATING EUROVISION AND EURORADIO

TECH 3326

AUDIO CONTRIBUTION OVER IP

REQUIREMENTS FOR INTEROPERABILITY

SOURCE: FNS-ACIP

STATUS: TECH SPEC. REV. 4

Geneva
November 2014

Page intentionally left blank. This document is paginated for two sided printing

Scope

This document sets out a minimum set of requirements necessary to ensure interoperability between equipment intended for the transport of contribution-quality audio over IP networks.

In this document the following bold, uppercase words have special meanings.

MUST and **SHALL** identify mandatory elements that need to be implemented or followed in order to achieve interoperability.

SHOULD and **RECOMMENDED** identify elements that are not mandatory, but whose implementation is advisable.

MAY and **OPTIONAL** identify facultative elements which, if present, SHOULD be implemented as specified for better interoperability with other equipment implementing the same elements.

Feedback on this document is invited; it should be sent to Mathias Coinchon (coinchon@ebu.ch), project manager of the EBU FNS-ACIP group.

Page intentionally left blank. This document is paginated for two sided printing

Contents

Audio contribution over IP
Requirements for Interoperability

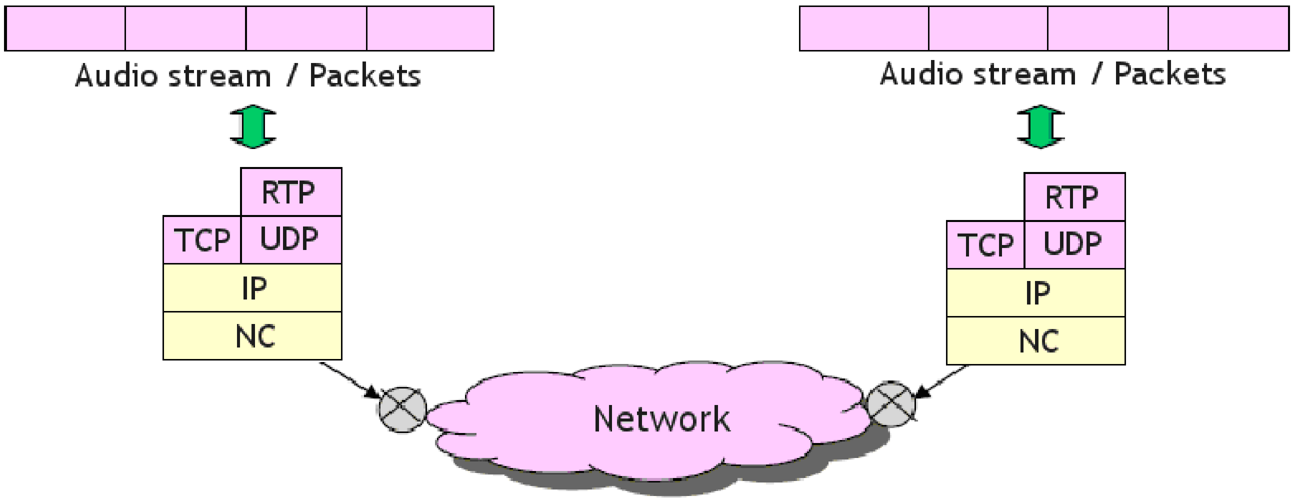
EBU Committee	First Issued	Revised	Re-issued
TC	2007	Aug. 2014	

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

1. Audio over IP Networks

1.1 Introduction

Increasingly, broadcasters are using IP connections for the purposes of streaming high-quality broadband audio to their production centres. This is in part accounted for by the fact that several countries are withdrawing ISDN services, which have been heavily used for contribution in the past.



About 15 to 20 manufacturers currently provide units capable of transferring audio over both ISDN and IP connections, and efforts must be made to achieve interoperability between units from different manufacturers.

1.2 Types of audio contribution

Different types of audio contribution in broadcasting can be identified:

- Unidirectional with no return channel (example: contribution by satellite).
- Bidirectional where the return audio is narrowband and for the purposes of cueing the contribution (examples: concert, football commentary). Latency is not an issue.
- Bidirectional with bidirectional broadband audio (examples: interview, discussion). Latency

is an issue.

