EBU

OPERATING EUROVISION AND EURORADIO

TECH 3347

AUDIO OVER IP

PRODUCTION INTERCOMS

REQUIREMENTS FOR INTEROPERABILITY
Rev. 1

Technical Specification
Source: FNS-I3P

October 2012

Conformance Notation

This document contains both normative text and informative text.

All text is normative except for that in the Introduction, any section explicitly labelled as ‘Informative’ or individual paragraphs which start with ‘Note:’.

Normative text describes indispensable or mandatory elements. It contains the conformance keywords ‘shall’, ‘should’ or ‘may’, defined as follows:

‘Shall’ and ‘shall not’:	Indicate requirements to be followed strictly and from which no deviation is permitted in order to conform to the document.
‘Should’ and ‘should not’:	Indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others. OR indicate that a certain course of action is preferred but not necessarily required. OR indicate that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.
‘May’ and ‘need not’:	Indicate a course of action permissible within the limits of the document.

Default identifies mandatory (in phrases containing “shall”) or recommended (in phrases containing “should”) presets that can, optionally, be overwritten by user action or supplemented with other options in advanced applications. Mandatory defaults must be supported. The support of recommended defaults is preferred, but not necessarily required.

Informative text is potentially helpful to the user, but it is not indispensable and it can be removed, changed or added editorially without affecting the normative text. Informative text does not contain any conformance keywords.

A conformant implementation is one that includes all mandatory provisions (‘shall’) and, if implemented, all recommended provisions (‘should’) as described. A conformant implementation need not implement optional provisions (‘may’) and need not implement them as described.

Feedback on this document is invited; it should be sent to Mathias Coinchon (coinchon@ebu.ch), coordinator of the EBU I3P group.

Contents

Audio over IP – Production Intercoms
Requirements for Interoperability

EBU Committee	*First Issued*	*Revised*	*Re-issued*
TC	Feb. 2011	Rev.1, Oct. 2012	

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

1. Introduction

1.1 Production intercoms - need for interoperability

Intercom systems are used to establish real-time communication networks between different locations. For instance, a live TV event requires interconnection with one or more studios, one reporter in the field and one OB van to the main control room. This system will use one or more intercom matrices, many dedicated panels, some handsets and phones.

To connect the main intercom system to the outside, broadcasters used to transport intercom audio over ISDN or PSTN lines through dedicated interfaces but more and more also use IP interfaces to do so. This is in part accounted for the fact that several countries are withdrawing ISDN services, which have been heavily used for intercom in the past, and also due to the fact that IP links are often more affordable and easier to access.

More than 10 manufacturers currently provide units capable of interconnecting their intercom system over IP connections, and efforts must be made to achieve interoperability between units from different manufacturers.

Interoperability is here considered at an audio level. GPI and data transport may be discussed in the future but are not in the scope of this document.

1.2 Intercom use cases

In broadcast environments, Intercom interconnections may be used to transport different types of signals:

1. pure intercom
    a. full duplex point to point communication
    b. conference call (alternative point-to-multipoint)
2. mix-minus
    a. half-duplex
    b. or return feed of a full-duplex communication
3. commentary feed
    a. half –duplex
    b. main feed of a full-duplex communication

Consequently audio quality may differ with usage, but will also depend on habits and available bandwidth.

Where and why do we need interoperability?

1. International events:
    a. International Broadcasting Centre – link with different broadcasters on-site (matrix to matrix or matrix to 4-wire)
    b. Broadcaster's link with headquarters (matrix to matrix)

2. Everyday use:
    a. Communication between two studios (matrix-to-matrix)
    b. Communication between one studio and an OB van (matrix-to-matrix)
    c. Communication with a remote studio (matrix to matrix or matrix to panel)
    d. Communication with a third-party broadcaster or service provider (matrix-to-matrix or matrix-to-4wire)
    e. Communication with a VoIP Phone inside the company
    f. Communication with a reporter on the field (matrix-to-phone / matrix-to-softphone)

1.3 Types of devices

Different kinds of devices may be used to connect an intercom system to an IP network and are considered in this document:

- IP interface integrated into the main matrix system
- IP interface integrated into a peripheral device
- external dedicated device
- software application

1.4 What are the requirements for interoperability?

The requirements necessary to achieve audio interoperability between production intercom 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.

