SMPTE STANDARD

Professional Media over Managed IP Networks: AES3 Transparent Transport

Table of Contents Page

Approved June 24, 2022

Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in its Standards Operations Manual. This SMPTE Engineering Document was prepared by Technology Committee 32NF.

This revision updates the 2018 original publication, including updating the normative references to current revisions, and some normative language adjustments based on the related PICS development. Based on feedback from field implementations, some notes and an annex have been added about the differences between AES3 and AES10. A section has been added for the purpose of IANA RTP payload format registration.

Intellectual property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

This clause is entirely informative and does not form an integral part of this Engineering Document.

The capability and capacity of IP networking equipment has improved steadily, enabling the use of IP switching and routing technology to transport and switch video, audio, and metadata essence within television facilities. Existing standards such as SMPTE ST 2022-6 have gained some amount of use in this application, but there was a desire in the industry to switch different essence elements separately.

This family of SMPTE engineering documents builds on the work of Video Services Forum (VSF) Technical Recommendations TR03 and TR04, and of AES67, documenting a system for transporting various essence streams over IP networks and capturing the timing relationships between those streams. The system is designed to be extensible to a variety of essence types.

SMPTE ST 2110-10 covers the system as a whole, the timing model, and common requirements across all essence types. Other documents will cover specific media essence formats.

SMPTE ST 2110-31 (this part) documents the transport of AES3 signals in transparent manner, preserving the V, U, C, and P bits. It builds upon a similar format defined within the Ravenna technology.

1 Scope

This Standard specifies the real-time, RTP-based transport of AES3 signals over IP networks, referenced to a network reference clock.

2 Conformance notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any clause explicitly labeled as "Informative" or individual paragraphs that start with "Note:”

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword “reserved” indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword “forbidden” indicates “reserved” and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document 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.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; then formal languages; then figures; and then any other language forms.

3 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this engineering document. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this engineering document are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

AES3-3-2009 (r2019) AES standard for digital audio — Digital input-output interfacing — Serial transmission format for two-channel linearly represented digital audio data, Part 3: Transport

AES67-2018 AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability

IETF RFC 3190 Kobayashi, K., Ogawa, A., Casner, S., and C. Bormann, RTP Payload Format for 12-bit DAT Audio and 20- and 24-bit Linear Sampled Audio, DOI 10.17487/RFC3190, January 2002

IETF RFC 3550 Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, DOI 10.17487/RFC3550, July 2003, https://www.rfc-editor.org/info/rfc3550

IETF RFC 3551 Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, DOI 10.17487/RFC3551, July 2003, https://www.rfc-editor.org/info/rfc3551

IETF RFC 4566 Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", DOI 10.17487/RFC4566, July 2006, https://www.rfc-editor.org/info/rfc4566

IETF RFC 8285 Singer, D., Desineni, H., and R. Even, Ed., "A General Mechanism for RTP Header Extensions", DOI 10.17487/RFC8285, October 2017, https://www.rfc-editor.org/info/rfc8285

SMPTE ST 2110-10:2022 Professional Media over Managed IP Networks: System Timing and Definitions

SMPTE ST 2110-30:2017 Professional Media over Managed IP Networks -- PCM Audio

4 Terms and definitions

For the purposes of this document, the terms and definitions of SMPTE ST 2110-10 and the following apply.

4.1 AES3 Block

block as defined in AES3-3

4.2 AES3 Frame

frame as defined in AES3-3

4.3 AES3 Subframe

subframe as defined in AES3-3

5 AES3 transport RTP format

5.1 Historical overview (informative)

SMPTE ST 2110-30 defines a method of transporting PCM audio data based upon AES67. While PCM audio and integration with AES67 are important in many television production environments, there also exist cases where transparent transport of entire AES3 signals (including the V, P, and U bits and the channel status information) is beneficial. The modern television ecosystem has leveraged the prevalence of AES3 signal transport to encapsulate many different data items; for example, SMPTE developed SMPTE ST 337 defining a general method for the encapsulation of various payloads into the AES3 transport, and also developed SMPTE ST 338 to manage the namespace of such payloads. Other use cases also exist in the industry where AES3 is used to transport non-PCM or even non-audio information.

This standard defines an RTP payload format for the transport of AES3 signals over IP.

5.2 RTP encapsulation

The sequence of AES3 Subframes inside the AES3 signals shall be transported using RTP as specified in IETF RFC 3550, and subject to the requirements and constraints of SMPTE ST 2110-10, subject to the constraints and payload definition below.

The technical metadata necessary to receive and interpret the RTP stream shall be communicated via SDP as defined in clause 6.