1.3 Types of equipment

Different types of equipment can be identified:

- General contribution equipment: Equipment meant for all type of contribution (fixed or remote).
- Portable contribution equipment: Equipment meant mainly for monophonic speech contribution at low bitrates.

The complexity and processing power of portable equipment is less than that of general contribution equipment and the corresponding requirements for interoperability are less stringent, and are noted where applicable in this document.

1.4 Where are 'Requirements' needed?

The requirements necessary to achieve interoperability between Audio over IP contribution devices concern:

- **Transport protocols** to be used on top of IP, including port definition and packet loss recovery mechanisms.
- Audio coding algorithms to be implemented.
- **Audio frame encapsulation**: definition of framing and encapsulation of audio frames into transport layer frames.
- **Signalling**: defines connection setup and termination procedure, signals parameters for the receiver (audio coding used, etc.). Unidirectional signalling is also considered.

For each of these areas this document defines the minimum requirements to achieve interoperability between devices. Best practices and other considerations not directly linked to interoperability are published in EBU Tech 3329.

Note: If SNMP (Simple Network Management Protocol) is used, IEC 62379 should be applied.

2. Transport protocols

2.1 Network layer

2.1.1 Network socket

IP version 4 as defined in RFC791 **MUST** be used.

IP multicast **SHOULD** be available for sending and receiving according to RFC1112. Support of IGMPv2 is **RECOMMENDED**.

IP version 6, as defined in RFC2460, **SHOULD** be supported.

2.2 RTP: Realtime Transport Protocol

2.2.1 UDP

UDP as defined in RFC768 **MUST** be used. The checksum in the UDP header **MUST** be used.

2.2.2 RTP standards

Realtime Transport Protocol (RTP) over UDP **SHALL** be used as transport protocol as defined in:

- RFC3550 'RTP: A Transport Protocol for Real-Time Applications
- RFC3551 'RTP: Profile for Audio and Video Conferences with Minimal Control

2.2.3 RTCP

The implementation of RTP Control Protocol is **RECOMMENDED**, in particular, Sender Report (SR) and Receiver Report (RR) messages. On unidirectional links only Sender Report is used.

Note: These can be useful for audio synchronisation issues ('lipsync') and active recovery (retransmission). On unidirectional links only Sender Report is used.

2.2.4 CSRC

CSRC is generally absent as it is only employed when an RTP mixer is used. The presence or absence of CSRC **MUST** not affect the receiver.

2.2.5 Port assignment

Port 5004 (RTP) and port 5005 (RTCP) **SHOULD** be used as default ports. It is recommended that streams are sent and received on the same port.

When more than one channel is used, it is possible to multiplex streams on the same port or use different ports. In the case that different ports are used, the port assignment **SHOULD** be as follows:

- Channel 1 RTP: 5004
- Channel 1 RTCP: 5005
- Channel 2 RTP: 5014
- Channel 2 RTCP: 5015
- Channel 3 RTP: 5024
- Channel 3 RTCP: 5025

2.2.6 Forward Error Correction

If FEC is required, RTP Payload Format for Generic Forward Error Correction as described in RFC5109 **SHOULD** be used.

Port 5006 **SHOULD** be used as a default. Non-FEC capable receivers **MUST** not be affected by FEC packets and **MUST** ignore them.

Support of Forward Error Correction according to RFC2733 is **OPTIONAL**.

2.2.7 Retransmission

Active recovery by using retransmission according to RFC4588 **MAY** be used. Support of Extended RTP profile as defined in RFC4585 is necessary for implementing retransmission.

2.3 TCP: Transmission Control Protocol

2.3.1 TCP standard

Transmission Control Protocol (TCP) as defined in RFC793 **MAY** be implemented in addition to RTP.

Note: On network paths with IP address translation TCP may be better suited as it is state-full compared to UDP which is stateless. Congestion avoidance mechanisms of TCP may lead to problems with continuous stream on network with packet loss and long round trip delay. Transmission overhead is higher with TCP.

2.3.2 Port assignment

5004 **SHOULD** be used as a default port for 'RTP over TCP' transport.

2.3.3 TCP encapsulation

If TCP transport is implemented, the framing **SHOULD** be done according to RFC4571 for 'RTP over connection oriented transport'.

