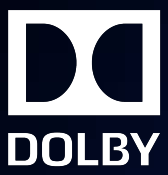


Dolby AC-4 and HTTP Live Streaming

Specification

7 July 2020
Confidential information

Notices

Copyright

Dolby Laboratories, Inc.
1275 Market Street
San Francisco, CA 94103-1410 USA
Telephone 415-558-0200
Fax 415-645-4000
http://www.dolby.com

Confidential information

Confidential information for Dolby Laboratories Licensees only. Unauthorized use, sale, or duplication is prohibited.

Confidential information for OEM customers of Dolby Laboratories only. Unauthorized use, sale, or duplication is prohibited.

Trademarks

Dolby and the double-D symbol are registered trademarks of Dolby Laboratories.

The following are trademarks of Dolby Laboratories:

Dialogue Intelligence™
Dolby®
Dolby Advanced Audio™
Dolby Atmos®
Dolby Audio™
Dolby Cinema®
Dolby Digital Plus™
Dolby Digital Plus Advanced Audio™
Dolby Digital Plus Home Theater™
Dolby Home Theater®
Dolby Theatre®
Dolby Vision®
Dolby Vision IQ™
Dolby Voice®
Feel Every Dimension™
Feel Every Dimension in Dolby™
Feel Every Dimension in Dolby Atmos™
MLP Lossless™
Pro Logic®
Surround EX™

All other trademarks remain the property of their respective owners.

Contents

1

Introduction to Dolby AC-4 and HTTP Live Streaming specification

This information set specifies the required data formatting and signaling between the server (sender) and the clients (receivers) to enable Dolby AC-4 to be used as an audio format within HTTP Live Streaming A/V delivery applications in conjunction with packed audio or ISO base media file format.

1.1 About this documentation

This documentation is written for a system designer to develop a product that can properly multiplex Dolby AC-4 bitstreams into a fragmented mp4 file or audio-only elementary stream segments.

Note: This documentation covers only delivering Dolby AC-4 bitstreams via HTTP Live Streaming (HLS). For a full description of HLS, see *HTTP Live Streaming 2nd Edition*.

Note: This documentation covers only delivering Dolby AC-4 content within the HLS compliant ISO base media file format. For a full description of HLS, see *HTTP Live Streaming 2nd Edition*.

This documentation provides information regarding:

- Signaling of Dolby AC-4 audio streams within the HLS playlist file
- Storage of Dolby AC-4 bitstreams within ISO base media file format files
- Storage of Dolby AC-4 bitstreams within HLS compliant audio-only elementary stream segments

In addition, the following information is covered:

- The data required to identify a Dolby AC-4 bitstream within an ISO base media file format file and a playlist file
- The steps required to properly packetize a Dolby AC-4 bitstream for multiplexing and storage in an ISO base media file format file
- The buffering consideration for demultiplexing a Dolby AC-4 bitstream from an ISO base media file format file

1.2 Conventions used

Modal verbs are used in this documentation to differentiate between mandatory requirements and recommendations.

- “Must”: indicates that the corresponding requirement or instruction is mandatory.
- “Should”: indicates that the corresponding statement or instruction is a recommendation and is not mandatory.

1.3 Resources

Standards and Dolby documents provide additional information to assist you in designing your product.

Standards

- ETSI TS 103 190-2 v1.2.1, *Digital Audio Compression (AC-4) Standard, Part 2; Immersive and Personalized Audio*, available from http://www.etsi.org. This documentation extends the AC-4 codec with a number of new use cases relevant for next-generation audio services.
- ISO/IEC 14496-12:2015, *Information Technology—Coding of Audio-Visual Objects, Part 12: ISO Base Media File Format*, available from http://www.iso.org. This documentation is Part 12 of the MPEG-4 specification and describes storage of content in a media file.
- *HTTP Live Streaming 2nd Edition* , available from https://datatracker.ietf.org. (Search for "HTTP Live Streaming".)
- RFC 6381, *The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types*, August 2011, available from http://tools.ietf.org/html.
- ISO 639-2:1998, *Codes for the Representation of Names of Languages, Part 2: Alpha-3 Code*, as maintained by the ISO 639/Joint Advisory Committee, available from http://www.iso.org.

Other documents

HLS Authoring Specification for Apple Devices, available from https://developer.apple.com/documentation/http_live_streaming/hls_authoring_specification_for_apple_devices.

1.4 Contacting Dolby

Support services are available to address any questions and to provide advice about integrating Dolby technology into your product.

For product design or testing, contact Dolby at systemsupport@dolby.com. By utilizing Dolby expertise, especially during the design process, many problems that might require design revisions before a product is approved can be prevented.

Dolby is also available to review product plans, including preliminary design information, markings, displays, and control and menu layouts, with the goal of preventing problems early in the product development cycle.

If you have comments or feedback about this documentation, send us an email at documentation@dolby.com.

2

Dolby AC-4 packetization into a fragmented MP4 file

The process of analyzing and preparing a Dolby AC-4 bitstream for storage within a fragmented MP4 file is described. AC-4 sync frames must be converted to raw AC-4 frames before packetization.

2.1 Constraints on Dolby AC-4 elementary streams within HTTP Live Streaming compliant test streams

A Dolby AC-4 elementary stream must meet certain conditions to be delivered via HTTP Live Streaming.

If your product encodes Dolby AC-4 elementary streams, ensure that the product uses the proper settings to meet these requirements. We recommend that an HTTP Live Streaming multiplexer or segmenter reject a bitstream that is not in compliance to prevent problems with subsequent decoding.

A compliant Dolby AC-4 elementary stream must adhere to these requirements:

- The value of the `frame_rate_index` parameter must remain constant.
- The value of the `presentation_config` parameter must remain constant.
- The value of the `fs_index` parameter must remain constant.
- The value of the `channel_mode` parameter must remain constant.
- The value of the `content_classifier` parameter must remain constant.
- Constraints described in the Annex E of ETSI TS 103 190-2 must be applied to the Dolby AC-4 bitstream for storage within the ISO base media file format.
- The Dolby AC-4 random access points must be encoded correctly to ensure compliance with the requirements described in Packetizing the Dolby AC-4 bitstream on page 9.
- We strongly recommend that Dolby AC-4 elementary streams be encoded with the same frame rate as the associated video frame rate to ensure constant alignment of video and audio access units in order to utilize the features of A/V alignment.