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

1.5 General recommendation about end-to-end delay

End to end latency (delay) depends on the addition of audio coding algorithm used, buffer management and IP network characteristics. It would be outside the scope of this document to try to set latency.

However, as a best practice for setting up an intercom system, a survey done with users of intercom systems has shown that such end-to-end latency value are expected

- Less than 100 ms for remote intercom system
- Less than 50 ms for in-house, internal intercom system

ITU-T G.114 figures on end-to-end delay for normal phone conversation can be used as guidelines.

2. Transport protocols

2.1 Network layer

2.1.1 Network socket

IP version 4 as defined in RFC 791 **MUST** be used.

IP version 6 as defined in RFC 2460 **SHOULD** be supported.

2.2 RTP: Real-time Transport Protocol

2.2.1 UDP

UDP as defined in RFC 768 MUST be used. The checksum in the UDP header **MUST** be used.

2.2.2 RTP standards

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

- RFC 3550 'RTP: A Transport Protocol for Real-Time Applications
- RFC 3551 'RTP: Profile for Audio and Video Conferences with Minimal Control

2.2.3 RTCP

The implementation of RTP Control Protocol is **OPTIONAL** Receiver that do not interpret RTCP messages **MUST** not be affected.

2.2.4 Port assignment

Port 5004 (RTP) and port 5005 (RTCP) **SHOULD** be used as default ports. It is recommended that the source port for an outgoing RTP/RTCP connection **SHOULD** be the same port as the port announced via SDP.

When more than one channel is used, it is possible to multiplex streams on the same IP port or use different IP ports. In the case that different ports are used, the IP 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

2.2.5 Forward Error Correction

Forward Error Correction is not considered in this specification but may be used when interconnecting with other devices (for example: ACIP compliant codecs). Non-FEC capable receivers **MUST NOT** be affected by FEC packets and **MUST** ignore them.

2.3 TCP: Transmission Control Protocol

2.3.1 TCP standard

Transmission Control Protocol (TCP) as defined in RFC 793 **MAY** be implemented in addition to RTP.

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

2.3.2 Port assignment

Port 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 RFC 4571 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 (RFC 3551)

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

RTP encapsulation according to RFC 3551 **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 RFC 3551 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 (RFC 3551) for improved compatibility with voice over IP systems. Receiver and transmitter **MUST** be able to deal with other amount of

audio time per RTP packet.

Sender **MUST** signal 'rtpmap' field in the SDP message. Receiver must accept SDP messages with or without 'rtpmap'. 'ptime' field **SHOULD** be used to signal audio duration used in one RTP packet.

Example: *G.711 Mu-Law: "a=rtpmap:0 PCMU/8000"*
a=ptime:80

3.1.2 ITU G.722

MIME Subtype name: G722

RTP Payload type: 9 (RFC 3551).

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

RTP encapsulation according to RFC 3551 **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 RFC 1890 and **MUST** remain for backward compatibility. The octet rate is 8 kHz.

20ms of audio per RTP packet **SHOULD** be used as default (RFC 3551). Receiver and transmitter **MUST** be able to deal with other amount of audio time per RTP packet.

Sender **MUST** signal 'rtpmap' field in the SDP message. Receiver **MUST** accept SDP messages with or without 'rtpmap'. 'ptime' field **SHOULD** be used to signal audio duration used in one RTP packet.

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

3.2 Recommended audio codecs

3.2.1 Speex

Speex is a free open source audio compression format designed for speech: (http://speex.org).

Support of Speex narrowband (8 kHz) and wideband (16 kHz) is **RECOMMENDED**.

For Speex, RTP audio frame encapsulation as defined in RFC 5574 **MUST** be used.

MIME Subtype name with RFC 5574: speex

RTP Payload Type: dynamic

Required parameters = rate (RTP timestamp clock rate, which is equal to the sampling rate in Hz. The sampling rate **MUST** be 8000, 16000, or 32000).