3. Audio coding

3.1 Mandatory audio codecs

3.1.1 ITU G.711

MIME Subtype name: PCMA and PCMU

RTP Payload type: 8 for PCMA and 0 for PCMU (RFC3551)

ITU G.711 audio coding standard with a bitrate of 64 kbit/s **SHALL** be implemented.

RTP encapsulation according to RFC3551 **SHALL** be used. Payload format names are 'PCMA' for A-law with RTP Payload type '8' and 'PCMU' with RTP Payload Type '0' for mu-law.

As defined in RFC3551, each G.711 octet **SHALL** be octet-aligned in an RTP packet. The sign bit of each G.711 octet **SHALL** correspond to the most significant bit of the octet in the RTP packet.

20 ms of audio per RTP packet **SHOULD** be used as default (RFC3551) for improved compatibility with voice-over-IP systems. Both receiver and transmitter **MUST** be able to deal with other durations of audio per RTP packet.

Sender MUST signal 'rtpmap' field in the SDP message. Receiver **MUST** accept SDP messages with or without 'rtpmap'.

Example: G.711 Mu-Law: "a=rtpmap:0 PCMU/8000"

3.1.2 ITU G.722

MIME Subtype name: G722

RTP Payload type: 9 (RFC3551).

ITU G.722 audio coding standard with a bitrate of 64 kbit/s **SHALL** be implemented.

RTP encapsulation according to RFC3551 **SHALL** be used. Profile name is 'G722' and RTP Payload Type is '9'. The first bit transmitted in the G.722 octet, which is the most significant bit of the higher sub band sample, **SHALL** correspond to the most significant bit of the octet in the RTP packet.

According to RFC3351: Even though the sampling rate for G.722 audio is 16 kHz, the RTP clock rate for the G.722 payload format is 8 kHz because this value was erroneously assigned in RFC1890 and **MUST** remain for backward compatibility. The octet rate is 8 kHz.

20 ms of audio per RTP packet **SHOULD** be used as default (RFC3551). Both receiver and transmitter **MUST** be able to deal with other durations of audio per RTP packet.

Sender MUST signal 'rtpmap' field in the SDP message. Receiver must accept SDP messages with or without 'rtpmap'.

Example: G.722 : "a=rtpmap:9 G722/8000"

3.1.3 ISO MPEG-1/2 Layer II

MIME Subtype name: MPA

RTP Payload Type: 14 (RFC3551) or dynamic.

ISO/IEC 11172-3 MPEG-1 Layer II and ISO/IEC 13818-3 MPEG-2 Layer II coding **SHALL** be implemented.

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

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

RTP audio frame encapsulation according to RFC2250 **MUST** be used with encapsulation of MPEG audio elementary streams.

An integral number of frames **MUST** be contained within the packet.

In all case the encoder **MUST** signal MPEG parameters in the SDP files.

The decoder **MUST** support both static payload type 14 and dynamic payload types. MPEG parameters **MUST** still be signalled in SDP file with a= rtpmap: and a=fmtp: fields according to RFC3555 and RFC4566. The reason is to ensure that signalling layer get all the necessary information about the stream.

Sampling frequency specified in the a=rtpmap: SDP field is 90 kHz in all cases and corresponds to the RTP clock.

Audio sampling frequency parameter 'samplerate' and other parameters 'layer' (layer 2) 'mode' (for joint-stereo, etc) 'bitrate' (in bit/s) are specified in a=fmtp: SDP field. Note that 'single_channel' is used in parameter 'mode' when signalling a mono stream.

Note that in the case of MPA, the number of channels **MUST** not appear in rtpmap:

The RTP payload clock is always 90 kHz.

SDP example: MPEG Layer II stereo, 48 kHz sampling at 256 kbit/s,

> "a=rtpmap:96 MPA/90000"
>
> "a=fmtp:96 layer=2;bitrate=256000;samplerate=48000;mode=stereo"

3.1.4 16 bit PCM

MIME Subtype name: L16 (RFC3555) Linear audio.

Sampling frequencies to be supported: 32 kHz, 48 kHz

For 16 bit per sample quantization, RTP frame encapsulation **MUST** be done according RFC3551. Samples are transmitted in network byte order (most significant byte first).

For multiple channels, section 4.1 of RFC3551 gives instructions.