Related information

A/V alignment and segmentation on page 15

2.2 Content of the MP4 sample

A Dolby AC-4 MP4 sample corresponds to one raw Dolby AC-4 frame.

For detailed information on the Dolby AC-4 MP4 sample content, refer to Annex E of ETSI TS 103 190-2.

2.3 Reading Dolby AC-4 frames from an .ac4 file

In an *.ac4* file, each `raw_ac4_frame` is encapsulated in an `ac4_syncframe`, as defined in ETSI TS 103 190-1. An AC-4 sync frame must first be converted to a raw AC-4 frame by stripping its sync word, frame size, and optional cyclic redundancy check (CRC) words before packetization.

Procedure

1. To locate the start of an AC-4 raw frame, search for the 16-bit sync word. The value of the sync word can be either `0xAC40` or `0xAC41`.
2. Read the 16 bits immediately succeeding the sync word.
   - If the 16-bit value is not equal to `0xFFFF`, the value is the frame size.
   - If the 16-bit value is equal to `0xFFFF`, read the next 24 bits as the frame size.
3. According to the frame size, read the next bits succeeding the frame size as the raw AC-4 frame.
4. **Optional:** If the sync word is `0xAC41`, the AC-4 simple transport frame includes a 16-bit CRC word based on the `frame_size` element and the `raw_ac4_frame` element. Read the 16 bits immediately succeeding the raw AC-4 frame, and verify that the value matches the CRC calculated using the IBM-CRC-16 polynomial: $x^{16}+x^{15}+x^{2}+1$.
5. Read the next 16 bits.
   - If the sync word is `0xAC40`, they are the 16 bits immediately succeeding the raw AC-4 frame.

- If the sync word is `0xAC41`, they are the 16 bits immediately succeeding the CRC word.

6. Verify that the 16-bit word is also a sync word; otherwise, this is not a proper AC-4 bitstream.

2.4 Packetizing the Dolby AC-4 bitstream

A multiplexer packetizes only one raw AC-4 frame as an MP4 sample.

About this task

A seamless switch can be accomplished only at a random access point. The first sample of each segment must contain a random access point. A Dolby AC-4 elementary stream contains independently decodable frame (I-frame)s to indicate random access points. Similar to the length of GOP for video streams, the I-frame interval for a Dolby AC-4 stream indicates the number of frames between two I-frames.

The multiplexer must ensure that the first MP4 sample of each Dolby AC-4 media segment is an I-frame. A Dolby AC-4 media segment may contain one or more random access points.

Procedure

1. The multiplexer scans the bitstream to find the first I-frame (a frame with the `b_iframe_global` value of `1`). This I-frame begins a Dolby AC-4 media segment.
2. The multiplexer continues to scan the bitstream for preceding I-frames and keeps track of the number of frames between two I-frames(I-frame interval).

   The I-frame interval should be constant.
3. When the multiplexer reaches the end of an I-frame interval and accumulates a media segment size worth of AC-4 frames, the multiplexer adds the AC-4 frames to an MP4 segment.

   The media segment size is determined by packager settings and A/V alignment and segmentation rules.
4. The multiplexer begins the next media segment with the next I-frame.

Related information

Random access point on page 28

2.5 Signaling Dolby AC-4 bitstreams in an ISO base media file format

Dolby AC-4 uses the basic structures defined within ISO/IEC 14496-12 to signal audio tracks and uses specific extensions within compliance to ISO base media file format to provide detailed information on the characteristics of a Dolby AC-4 bitstream.

The information is included in the `stbl` box of the `moov` box of the ISO base media file. The locations and hierarchy of the boxes that must be included in `stbl` to identify a Dolby AC-4 stream within an ISO base media file are listed in the table.

Table 1: Sample table box hierarchy for Dolby AC-4 audio tracks

Nesting level				
4	**5**	**6**	**7**	**Reference**
`stbl`				ISO/IEC 14496-12
	`stsd`			
		`AC4SampleEntry` (header type set to ac−4 or enca)		ETSI TS 103 190-2

Table 1: Sample table box hierarchy for Dolby AC-4 audio tracks (continued)

Nesting level				
4	**5**	**6**	**7**	**Reference**
			AC4SpecificBox (header type set to dac4)	ETSI TS 103 190-2
			Zero or more AC4PresentationLabelBox (header type set to lac4)	ETSI TS 103 190-2
	stts			ISO/IEC 14496-12
	stsc			
	stsz			
	stz2			
	stco			
	co64			

In this table, the value of the nesting level provided for each box is based on the structure of the complete media file, beginning with a nesting value of 0 for the ftyp and moov boxes. The table also includes a reference to where each box is defined.

The value for the AC4SampleEntry box header type depends on whether the file is encrypted. For an unencrypted file, the AC4SampleEntry box header type value is ac−4; for an encrypted file, the AC4SampleEntry box header type value is enca.

The value of the AC4SpecificBox header type is dac4.

The stbl box of a Dolby AC-4 audio track must contain a sync sample box (stss), unless all samples are sync samples (I-frames). The stss box must reference all the sync samples. Random access points are signaled by the sync sample box (stss) in an ISO base media file, or by setting the tr_flags in the track run box in ISO base media segments.

Unless otherwise stated here, all requirements specified in ISO/IEC 14496-12 and Annex E of ETSI TS 103 190-2 must be followed.

2.5.1 Signaling immersive stereo content

For every presentation in the ac4_toc signaling immersive stereo content, an additional regular stereo-compatible presentation must be included in the ac4_dsi_v1.

Dolby AC-4 bitstreams signal immersive stereo content by setting the following fields:

- presentation_version is set to 2.
- b_presentation_id is set to 1.
- presentation_id is any nonnegative number.
- channel_mode is set to 0b1111001 (signaling Dolby Atmos content in IMS) or 0b1111000 (any content in IMS), with ch_mode set to 1.
- b_pre_virtualized is set to 0.

To signal immersive stereo in the ISO base media file format, the ac4_dsi_v1 must include two ac4_presentation_v1_dsi instances for each immersive stereo presentation.

1. One instance uses presentation version 2 and signals the immersive stereo content. The corresponding field settings are as follows:
   - presentation_version is set to 2.