Optional parameters = ptime, maxptime, vbr, cng, mode (see RFC 5574)

Example supporting All Modes, Prefer Mode 4

m=audio 8088 RTP/AVP 97

a=rtpmap:97 speex/8000

a=fmtp:97 mode="4,any"

3.2.2 G.729 and variants

G.729 is an ITU standard codec; It offers toll quality speech at a reasonably low bit rate of 8 kbit/s.

Mime Subtype name : G729

RTP payload type :18

Required parameters : none. Optional parameters : ptime, maxptime : see RFC 4556

annexb : indicates that the voice activity detection (described in Annex B) is used or preferred.

Permissible values are “yes” and “no” (without quotes); “yes” is implied if this parameter is omitted.

Example: m=audio 5056 RTP/AVP 18 96

a=ptime:80

a=fmtp:18 annexb=no

a=rtpmap:18 G729/8000/1

G.729.1 is an 8 – 32 kbit/s scalable wideband (50 – 7000 Hz) speech and audio coding algorithm interoperable with G.729, G.729 Annex A, and G.729 Annex B.

The G.729.1 coder produces an embedded bitstream structured in 12 layers corresponding to 12 available bit rates between 8 and 32 kbit/s.

The first layer, at 8 kbit/s, is called the core layer and is bitstream compatible with the ITU-T G.729/G.729A coder. At 12 kbit/s, a second layer improves the narrowband quality. Upper layers provide wideband audio (50 – 7000 Hz) between 14 and 32 kbit/s, with a 2 kbit/s granularity allowing graceful quality improvements. Only the core layer is mandatory to decode understandable speech; upper layers provide quality enhancement and wideband enlargement.

Since G.729.1 is an extension of G.729, the offerer (i.e. the caller) **SHOULD** announce G.729 support in its "m=audio" line, with G.729.1 preferred. This will allow interoperability with both G.729.1 and G.729-only capable parties.

Example of such an offer:

m=audio 55954 RTP/AVP 98 18

a=rtpmap:98 G7291/16000

a=rtpmap:18 G729/8000

3.3 Optional audio codecs

3.3.1 G.723.1

G.723.1 is specified by ITU as “Dual-rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s”.

RTP payload format is defined in RFC3551.

MIME Subtype name: ‘G723’

SDP signalling is defined in RFC 3555.

3.3.2 MPEG-4 AAC family of codecs

(ISO/IEC 14496-3 MPEG-4 AAC, AAC-LD, AAC-ELD).

For MPEG-4, RTP audio frame encapsulation as defined in RFC 3640 MUST be used. High bitrate AAC profile **MUST** be supported and **SHOULD** be used. Support of interleaving is **OPTIONAL.**

AudiospecificConfig MIME as defined in RFC 3640 section 4.1.

MIME Subtype name with RFC 3640: mpeg4-generic

RTP Payload Type: dynamic

3.3.3 Enhanced - APTX

MIME Subtype names: ‘aptX’, ‘EaptX 16’, ‘EaptX 24’, ‘aptXLive’

ADPCM format from APT Corporation (www.aptx.com).

RTP payload format definition is in preparation by APT.

3.3.4 PCM L16

MIME Subtype name: L16 (RFC 3555) Linear audio.

Sampling frequencies recommended: 32 kHz, 48 kHz.

For 16 bit per sample quantization, RTP frame encapsulation **MUST** be done according to RFC 3551. Samples are transmitted in network byte order (most significant byte first).

For 32 kHz and 48 kHz sampling frequencies a dynamic payload type **SHOULD** be used and the sampling frequency **MUST** be specified in the SDP file according to RFC 3555.

4ms 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.

“a=rtpmap:97 L16/48000/2”

“a=fmtp:97”

“a=ptime:4”

3.3.5 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 RFC 4352 **SHOULD** be used

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

3.3.6 Other audio coding

Other audio codecs might be implemented but RTP audio frame encapsulation **SHOULD** be defined. An internet draft from the IETF ‘avt’ group is being written entitled ‘How to Write an RTP Payload Format’ (http://tools.ietf.org/html/draft-ietf-payload-rtp-howto-02).