For 32 kHz and 48 kHz sampling frequencies a dynamic payload type **MUST** be used and the sampling frequency **MUST** be specified in the SDP file according to RFC3555.

4 ms of audio per RTP packet **SHOULD** be used as default. 'a=ptime:' SDP field as defined in RFC4 566 SHOULD be signalled in SDP for improved compatibility. 'ptime' gives the length of time in milliseconds represented by the media in the RTP packet.

Implementation of this section is **OPTIONAL** for portable units.SDP example: Linear audio at 48 kHz, stereo, 4 milliseconds per packet

> "a=rtpmap:97 L16/48000/2"
>
> "a=fmtp:97"
>
> "a=ptime:4"

3.1.5 PCM at 12, 20 and 24 bit per sample

MIME Subtype name: DAT12, L20, L24

RTP Payload Type: Dynamic

For quantization of 20 and 24-bit per sample, RTP encapsulation as defined in RFC3190 **MUST** be used. 12-bit per sample quantization is **OPTIONAL**.

4 ms of audio per RTP packet **SHOULD** be used as default. 'a=ptime:' SDP field as defined in RFC4566 **SHOULD** be signalled in SDP for improved compatibility. 'ptime' gives the length of time in milliseconds represented by the media in the RTP packet. Implementation of this section is **OPTIONAL** for portable units.

3.2 Recommended audio codecs

3.2.1 MPEG-4 AAC, MPEG-4 AAC-LD

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

- For MPEG-4, RTP audio frame encapsulation as defined in RFC3640 **MUST** be used.
- High bitrate AAC profile (AAC-hbr) **MUST** be supported and **SHOULD** be used.
- Support of interleaving is **OPTIONAL.** AAC bitrate **MUST** be signalled in SDP field "b=TIAS:<audio bitrate in bit/s>". Due to some device incompatibilities it is also recommended to signal at the same time the bitrate as "bitrate=<audio bitrate in bit/s>" in the "a=fmtp:" set of AAC parameters.

MIME Subtype name with RFC3640: mpeg4-generic

RTP Payload Type: dynamic

See Annex 1 (Informative) for more details on AAC Transport.

3.2.2 Standard/Enhanced APT-X

MIME Subtype name: 'aptx'

ADPCM based format from CSR corporation (http://www.csr.com).

RTP audio frame encapsulation as defined in RFC7310 **MUST** be used.

3.3 Optional audio codecs

3.3.1 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. Bitrates 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

Support of RTP frame encapsulation according to RFC3119 is **RECOMMENDED**.

MIME Subtype name with RFC3119: MPA-ROBUST

RTP Payload Type: **MUST** be dynamic (RFC3119)

If MPEG encapsulation according to RFC2250 is used, the same SDP rules as for MPEG Layer II are applied (with 'layer' field set to 3).

3.3.2 MPEG-4 HE-AACv2

RTP encapsulation according to RFC3640 **SHOULD** be used. High bitrate profile **SHOULD** be used. Non-Backward Compatible signalling SHOULD be used.

See Annex 1 (Informative) for more details on AAC Transport.

3.3.3 Opus

Free Open source voice/music codec with bitrates ranging from 6 kbit/s to 510 kbit/s as defined in RFC6716. http://www.opus-codec.org/

RTP payload format draft: http http://www.ietf.org/id/draft-ietf-payload-rtp-opus-03.txt

3.3.4 AMR-WB/AMR-WB+

(Extended Adaptive Multi-Rate Wideband as defined in 3GPP TS 26.290).

3GPP originally developed the AMR-WB+ audio codec for streaming and messaging services in the Global System for Mobile communications (GSM) and third generation (3G) cellular systems. The codec is designed as an audio extension of the AMR-WB (G.722.2) speech codec.

If AMR-WB+ is implemented, RTP encapsulation according to RFC4352 **SHOULD** be used.

If AMR-WB (G722.2) is implemented, RTP encapsulation according to RFC4867 **SHOULD** be used. If exclusively G.722.2 is used for voice conversation applications, this RFC**SHOULD** be used. For streaming, RFC4352 can be used for both AMR-WB+ and AMR-WB.

3.3.5 Other audio coding

The rule in the EBU specification is that audio formats must have an RTP payload format defined and registered at IETF.