- b_presentation_id is set to 1.
- presentation_id is set to the same number as in the corresponding ac4_presentation_info of the Dolby AC-4 bitstream.
- dsi_presentation_ch_mode is set to 1.
- b_pre_virtualized is set to 1.
- If channel_mode for the Dolby AC-4 bitstreams is set to 0b1111001, dolby_atmos_indicator is set to 1; otherwise, dolby_atmos_indicator is set to 0.

2. A second instance uses presentation version 1 and signals a regular stereo-compatible presentation of the immersive stereo content. This ac4_presentation_v1_dsi instance must come after the instance that signals the immersive stereo content. The corresponding field settings are as follows:
   - presentation_version is set to 1.
   - b_presentation_id is set to 1.
   - presentation_id is set to the same number as in the corresponding ac4_presentation_info of the Dolby AC-4 bitstream.
   - dsi_presentation_ch_mode is set to 1.
   - b_pre_virtualized is set to 0.
   - dolby_atmos_indicator is set to 0.

Note: Signaling in the Media Presentation Description (MPD) file is based on the first instance only.

Signaling immersive stereo with Dolby Atmos

If the ac4_toc of a Dolby AC-4 bitstream contains an immersive stereo presentation (presentation_version = 2) and that presentation contains an AC-4 substream with channel_mode set to 0b1111001, then this immersive stereo presentation is created from Dolby Atmos content and can be signaled as Dolby Atmos.

To signal Dolby Atmos, the ac4_presentation_v1_dsi defined in ETSI TS 103 190-2 is extended, as highlighted in the following table. Dolby Atmos is signaled by setting the dolby_atmos_indicator to 1.

The extension to ac4_presentation_v1_dsi includes additional fields for signalling Dolby AC-4 capabilities.

- The de_indicator field indicates that the audio presentation includes dialogue enhancement metadata and must be set to 1
- The extended_presentation_id field is an optional extension to the presentation_id to increase its value range beyond 31. If the extension is needed, the b_extended_presentation_id field must be set to 1; otherwise, the b_extended_presentation_id field must be set to 0. For index values greater than 31, the corresponding value is signalled using only the 9 bits of the extended_presentation_id, and the bits of the presentation_id must be ignored when the bitstream is read.

Syntax	Word size (in bits)
ac4_presentation_v1_dsi(pres_bytes)	
{	
presentation_config_v1.....	5
if (presentation_config_v1 == 0x06)	
{	
b_add_emdf_substreams = 1;	
}	
else	

Syntax	Word size (in bits)
mdcompat;.....	3
b_presentation_id;.....	1
if (b_presentation_id)	
{	
presentation_id;.....	5
}	
dsi_frame_rate_multiply_info;.....	2
dsi_frame_rate_fraction_info;.....	2
presentation_emdf_version;.....	5
presentation_key_id;.....	10
b_presentation_channel_coded;.....	1
if (b_presentation_channel_coded)	
{	
dsi_presentation_ch_mode;.....	5
if (dsi_presentation_channel_mode in [11, 12, 13, 14])	
{	
pres_b_4_back_channels_present;.....	1
pres_top_channel_pairs;.....	2
}	
presentation_channel_mask_v1;.....	24
}	
b_presentation_core_differs;.....	1
if (b_presentation_core_differs)	
{	
add_pres_bytes;.....	16
pres_bytes += add_pres_bytes;	
}	
b_presentation_core_channel_coded;.....	1
if (b_presentation_core_channel_coded)	
{	
dsi_presentation_channel_mode_core;....	2
}	
}	
b_presentation_filter;.....	1
if (b_presentation_filter;)	
{	
b_enable_presentation;.....	1

Syntax	Word size (in bits)
n_filter_bytes;.....	8
for (i = 0; i < n_filter_bytes; i++)	
{	
filter_data;.....	8
}	
}	
if (presentation_config_v1 == 0x1f)	
{	
ac4_substream_group_dsi();	
}	
else	
{	
b_multi_pid;.....	1
if (presentation_config_v1 in [0, 1, 2])	
{	
ac4_substream_group_dsi();	
ac4_substream_group_dsi();	
}	
if (presentation_config_v1 in [3, 4])	
{	
ac4_substream_group_dsi();	
ac4_substream_group_dsi();	
ac4_substream_group_dsi();	
}	
if (presentation_config_v1 == 5)	
n_substream_groups_minus2;.....	3
n_substream_groups = n_substream_groups_minus2 + 2;	
for (sg = 0; sg < n_substream_groups; sg++)	
{	
ac4_substream_group_dsi();	
}	
}	
if (presentation_config_v1 > 5)	
n_skip_bytes;.....	7
for (i = 0; i < n_skip_bytes; i++)	
skip_data;.....	8
}	

Syntax	Word size (in bits)
}	
}	
b_pre_virtualized;.....	1
b_add_emdf_substreams;.....	1
}	
if (b_add_emdf_substreams)	
{	
n_add_emdf_substreams;.....	7
for (j = 0; j < n_add_emdf_substreams; j++)	
{	
substream_emdf_version;.....	5
substream_key_id;.....	10
}	
}	
b_presentation_bitrate_info;.....	1
if (b_presentation_bitrate_info)	
{	
ac4_bitrate_dsi();	
}	
b_alternative;.....	1
if (b_alternative)	
{	
byte_align;.....	0...7
alternative_info();	
}	
byte_align;.....	0...7
if (bits_read() <= (pres_bytes - 1) * 8)	
{	
de_indicator;.....	1
dolby_atmos_indicator;.....	1
reserved;.....	4
b_extended_presentation_id;..	1
if (b_extended_presentation_id)	
{	
extended_presentation_id;.....	9
}	
else	

Syntax	Word size (in bits)
{	
reserved;.....	4
}	
}	
}	

Note: The number of bits in byte_align pads the number of bits, counted from the start of ac4_presentation_v1_dsi to a multiple of eight.

2.6 Contents of AC4SpecificBox

The AC4SpecificBox contains the AC-4 decoder-specific information (ac4_dsi). The content of the AC4SpecificBox is derived from the AC-4 table of contents (ac4_toc) included in the Dolby AC-4 frame.

The process of deriving the contents of the ac4_dsi from the ac4_toc is described in section E.6 of ETSI TS 103 190-2.

Note: To signal immersive stereo content, the ac4_presentation_v1_dsi defined in ETSI TS 103 190-2 is extended. For detailed information, see *AC-4 DSI syntax changes to signal immersive stereo content*.