5.3 RTP header definitions

The fields of the RTP packet header shall be as specified in IETF RFC 3550 section 5.1 (RTP Fixed Header Fields), and shown in Figure 1. Fields not listed below are exactly as specified in IETF RFC 3550. The following additional constraints shall apply:

Payload Type (PT): 7 bits	Dynamically allocated payload type. The payload type shall be dynamically allocated in accordance with IETF RFC 3551
CSRC Count (CC): 4 bits	Set to zero (0)
Timestamp: 32 bits	RTP Timestamp as specified in SMPTE ST 2110-10 for PCM audio signals. For AES3 transparent subframes, the RTP timestamp is a sample of the RTP Clock at the time that the X or Z preamble of the first AES3 Subframe in the packet is presented to the encapsulator on the AES3 interface, or the equivalent value for an AES3 signal embedded into SDI.
Marker bit (M): 1 bit	Set to zero (0)
Extension bit (X): 1 bit	When this bit is set, an RTP header extension formatted as specified in IETF RFC 8285 shall be present immediately following the SSRC field. If this bit is not set, then there shall be no header extension present.

Unless otherwise specified, all multi-octet numeric values expressed in the RTP Header and in the RTP Payload Headers shall be expressed in Big-Endian Byte Order as defined in SMPTE ST 2110-10.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
|                           AM824 Subframe                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           AM824 Subframe                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
...
...
|                           AM824 Subframe                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1 — RTP Header with RTP Payload.

5.4 RTP Payload definition

The RTP Payload shall consist of an interleaved set of sequences of AM824 subframes. The contents of each AES3 Subframe shall be transported in one AM824 Subframe. The fields of the AM824 subframe are shown in Figure 2 and shall be as defined in the text below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|0|B|F|  PCUV |                  DATA24                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2 — AM824 subframe.

Block Start (B): 1 bit	When set to 1, indicates the first subframe of an AES3 Block.
Frame Start (F): 1 bit	When set to 1, indicates the first subframe of an AES3 Frame
PCUV: 4 bits	The P, C, U, and V bits of the AES3 Subframe, in order. The P bit is copied from timeslot 31 of the AES3 Subframe. The C bit is copied from timeslot 30, the U bit from timeslot 29, and the V bit from timeslot 28 of the AES3 Subframe.
DATA24: 24 bits	24 bits originating from time slots 4 to 27 of the AES3 Subframe. AES3 timeslot 27 is conveyed in AM824 subframe bit position 8.

AES3 Subframes 1 and 2 of each AES3 Frame shall be sequentially interleaved. If multiple AES3 signals are transmitted within the same RTP stream, then the AES3 Subframes from each AES3 signal shall be sequentially interleaved. Every packet within a stream shall contain data from the same number of AES3 signals. Each packet within a stream shall contain AES3 Subframes from the same number of periods of the sampling clock. The time period corresponding to each packet within the stream shall be signaled using the `ptime` attribute in the SDP, as defined in subclause 6.1, using one of the permitted values from Table 1.

Note 1 In AES3-3, the Block Start (B) bit set to 1 corresponds to the “Z” preamble and requires that the Frame Start (F) bit is also set to 1. When the Block Start (B) bit is set to 0, the Frame Start (F) bit is set to 1 in correspondence to the “X” preamble, and set to 0 in correspondence to the “Y” preamble. Additional notes about proper handling of the B, F, C, and U bits can be found in Annex A.

Note 2 The AES3 Subframe in AES3-3 is divided into timeslots numbered 0..31 where timeslot 0 is transmitted first, and multi-bit quantities such as the audio sample value are organized with the lowest-numbered bit in the bitfield carrying the least significant bit in the value. This RTP payload format is composed of 32-bit AM824 subframes in which the most significant bit of a multi-bit quantity is actually the lowest-numbered bit in the big-endian representation shown.

Note 3 The term “AM824” used here as a moniker for the subframe structure, and as the name for the payload in the SDP. The name originates in IEC 61883-6 where it is used in a similar application.

5.5 Media Clock and RTP Clock

AES3 streams in scope of this standard shall have a sampling frequency of 44.1 kHz, 48 kHz, or 96 kHz. The Media Clock and RTP Clock rate of the RTP Stream shall be the same as the sampling frequency. Senders and Receivers under this standard shall support a sampling frequency of 48 kHz, and sampling frequencies of 44.1 kHz or 96 kHz may be supported.

The offset between the RTP Clock and the Timestamp Reference Clock is zero as specified in SMPTE ST 2110-10.

The relationship between the sampling rate clock and the AES3 Frame shall be as defined in AES3-3-2009.