Any codec manufacturer is welcome to submit a payload format to the IETF payload group for publishing as an RFC. This is important for maintaining the present interoperability document in the longer term.

4. Signalling

4.1 Stream description

4.1.1 SDP

Session Description Protocol according to RFC4566 **MUST** be used for session description.

According to RFC4566 section 5, fields ‘v’, ‘o’, ‘s’, ‘c’, ‘t’ and ’m’ are mandatory:

‘v’ is the protocol version

‘o’ are the originator and session identifiers,

‘s’ is the session name (<u>single space</u> if not used – must not be empty!),

‘c’ is the connection data

‘t’ are the session times (0 0 if not used)

‘m’ is the media description

In FNS-ACIP case field ‘a’ for media attributes is almost always necessary with ‘a=rtpmap:’ for payload definition and sometimes ‘a=fmtp:’ for additional payload parameters.

Care **MUST** be taken to implement media field: ‘m=’ and attribute field: ‘a=’ correctly in order to ensure correct codec interpretation. On ‘a=rtpmap:’ codec MIME subtype name **MUST** be used followed by specified parameters such as clock rate and encoding parameters. The general ‘m’, and ‘a’ format for one audio stream using RTP is as follows:

m=audio <port> RTP/AVP <payload type>

a=rtpmap:<payload type> <encoding name>/<clock rate>[/<encoding parameters>]

Transport is not defined in SDP so a carriage protocol **MUST** be defined (e.g. SAP, RTSP, FTP).

Example of SDP:

Description of a 16 bit PCM, 48 kHz stereo session:

v=0

o=alice 2890844526 2890844526 IN IP4 host.anywhere.com

s= (single space)

c=IN IP4 host.anywhere.com

t=0 0

m=audio 5004 RTP/AVP 98

a=rtpmap:98 L16/48000/2

4.1.2 MIME subtypes

MIME subtype name according to RFC3555 **MUST** be used for audio description.

Other MIME type can be found at: http://www.iana.org/assignments/media-types/audio/

Examples:

‘G722’ for G.722

‘L16’ for 16 bit PCM

‘DAT12’ for 12 bit PCM

‘L20’ for 20 bit PCM

‘MPA’ for MPEG-1 or MPEG-2 audio

4.2 Session announcement

4.2.1 SAPv1

SAPv1 according to RFC2974 **SHOULD** be supported. It is used for unidirectional multicast links. A modification to RFC2974 to allow SAP with unicast addresses is proposed here.

4.3 Session management with SIP

4.3.1 Session Initiation Protocol (SIP)

SIP, according to RFC3261, **MUST** be used as the signalling method for bidirectional links. The SIP requests ‘INVITE’, ‘ACK’, ‘BYE’ and ‘OPTIONS’ **MUST** be supported for basic communication. SIP registrar **MUST** also be supported so ‘REGISTER’ message **MUST** be supported.

5060 **MUST** be used as a default port for establishment. This is important for direct calls without gateways. The user may be able to change the port in case of problems (filtering).

User agents

User agents **MUST** be capable of using profiles with multiple audio formats. Signalled as a list of formats in the SDP Media Description as defined in section 5.14 of RFC4566 - SDP: Session Description Protocol.

User agents **SHOULD** conform to RFC3515 - Session Initiation Protocol (SIP) Refer Method

Proxies/Registar

Registrar/proxy discovery: User-agents **MUST** follow RFC3263 - Session Initiation Protocol (SIP) Locating Sip Servers when looking up the SIP registrar and proxy for registration and invitations respectively.

User agents **MUST** have an option enabling "outbound proxy" meaning that instead of sending invite messages directly to the callee, or the proxy managing the callee, all invite messages are sent to the proxy that manages the caller.

When communicating using SIP proxies, all SIP messages **MUST** go through the proxy.

4.3.2 Codec negotiation

For codec negotiation between two units, the model described in RFC3264 (An Offer/Answer Model with the Session Description Protocol) **MUST** be used. User SHOULD be able to define prioritisation of codec.

4.3.3 Reconfiguration

Reconfiguration (change of audio codec) may be needed when a connection is already established. Reconfiguration **SHOULD** be supported.