2.7 A/V alignment and segmentation

Dolby AC-4 can adapt its frame rate to match commonly used video frame rates (for example, 23.976, 25, and 29.97 fps). Therefore, Dolby AC-4 frames and corresponding video access units can maintain temporal alignment so long as the same frame rate is used in both the audio and video encoders.

Temporally aligned audio and video are especially important at the beginning and particularly at the end of segments when trying to achieve seamless transition for certain use cases (such as ad insertion) without introducing artifacts through audio gaps or additional audio.

Dolby AC-4 I-frames should be placed temporally aligned with the I-frames of the video to enable seamless switching. Most importantly, the first I-frames in the video and audio segments should be temporally aligned. It is acceptable that the succeeding I-frames in the corresponding segments are not aligned.

In the following three figures, the segment length is two seconds. For the first audio and video pair, the I-frame intervals of both the video and audio streams are one second. The I-frames, frame rates, and segment sizes are all aligned. For the second and third audio and video pairs, the I-frame intervals for the video streams are two seconds, while the I-frame intervals for the audio streams are one second. The second I-frame in an audio segment is not aligned with any video I-frame. All of the three examples are suitable as all the first I-frames in video and audio segments are temporally aligned.

Figure 1: Alignment of I-frame intervals, frame rates, and segment sizes

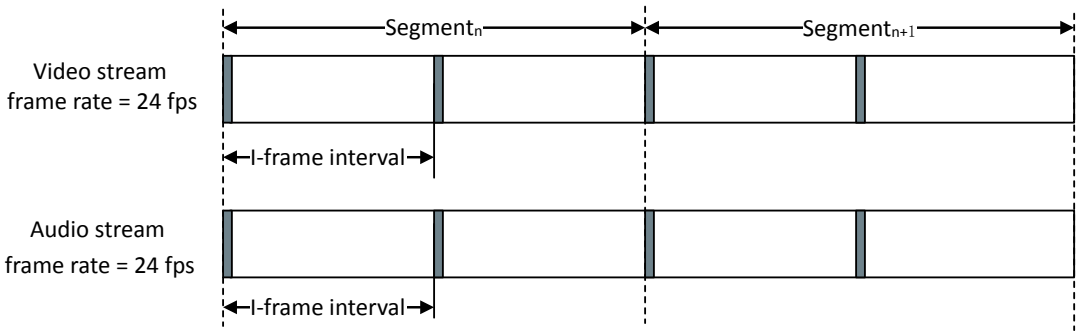


Figure 2: Alignment of frame rates and segment sizes

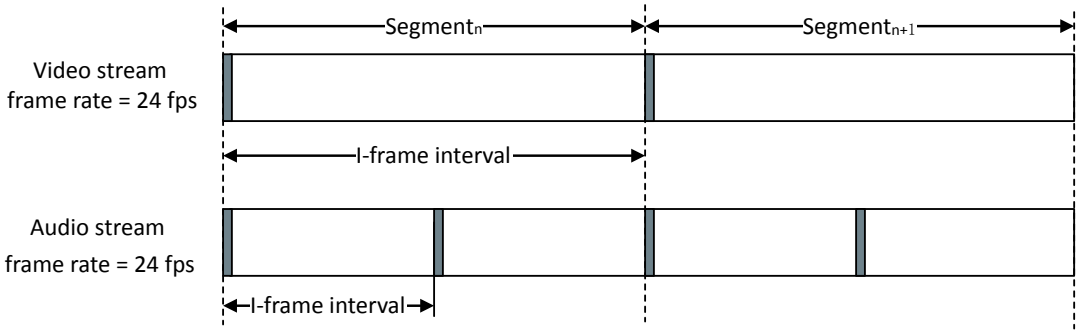
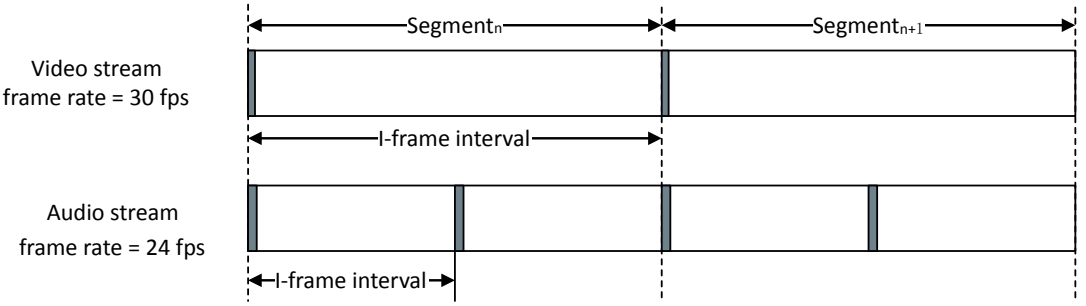
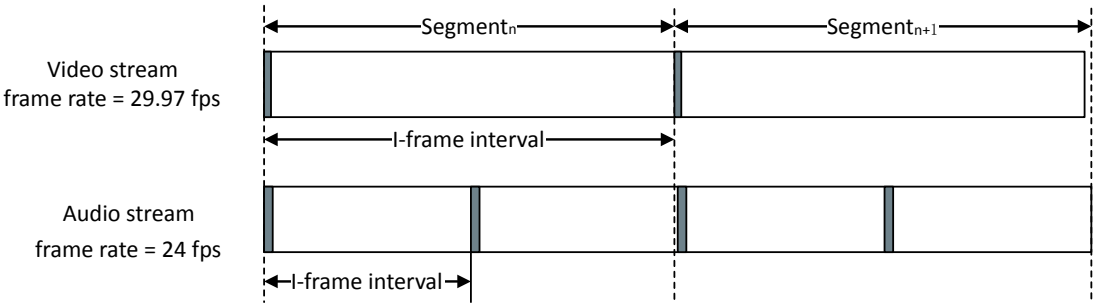


Figure 3: Alignment of segment sizes



The following figure shows an example where the first I-frames of the video segment$_{n+1}$ and audio segment$_{n+1}$ are not aligned, and thus this user case is not recommended.