3.4 General remarks

3.4.1 Silence management

If silence or voice detection is implemented, it **MUST** be possible to deactivate it. If the comfort noise feature is implemented, it **MUST** be possible to deactivate it.

3.4.2 Handling of unsupported RTP packets

Receivers **MUST** ignore packets that it does not understand.

4. Signalling

4.1 Stream description

4.1.1 SDP

Session Description Protocol according to RFC 4566 **MUST** be used for session description.

According to RFC 4566 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 (single space 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 N/I3P 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.

Example: the general ‘m’, ‘a’ format for one audio stream using RTP:

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 mono 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 L16/48000/1

4.1.2 MIME subtypes

MIME subtype name according to RFC 3555 **MUST** be used for audio description.

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

Example: 'G722' for G.722

4.2 Session management with SIP

4.2.1 Session Initiation Protocol (SIP)

SIP, according to RFC 3261, **MUST** be used as the signalling method for bidirectional links. Communication using SIP registrar **MUST** be supported. The SIP requests 'INVITE', 'ACK', 'BYE', 'REGISTER' and 'OPTIONS' **MUST** be supported for basic communication.

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

Port 5060 **MUST** be used as a default port for establishment. This is important for direct calls without gateways. When more than one channel is used on one interface, each channel **SHOULD** have its own SIP identifier and a session.

4.2.2 Codec negotiation

For codec negotiation between two units, the model described in RFC 3264 (An Offer/Answer Model with the Session Description Protocol) **MUST** be used. Codec negotiation **MUST** be supported in the unit.

4.2.3 Independent encoder/decoder settings

By default a call setup is with SDP parameters *a=sendrecv* for a bidirectional connection. The same codecs are used for encoder and decoder.

In the case that independent codecs are used for encoder and decoder, 2 SIP sessions **MUST** be used. One with *a=sendonly* and the other with *a=recvonly*. Support of independent encoder/decoder settings is **RECOMMENDED**.

4.2.4 SIP keepalive mechanisms

Support of the SIP keepalive mechanism is **RECOMMENDED**. If implemented, the SIP INVITE message **MUST** be used as a SIP keepalive mechanism according to RFC4028 by sending it periodically with the same call parameters.

4.2.5 Unexpected dropped connections

Generally, an unexpected connection break is quickly detected by the absence of media packets. If such an event occurs, a SIP INVITE **SHALL** be the mechanism to re-initiate connection.

4.2.6 Special case: permanent connections

In the case of permanent intercom connections, SIP **MAY NOT** be necessary and so **MAY NOT** be used. In this special case, only RTP ports and the encapsulation definitions for audio transportation are used. It means also that both ends must set all audio and IP parameters manually. Warning, this "manual mode" support does not mean that a product can avoid the integration of SIP to be compliant with this specification! SIP is a mandatory feature for implementation.

5. Bibliography

To find the following RFCs go to the IETF website at http://www.ietf.org/rfc.html .

1. RFC 791: Internet Protocol (version 4)
2. RFC 1112: Host Extensions for IP Multicasting
3. RFC 2460: Internet Protocol, Version 6 (IPv6)
4. RFC 768: UDP: User Datagram Protocol
5. RFC 3550: RTP: A Transport Protocol for Real-Time Applications
6. RFC 3551: RTP: Profile for Audio and Video Conferences with Minimal Control
7. RFC 2733: An RTP Payload Format for Generic Forward Error Correction
8. RFC 4588: RTP retransmission
9. RFC 3640: RTP Payload Format for Transport of MPEG 4 Elementary Streams
10. RFC 3190: RTP Payload Format for 12-bit DAT, 20- and 24-bit Linear Sampled Audio
11. RFC 4352: RTP Payload Format for the Extended Adaptive Multi-Rate Wideband (AMR WB+) Audio Codec
12. RFC 4566: SDP: Session Description Protocol
13. RFC 3555 MIME Type Registration of RTP Payload Formats
14. RFC 2974: Session Announcement Protocol
15. RFC 3264: An Offer/Answer Model with the SDP
16. RFC 3261: SIP: Session Initiation Protocol (SIPv2)

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