For reconfiguration to happen the sender **MUST** send a new SIP INVITE command with SDP signalling the new codec to be used. The receiver replies with a SIP ACK command to acknowledge the codec change. The sender **MUST** not apply reconfiguration before it has received acknowledgement.

5. Bibliography

The following RFCs may be found on the IETF website at http://www.ietf.org/rfc.html.

1. RFC768: UDP: User Datagram Protocol
2. RFC791: Internet Protocol (version 4)
3. RFC1112: Host Extensions for IP Multicasting
4. RFC2250: RTP Payload Format for MPEG1/MPEG2 Video
5. RFC2460: Internet Protocol, Version 6 (IPv6)
6. RFC2733: An RTP Payload Format for Generic Forward Error Correction
7. RFC2974: Session Announcement Protocol
8. RFC3119: A More Loss-Tolerant RTP Payload Format for MP3 Audio. Improvement of RFC2250 for MPEG1/2 Layer III.
9. RFC3190: RTP Payload Format for 12-bit DAT, 20- and 24-bit Linear Sampled Audio
10. RFC3261: SIP: Session Initiation Protocol (SIPv2)
11. RFC3264: An Offer/Answer Model with the SDP
12. RFC3550: RTP: A Transport Protocol for Real-Time Applications
13. RFC3551: RTP: Profile for Audio and Video Conferences with Minimal Control
14. RFC3555 MIME Type Registration of RTP Payload Formats
15. RFC3640: RTP Payload Format for Transport of MPEG4 Elementary Streams
16. RFC4352: RTP Payload Format for the Extended Adaptive Multi-Rate Wideband (AMRWB+) Audio Codec
17. RFC4566: SDP: Session Description Protocol
18. RFC4588: RTP retransmission

19 RFC6716: Definition of the Opus Audio Codec

20. RFC7310: RTP Payload Format for Standard apt-X and Enhanced apt-X Codecs

6. Glossary

3GPP	3rd Generation Partnership Project
AAC	Advanced Audio Coding
AAC-LD	Advanced Audio Coding Low Delay
AMR-WB	Adaptive Multi Rate - WideBand (G.722.2)
CSRC	Contribution Source (in RTP)
FEC	Forward Error Correction
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IP	Internet Protocol
MIME	Multipurpose Internet Mail Extensions
PCM	Pulse Coded Modulation
RFC	Request For Comments (IETF standard)

RTP	Realtime Transport Protocol.
RTCP	Realtime Control Protocol
RSTP	Realtime Streaming Protocol
SAP	Session Announcement Protocol
SDP	Session Description Protocol
SIP	Session Initiation Protocol
UDP	User Datagram Protocol

Annex 1: Parameters for AAC transport (Informative)

Raw AAC

Raw AAC consists of two main structural components:

- Audio Specific Config (ASC): contains initial information for the decoder, e.g. the Audio Object Type.
- Access Units (AUs): contain the compressed audio data, e.g. scale factor bands.

SDP

RFC3640 (payload format for elementary MPEG-4 streams) is indicated by 'RTP/AVP'and 'mpeg4-generic'

- StreamType is always 5 for AAC (AudioStream ISO/IEC 14496-1, Table 5)
- Profile-Level-ID is defined in ISO/IEC 14496-3, Table 1.14. It is an informative field that depends on AAC system used and complexity. Profile-Level-ID can be (in decimal) 16 (High Quality Audio Profile), 4 (Main Audio Profile), 46 (HE AAC Profile), 49, 22 (Low Delay Audio Profile).
- SDP needs to repeat the mandatory values from 'AAC-hbr'.

Example:

SDP for AAC-LC 44.1 kHz stereo at 128 kbit/s:

v=0

o=encoder 790729066 1212747664 IN IP4 192.168.1.2

s=Test Session

i=Session for Demo purposes from Fraunhofer IIS

e=jochen.issing@iis.fraunhofer.de

c=IN IP4 0.0.0.0

t=3421643750 0

a=control:*

a=range:npt=0.0-

m=audio 9202 RTP/AVP 101

b=TIAS:128000

a=rtpmap:101 mpeg4-generic/44100/2

a=fmtp:101 streamtype=5; profile-level-id=16; config=1210; mode=AAC-hbr; constantduration=1024; sizelength=13; indexlength=3; indexdeltalength=3; bitrate=128000

a=control:trackID=2000