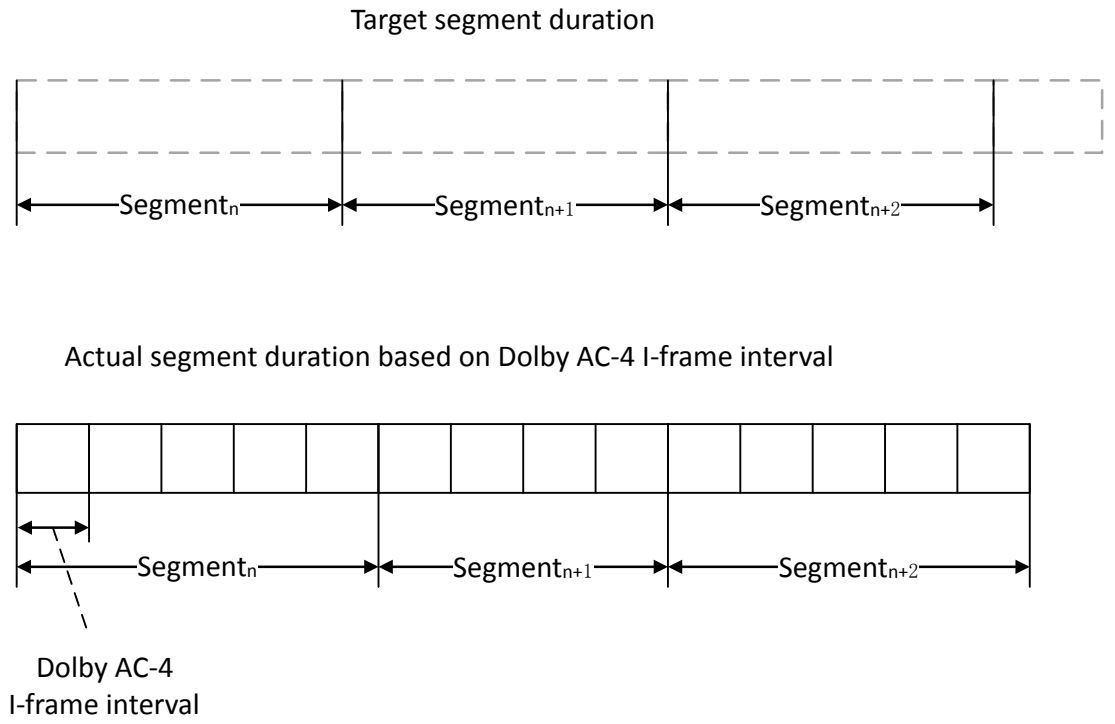
Figure 4: I-frames in segment$_{n+1}$ are not aligned



If frame alignment cannot be achieved, or if the target segment duration is not an integer multiple of the Dolby AC-4 I-frame interval, Dolby AC-4 segment durations are allowed to fluctuate to maintain close alignment with video segments or the target segment duration. The following figure shows an example where an Dolby AC-4 segment (segment$_{(n+1)}$ in the figure) is shorter by one Dolby AC-4 I-frame interval to maintain close segment alignment

Note: The actual segment length must not exceed the target duration by more than 0.5 seconds.

Figure 5: Variable Dolby AC-4 segment durations to maintain segment alignment



In an HLS playlist, the `#EXT-X-TARGETDURATION` parameter is used to determine the target segment duration. The `#EXTINF` parameter indicates the actual length of each segment, as shown in the following example.

```
#EXTM3U
#EXT-X-TARGETDURATION:8
#EXT-X-VERSION:7
#EXT-X-MEDIA-SEQUENCE:1
#EXT-X-PLAYLIST-TYPE:VOD
#EXT-X-INDEPENDENT-SEGMENTS
#EXT-X-MAP:URI="main.mp4",BYTERANGE="1118@0"
#EXTINF:7.98333,
#EXT-X-BYTERANGE:1700094@1118
main.mp4
#EXTINF:8.00000,
#EXT-X-BYTERANGE:1789481@1701212
main.mp4
#EXTINF:8.00000,
#EXT-X-BYTERANGE:1777588@3490693
main.mp4
#EXTINF:8.00000,
#EXT-X-BYTERANGE:1752144@5268281
main.mp4
#EXTINF:7.26667,
#EXT-X-BYTERANGE:1563219@7020425
main.mp4
#EXTINF:8.00000,
#EXT-X-BYTERANGE:1801953@8583644
```

2.7.1 Recommended frame rates

Recommended Dolby AC-4 frame rates are listed in following table.

Highest video frame rate used (in fps)	**Recommended audio frame rates (in fps, in the order of preference)**
120	30, 24 [a], 25 [a]
119.88	29.97
100	25, 24 [a]
60	30, 24 [a], 25 [a]
59.94	29.97
50	25, 24 [a]
48	24, 25 [a]
30	30, 24 [a], 25 [a]
29.97	29.97
25	25, 24 [a]
24	24, 25 [a]
23.976	23.44 (native)

[a] Recommended only if perfect alignment of audio and video segments can be achieved using the frame rate.

3

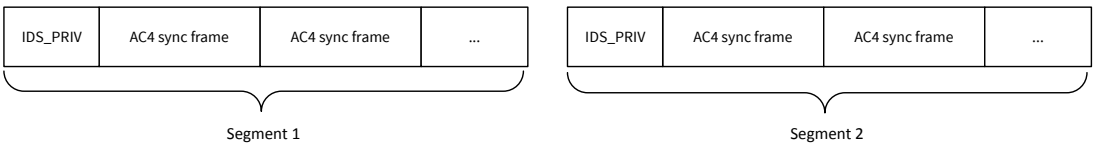
Dolby AC-4 packetization into HTTP Live Streaming compliant audio-only elementary stream segments

When multiplexing a Dolby AC-4 audio elementary stream into HTTP Live Streaming compliant audio-only elementary stream segments, the product must meet certain conditions.

- Use Dolby AC-4 Sync frames.
- Ensure that the first Dolby AC-4 sync frame packaged in each audio elementary stream segment is a Random access point.
- Each audio elementary stream segment must signal the time stamp of its first sample with an `ID3 PRIV` tag at the beginning of the segment.
- The segments should have a file extension of *.ac4*.

The structure of an audio-only elementary stream segment is shown in the figure.

Figure 6: Audio-only elementary stream segments



The syntax of the `ID3 PRIV` tag is described in this table.