5.6 Timing

Senders and receivers shall observe the timing provisions of AES67-2018 section 7.5 titled “Sender timing and receiver buffering”.

6 Session Description

6.1 General

Streams under this standard shall be signaled using Session Description Protocol (SDP) in accordance with SMPTE ST 2110-10.

Streams under this standard shall be signaled in the SDP using the media type “audio” and the media subtype “AM824”.

The number of AES3 Subframe sequences multiplexed within the payload shall be signaled in the SDP object on the `a=rtpmap` line, using the syntactic field which typically communicates the number of channels in an audio signal, as shown below:

```
a=rtpmap:<pt> AM824/<clock-rate>/<nchan>
```

where

`<clock-rate>` is the RTP Clock Rate (equal to the sampling rate)

`<nchan>` is the number of AES3 Subframes multiplexed together within the payload

`<pt>` is the dynamically assigned RTP Payload Type found in the RTP header

The `<clock-rate>` parameter shall take one of the values 44100, 48000, or 96000.

Since this standard transports AES3 signals, and each AES3 signal contains two sequences of AES3 Subframes, the number of AES3 Subframe sequences `<nchan>` expressed in the SDP object shall always be an even number.

Senders under this standard shall signal a `ptime` attribute in the SDP, as defined in IETF RFC 4566 and shown as `<packet-time>` below.

```
a=ptime:<packet-time>
```

The `<packet-time>` parameter shall take one of the values from the Table 1, based on the prevailing sampling rate and the desired number of AES3 Subframe sequences interleaved together within the RTP stream.

Table 1 — Permitted values of Packet Time for each RTP Clock rate.

<packet-time> (ms)	**<clock-rate> RTP Clock (Hz)**	**Number of periods of the RTP Clock per packet**
1	48000	48
0.12	48000	6
0.08	48000	4
1	96000	96
0.12	96000	12
0.08	96000	8
1.09	44100	48
0.14	44100	6
0.09	44100	4

Note 1 For the avoidance of doubt, the `<packet-time>` value in the context of this RTP Payload specification indicates the approximate amount of time required (in milliseconds) for the transmission of the underlying AES3 Subframes when formatted into AES3-3 signals at the signaled `<clock-rate>`.

Note 2 The parameter values in Table 1 are consistent with AES67:2018 section 8.1.

6.2 Channel Order signaling

For AES3 Subframes containing PCM audio, Senders may signal the channel order in the SDP using the Channel Order Convention specified in SMPTE ST 2110-30. Because this standard can also transport non-PCM audio signals, the additional Channel Grouping Symbol listed in Table 2 may be used in addition to those specified in ST 2110-30.

Table 2 — Additional Channel Order Convention Grouping Symbols.

Channel Grouping Symbol	Quantity of AES3 Subframe sequences in group	Description of group
`AES3`	2	Either Non-PCM signals or PCM signal, or one subframe of each.

7 Conformance Levels

All Receivers under this standard shall implement the Level A requirements in Table 3. Senders and Receivers conforming to this standard should specify the Table 3 Conformance Levels that are supported when indicating conformance to this standard.

Table 3 — Conformance Levels.

Level	Receiver shall support			
	Sampling Clock Rate (Hz)	**Signaled Packet Time (ms)**	**Maximum Number of AES3 Subframe Sequences per packet**	**Maximum Number of AES3 signals**
A	48000	1	6	3
AX	48000 44100 96000	1 1.09 1	6 6 2	3 3 1
B	48000 48000	1 0.12	6 8	3 4
BX	48000 48000 44100 44100 96000 96000	1 0.12 1.09 0.14 1 0.12	6 8 6 8 2 4	3 4 3 4 1 2
C	48000 48000	1 0.12	6 60	3 30
CX	48000 48000 44100 44100 96000 96000	1 0.12 1.09 0.14 1 0.12	6 60 6 60 2 30	3 30 3 30 1 15
D	48000 48000 48000	1 0.12 0.08	6 60 80	3 30 40

Level	Receiver shall support			
	Sampling Clock Rate (Hz)	Signaled Packet Time (ms)	Maximum Number of AES3 Subframe Sequences per packet	Maximum Number of AES3 signals
DX	48000	1	6	3
	48000	0.12	60	30
	48000	0.08	80	40
	44100	1.09	6	3
	44100	0.14	60	30
	44100	0.09	80	40
	96000	1	2	1
	96000	0.12	30	15
	96000	0.08	40	20

Note The signaled packet time is rounded to 2 decimal places, with midway values such as 0.125 rounded down.