Syntax	Word Size in Bits	Identifier	Value
`ID3_tag(){`			
`ID3V2_header()`			
`PRIV_frame()`			
`}`			
`ID3V2_header(){`			
`ID3V2_file_identifier`	`3x8`	`string`	`“ID3”`
`ID3V2_version`	`2x8`	`uimsbf`	`0x0400`
`ID3V2_flags`	`8`	`bslbf`	`0x00`
`ID3V2_size`	`4x8`	`bslbf`	`4x0b0xxxxxxx(0x0000003f)`
`}`			
`PRIV_frame(){`			

Syntax	Word Size in Bits	Identifier	Value
frame_ID	4x8	string	"PRIV"
frame_size	4x8	uimsbf	4x0b0xxxxxxx(0x00000035)
frame_status_flags	8	bslbf	0x00
frame_format_flags	8	bslbf	0x00
Owner_identifier	45x8	string	"com.apple.streaming. transportStreamTimestamp"
Private_data	8x8	bslbf	The upper 31 bits are set to zero,the lower 33 bits is the pts value
}			

4

HTTP Live Streaming playlist files with Dolby AC-4

Requirements and recommendations for a playlist that lists entries for a presentation that includes Dolby AC-4 content are described.

4.1 CHANNELS attribute

All EXT-X-MEDIA parameters for audio components should have a CHANNELS attribute.

The value of the CHANNELS attribute is the count of the audio channels as a decimal-integer. For example, the EXT-X-MEDIA parameter for a 5.1-channel Dolby AC-4 audio component should have a CHANNELS attribute with the value of 6.

4.2 Codec type indication for Dolby AC-4

The codec type must be indicated in an HLS playlist.

For media streams containing Dolby AC-4, we recommend using the CODEC attribute of the EXT-X-STREAM-INF parameter to indicate the codec type. The CODEC must include a dot-separated four-part value for media presentations using Dolby AC-4, as shown in this example:

```
#EXT-X-STREAM-INF:BANDWIDTH=3464568,CODECS="avc1.640028,ac-4.02.01.01"
example.m3u8
```

The FourCC ac-4 part indicates that the codec is Dolby AC-4. The latter three parts are represented by two-digital hexadecimal numbers and indicate these parameters sequentially:

- The bitstream_version as indicated in the ac4_toc(), which is defined in ETSI TS 103 190-2
- The presentation_version as indicated for the presentation in the ac4_toc(), which is defined in ETSI TS 103 190-2
- The mdcompat parameter as indicated for the presentation in the ac4_toc(), which is defined in ETSI TS 103 190-2

An object type indicator (OTI) value must not be appended to the ac-4 string. [1]

4.3 Codec type indication for immersive stereo content

For HTTP Live Streaming streams containing Dolby AC-4 tracks with immersive stereo content, the CODEC attribute of the EXT-X-STREAM-INF parameter follows the same requirements for regular Dolby AC-4 content. In addition, a CHANNELS attribute must be included in the #EXT-XMEDIA parameter.

The CHANNELS attribute must contain an IMSA identifier to indicate immersive stereo content. For immersive stereo content other than Dolby Atmos, the value of the CHANNELS attribute is 2/IMSA; for Dolby Atmos immersive stereo content, the recommended value of the CHANNELS attribute is 2/IMSA,ATMOS.

For example, for a media presentation containing a Dolby AC-4 bitstream with immersive stereo content, the CHANNELS attribute of the #EXT-X-MEDIA parameter must include the IMSA identifier.

```
#EXT-X-MEDIA:TYPE=AUDIO,GROUP-
ID="ac4_IMS",NAME="English",LANGUAGE="en",DEFAULT=YES,AUTOSELECT=YES,
CHANNELS="2/IMSA",URI="audio_ims_index.m3u8"
```

4.4 Examples of HTTP Live Streaming playlist files with Dolby AC-4

Examples of HTTP Live Streaming playlist files that contain Dolby AC-4 bitstreams are provided.

The following example is a master playlist listing two media playlists. One playlist contains an English Dolby AC-4 audio bitstream. The second playlist contains a French Dolby AC-4 audio bitstream.

[1] The object type indicator value is not defined for Dolby AC-4.

```
#EXTM3U
#EXT-X-VERSION:4
#EXT-X-MEDIA:TYPE=AUDIO,GROUP-ID=“ac4",NAME="English",
DEFAULT=YES,AUTOSELECT=YES,LANGUAGE="en",CHANNELS="6",URI="main/english-ac4.m3u8"
#EXT-X-MEDIA:TYPE=AUDIO,GROUP-ID=“ac4",NAME=“Français",
DEFAULT=YES,AUTOSELECT=YES,LANGUAGE=“fr",CHANNELS="6",URI=“main/french-ac4.m3u8"
#EXT-X-STREAM-INF:PROGRAM-
ID=1,BANDWIDTH=2077692,CODECS="avc1.42c01e,ac-4",AUDIO=“ac4",RESOLUTION=640x480
example.m3u8
```

The following example is a playlist containing an English Dolby AC-4 audio bitstream with Dolby Atmos immersive stereo content.

```
#EXTM3U
#EXT-X-VERSION:7
#EXT-X-INDEPENDENT-SEGMENTS
#EXT-X-MEDIA:TYPE=AUDIO,GROUP-
ID="ac4_IMS",NAME="English",LANGUAGE="en",DEFAULT=YES,AUTOSELECT=YES,CHANNELS="2/
IMSA,ATMOS",URI="audio_ims_index.m3u8"
#EXT-X-STREAM-INF:AVERAGE-
BANDWIDTH=6303166,BANDWIDTH=11265828,CODECS="avc1.640028,ac-4",RESOLUTION=1920x1080,FRAME
-RATE=25.000,CLOSED-CAPTIONS=NONE,AUDIO="ac4_IMS"
video_6000k/prog_index.m3u8
```

5

Demultiplexing Dolby AC-4 bitstreams from an ISO base media file segment

Buffering must be considered when product designers implement an ISO base media file format file demultiplexer that supports Dolby AC-4 bitstreams.

5.1 Buffering considerations

The size of the Dolby AC-4 audio main buffer in the receiver is assumed to be 120 KB per instance of the Dolby AC-4 decoder. (That is, for simultaneous decoding of two Dolby AC-4 bitstreams, each Dolby AC-4 decoder must have a main buffer 120 KB in size). The Dolby AC-4 bitstream encoding must be constrained so that a buffer of this size is always sufficient.

The 120 KB buffer size originates from the idea that Dolby AC-4 needs to buffer at most six frames until it can start decoding (according to the buffer signaling model). Some jitter is always possible, so one extra frame is added for safety, which then totals seven. The required buffer size is dependent on the bit rate and the frame rate. Therefore, the minimum buffer size can be calculated with the equation:

$N = 7 \times R_{avg} / F_{frame}$

In this equation, R_{avg} and F_{frame} are the total bit rate and the frame rate of a Dolby AC-4 stream. For example, the data rate is 1,536 kbps and the frame rate is 23.976 fps, which makes the minimum buffer size 56.056 KB.

Note: The specified buffer size takes additional overhead into account; this is not reflected in the equation.

6

HTTP Live Streaming and Dolby AC-4 overview

Introduction about HLS and Dolby AC-4

- HTTP Live Streaming
- Raw AC-4 frame
- Random access point
- Sync frames

6.1 HTTP Live Streaming

HLS is a protocol developed by Apple for transferring unbounded streams of multimedia data. HLS works by breaking the content into a sequence of small HTTP-based file segments, each of which can be downloaded independently. Alternative segments at different data rates are available, allowing streaming to adapt to different network conditions.

6.1.1 HTTP Live Streaming playlist files

An HTTP Live Streaming playlist file provides information for a playback device about the available content for a multimedia presentation.

Specifically, HTTP Live Streaming playlist files provide information about:

- The media segments that comprise the multimedia presentation
- Available content renditions that can be used in the content selection process (for example, different languages)
- Available content variants (for example, content rendered for different bandwidths)

HTTP Live Streaming playlist files are regular M3U playlists, extended by the addition of information specific to HTTP Live Streaming. This extended information is contained in lines that start with an #EXT prefix inside the playlist. A playlist contains uniform resource identifiers (URIs) that point to media files or to other playlists.

The HTTP Live Streaming protocol defines a specific type of playlist, known as a variant playlist, to allow a client device to select from different versions of the same piece of content. For example, these versions can be encoded at different bit rates to enable the client to switch to a lower data rate (for example, when the available delivery bandwidth is reduced) or to switch to a higher data rate to improve audio and video quality. A variant playlist may also point to alternative content for a presentation, such as an alternative language version of the content.

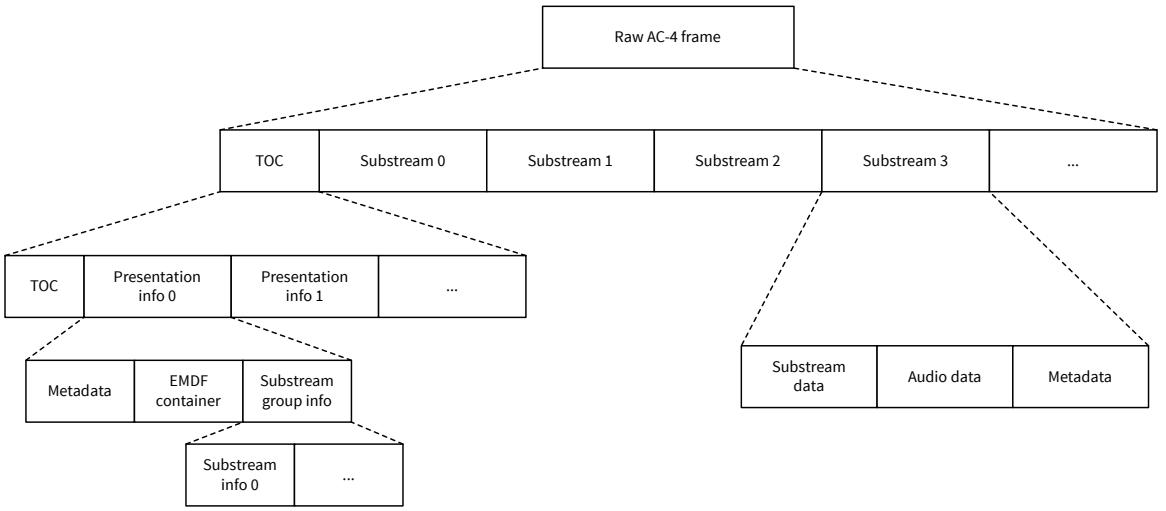
6.2 Raw AC-4 frame

Each raw AC-4 frame is composed of a table of contents and at least one substream.

The table of contents can be considered as the bitstream inventory where all important information for the overlaying system resides. Each table of contents contains the table-of-contents data and at least one presentation. An audio presentation informs the decoder about which parts of an AC-4 stream are intended to be played back simultaneously at a given point in time. Presentations consist of substream groups. Substream groups consist of one or more individual substreams. The substream group carries properties common to all substreams contained in the substream group. Substreams in one substream group are either all channel coded or all object coded. Substreams are decodable units that represent a specific channel configuration (mono, stereo, 5.1, or 7.1). All of the payload information (such as audio data and metadata) is stored in substreams.

The raw AC-4 frame structure is shown in the following figure, as described in ETSI TS 103 190-2.

Figure 7: Raw AC-4 frame structure



6.3 Random access point

The Dolby AC-4 decoder features a seamless switch between Dolby AC-4 elementary streams of the same media content and different configuration options.

Configuration changes can include:

- Bit-rate changes
- Channel-mode changes
- Sampling-rate changes where the higher sampling rate is an integer multiple of the lower sampling rate (for example, from 96 to 48 kHz)
- Frame-rate changes where the higher frame rate is a factor of two or four times the lower frame rate (for example, from 48 to 24 fps and vice versa)
- Any combinations of these

A seamless switch means that the audio output has no audible artifacts or distortion introduced by audio gaps or additional audio during and after switching.

A seamless switch can be accomplished only at a random access point. Comparable to random access frames in H.264 video streams, the Dolby AC-4 elementary stream contains I-frames to indicate random access points. The `b_iframe_global` flag inside the table of contents indicates an I-frame. Unlike video I-frames, audio I-frames are for signaling only. Audio can still be decoded if an audio I-frame is not present at a random access point. When a switch is not immediately followed by an I-frame, the Dolby AC-4 decoder performs partial decoding until it receives the next I-frame. Partial decoding may lead to degraded audio quality, including short periods of silence, limited audio bandwidth, or a downmix of the content.