8 IANA RTP payload format registration information

8.1 Overview

This clause provides information to support the registration of the payload subtype name “AM824” with the Internet Assigned Numbers Authority (IANA) in accordance with IETF RFC 4855. The template defined in IETF RFC 6838 is used.

8.2 Media Type definition

8.2.1 Type and Subtype Names

Type Name: audio

Subtype Name: AM824

8.2.2 Required parameters

`rate`: RTP timestamp clock rate, as specified in subclause 5.5 of this document

`ptime`: The length of time (in milliseconds) represented by the media in this packet, as specified in subclause 6.1 of this document.

`channels`: The number of AES3 Subframes multiplexed together for each sample period, signaled as specified by the placeholder `<nchan>` in subclause 6.1 of this document.

8.2.3 Optional parameters

`channel-order`: The channel-order parameter as specified in IETF RFC 3190, using the channel-order convention defined in SMPTE ST 2110-30, as extended in this document.

8.2.4 Encoding considerations

This media type is framed and binary; see IETF RFC 6838 section 4.8.

8.2.5 Security considerations

RTP packets using the payload format defined in this specification are subject to the same considerations as outlined in IETF RFC 3190 section 9. Those considerations are incorporated here by reference.

8.2.6 Interoperability considerations

Care was taken in the development of this payload specification to retain compatibility with existing implementations in predecessor systems.

8.2.7 Published specification

The IANA registration should cite this document as the authoritative reference document for the registration of this media type and payload format.

8.2.8 Applications

This media type is used by professional equipment commonly found in the television production and distribution industry, for the transport and intercommunication of signals traditionally carried within AES3 bit-serial transports.

8.2.9 Additional information

Deprecated alias names for this type: N/A

Magic number(s): N/A

File Extension(s): N/A

Macintosh file type code(s): N/A

8.2.10 Contact information

Comments or questions about this document can be addressed to the SMPTE Standards Vice President, at svp@smpte.org.

8.2.11 Intended usage and restrictions

This media type is intended for common use. It is restricted to use within the context of RTP framing and is only defined for transfer via RTP as defined in IETF RFC 3550. Transport within other framing protocols is not defined.

8.2.12 Change control

This document is change-controlled by the Society of Motion Picture and Television Engineers (SMPTE).

Annex A Regarding differences between AES3 and AES10 (informative)

While the scope of this document is limited to AES3 signals, implementers are cautioned that the contents of AES3 signals are often exchanged or interoperated with AES10 (MADI) environments.

AES3 specifies alignment of the Block Start (B), Channel Status Data (C) and User Data (U) bits between the two channels within an AES3, while AES10 can carry channel status data (C) and user data (U) not aligned between pairs of channels; further, in AES10 the Block Start (B) bit set to 1 can occur in any subframe independently to indicate the start of the channel status data and user bit data block in that subframe.

A resilient receiver implementation ought to tolerate a Block Start (B) bit occurring in conjunction with a Frame Start (F) bit set to 0 (see also AES3-2-2009, section 5.2).

Devices receiving or transmitting AES3 Frames derived from/destined to AES10 interfaces also need to take account of the reversed definition of the Frame Start (F) bit in AES10 (AES Subframe Ā / B). In order to meet the requirements of this standard, devices could need to adjust (reverse) the Frame Start (F) bit accordingly upon conversion from AES10 into stream data compliant with AES3 and vice-versa.

Bibliography (informative)

AES3-2-2009 (r2019) AES standard for digital audio — Digital input-output interfacing — Serial transmission format for two-channel linearly represented digital audio data, Part 2: Metadata and Subcode

AES10-2008 (r2019) AES Recommended Practice for Digital Audio Engineering – Serial Multichannel Audio Digital Interface (MADI)

IETF RFC 4855 Casner, S., Media Type Registration of RTP Payload Formats", DOI 10.17487/RFC4855, February 2007, https://www.rfc-editor.org/info/rfc4855

IETF RFC 6838 Freed, N., Klensin, J., and T. Hansen, Media Type Specifications and Registration Procedures", IETF BCP 13, DOI 10.17487/RFC6838, January 2013, https://www.rfc-editor.org/info/rfc6838

ISO/IEC 61883-6 Ed. 2 2005 Consumer audio/video equipment – Digital interface – part 6: Audio and Music Data Transmission Protocol

VSF TR-03 "Transport of Uncompressed Elementary Stream Media over IP", November 12, 2015, http://www.videoservicesforum.org/download/technical_recommendations/VSF_TR-03_2015-11-12.pdf

SMPTE ST 337:2015 Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface

SMPTE ST 338:2015 Format for Non-PCM Audio and Data in AES3 — Data Types