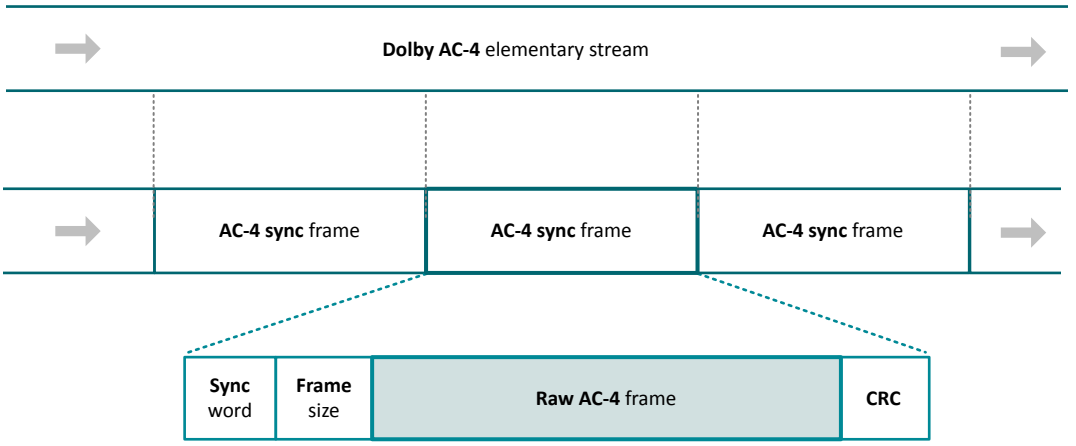
6.4 Sync frames

A Dolby AC-4 bitstream in the AC-4 sync frame transport format (as specified in ETSI TS 103 190) consists of Dolby AC-4 synchronization frames. The AC-4 sync frame transport format must be used for MPEG packed audio.

Each frame begins with a sync word and optionally ends with a CRC word. The sync word allows a decoder to easily identify frame boundaries and begin decoding. The CRC word allows a decoder to detect the occurrence of bitstream errors and to perform error concealment when it detects an error. The actual codec frame is referred to as the raw AC-4 frame.

This figure shows the general structure of the AC-4 frames as specified in ETSI TS 103 190.

Figure 8: The AC-4 sync frame transport format



Possible values for the sync word are:

- `0xAC40`: No CRC field is included.
- `0xAC41`: A CRC field is included.

The CRC is based on the raw AC-4 frame. The frame size and sync word fields are not part of the CRC calculation.

The frame size signals the size of the raw AC-4 frame and can be either two or five bytes in size. When a five-byte frame size field is used, the first two bytes are set as `0xFF` and the actual size value is written into the remaining three bytes.

Each AC-4 sync frame element with the sync word `0xAC41` can contain a 16-bit CRC word. The CRC applies to the complete AC-4 sync frame element, not including the sync word. Decoding of the CRC word allows errors to be detected.

Glossary

bslbf
Bit string, left bit first. A bit string whose elements are written from left to right.

CBR
Constant bit rate. Describes an encoding method that produces output data at a constant rate.

CRC
Cyclic redundancy check.

Dolby AC-4
An audio coding system that enables high audio quality at very low bit rates. It consists of a channel-based audio codec (lossy, low bit-rate audio codec), object-based audio extensions, and system integration components.

elementary stream
A bitstream that is the output of an audio or video encoder and contains only one type of data, such as audio or video.

HLS
HTTP Live Streaming. An adaptive streaming protocol developed by Apple for delivery of media content in various software environments.

immersive stereo
A technology that delivers a virtualized immersive experience to headphones or stereo speakers through a Dolby AC-4 bitstream with appropriate stereo content and metadata that converts the stereo signal into the virtualized experience.

I-frame
Independently decodable frame. An I-frame is a single frame that contains all the data needed to decode the frame. Dolby AC-4 uses the concept of I-frames to achieve greater coding efficiency. An I-frame is indicated by the value of the `b_iframe_global` bitstream parameter.

MPEG-2 transport stream
As defined in ISO/IEC 13818-1, a packetized bitstream that is used to transmit audio and video information. An MPEG-2 transport stream is made up of multiplexed program elementary streams.

MPEG-DASH
MPEG Dynamic Adaptive Streaming over HTTP. An adaptive bit-rate streaming protocol that enables high-quality streaming of media content over the Internet delivered from HTTP.

OTI
Object type indicator.

PAT
Program Association Table. Program Association Table of an MPEG-2 transport stream.

PCM
Pulse code modulation. A method that is used to convert analog signals into digital, binary, coded pulses by sampling the analog signal, quantizing each sample independently, and converting the resulting quantized values into a digital signal.

PCR
Program clock reference. A periodically transmitted value of 42 bits that provides a sample of the system time clock in the encoder and which is used to properly demultiplex packets and to ensure that audio and video are synchronized.

PES
Packetized elementary stream. An elementary stream that is split into small chunks (packets) for transmitting and combining multiple streams within a transport stream. Each PES is identified by a unique packet identifier (PID).

PID
Packet identifier. A unique code that identifies a packetized elementary stream (PES) within a transport stream. The PID is contained in the transport stream packet header and is listed in the service information (SI) for a transport stream.

PMT
Program Map Table. A table within an MPEG-2 transport stream that defines the set of elementary streams associated with a specific program.

presentation
References to AC-4 substreams to be decoded and presented simultaneously.

presentation configuration
Set of metadata to describe how a presentation must be decoded.

PTS
Presentation time stamp. The presentation time stamp is contained in the packetized elementary stream (PES) packet header that indicates when an access unit should be decoded and presented for output. The PTS is used in combination with other time stamp parameters to synchronize audio and video.

random access point
Within a media file, a random access point indicates the location at which to begin reading and decoding the coded bitstream.

raw AC-4 frame
The actual codec frame that consists of a table of contents plus several byte-aligned substreams.

substream
A decodable unit that represents a specific channel configuration (mono, stereo, or 5.1) and contains audio data and corresponding metadata.

UIMBSF
Unsigned integer, most-significant bit first.

URI
Uniform Resource Identifier. A group of characters identifying a resource on a network (typically, the Internet).

VBR
Variable bit rate. Describes an encoding method that outputs data at a variable rate.

Dolby Laboratories, Inc. 1275 Market Street, San Francisco, CA 94103-1410 USA.