CTA Specification

Web Application Video Ecosystem – Content Specification

CTA-5001-B

December 2019

Consumer Technology Association™

NOTICE

(Formulated under the cognizance of the CTA **WAVE Project** in cooperation with the W3C; for information please see cta.tech/WAVE.)

Published by
CONSUMER TECHNOLOGY ASSOCIATION
Technology & Standards Department
www.cta.tech

Table of Contents

Figures

Tables

FOREWORD

This document was developed by the Web Application Video Ecosystem (WAVE) Project of the Consumer Technology Association[1]. The WAVE Project is a broad industry initiative of content, technology, infrastructure and device companies, all working together towards commercial Internet video interoperability based on industry standards.

[1] See https://cta.tech/WAVE

Web Application Video Ecosystem – Content Specification

2019 Edition

1 Scope

This Web Application Video Ecosystem Content Specification (WAVE Content Spec) is derived from the ISO/IEC standard, “Common Media Application Format for Segmented Media” [CMAF].

The scope of the WAVE Content Spec is the encoding and packaging of segmented media for delivery and decoding on end user devices in adaptive multimedia presentations. The WAVE Content Spec extends [CMAF] by referencing additional Media Profile specifications.

The WAVE Content Specification defines WAVE Programs that contain a sequence of one or more CMAF Presentations with encoding constraints to enable continuous rendering and user experience on targeted devices. This 2019 Edition additionally defines a Baseline Splice Constraint Profile targeting the majority of existing adaptive streaming devices.

The purpose of the Content Specification is two-fold: firstly, to decouple the preparation of content from the devices which will eventually consume it, so it can be used with different protocols and platforms; and secondly, to assist content distributors in preparing content compatible with the digital media playback devices they wish to target.

WAVE Media Profiles, Presentation Profiles and Splice Constraint Profiles define options allowed in content. Corresponding functionality is required in compatible Web applications and devices to request, decode, decrypt, and render the content as intended. The application, device, and test requirements needed for interoperability are specified in other WAVE specifications.

2 References

2.1 Normative References

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the dated edition applies. For undated references, the latest edition of the referenced document (including any amendments and corrigenda) applies.

NOTE *ISO/IEC 23000-19 is an evolving specification. At the time of this writing amendments have been approved which specify additional Media Profiles and provide corrigenda to the original draft. For those with access to the ISO/IEC JCT SC 29/WG 11 repository, amendments not yet published by ISO are available using the bracketed document number. CTA members may also access unpublished MPEG drafts in the WAVE content specification repository, made possible by the CTA WAVE – ISO/IEC JTC 1/SC 29/WG 11 liaison agreement.*

Additional amendments are anticipated. To ensure Web media delivery interoperability, it is recommended that you download the latest CTA WAVE Content Specification and stay current with ISO/IEC 23000-19 and other referenced specifications, including errata, corrigenda, new editions or amendments.

[AC4] ETSI TS 103 190-2 V1.2.1, *Digital Audio Compression (AC-4) Standard; Part 2: Immersive and personalized audio*, Annex H. http://www.etsi.org/deliver/etsi_ts/103100_103199/10319002/01.02.01_60/ts_10319002v010201p.pdf

[CMAF] ISO/IEC 23000-19, *Information technology — Coding of audio-visual objects — Part 19: Common media application format (CMAF) for segmented media*. 2nd Edition.

[DTS1] ETSI TS 102 114 v1.5.1, *DTS Coherent Acoustics; Core and Extensions with Additional Profiles*, Annex H. http://www.etsi.org/deliver/etsi_ts/102100_102199/102114/01.05.01_60/ts_102114v010501p.pdf

[EAC3] ETSI TS 102 366 V1.4.1 (2017-09), *Digital Audio Compression (AC-3, Enhanced AC-3) Standard*, Annex J. http://www.etsi.org/deliver/etsi_ts/102300_102399/102366/01.04.01_60/ts_102366v010401p.pdf

2.2 Informative References

The following documents contain provisions that, through reference in this text, constitute informative provisions of this document. At the time of publication, the editions indicated were valid. All documents are subject to revision, and parties to agreements based on this document are encouraged to investigate the possibility of applying the most recent editions of the documents listed here.

[CENC] ISO/IEC 23001-7:2016, *Information technology – MPEG systems technologies – Part 7: Common encryption in ISO base media file format files*, https://www.iso.org/standard/68042.html

[DASH] ISO/IEC 23000-1:2019, *Information technology — Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats*, 3rd Edition. https://www.iso.org/standard/75485.html

[EBU3380] *EBU-TT-D Subtitling Distribution Format*, EBU Tech 3380, 30 October 2017, https://tech.ebu.ch/publications/tech3380/.

[EME] *Encrypted Media Extensions*, W3C Recommendation 18 September 2017, http://www.w3.org/TR/encrypted-media/.

[HLS] RFC 8216 "HTTP Live Streaming", August 2017, https://tools.ietf.org/html/rfc8216.

[MSE] *Media Source Extensions*, W3C Recommendation 17 November 2016, http://www.w3.org/TR/media-source/.

[TTMLREG] *TTML Media Type Definition and Profile Registry*, W3C Working Group Note 17, January 2017, http://www.w3.org/TR/ttml-profile-registry/.

[WAVEDPC] CTA-5003, *Web Application Video Ecosystem (WAVE) Device Playback Capabilities*, Consumer Technology Association (CTA), December 2018.

[WAVEWMA] *Web Media API Snapshot 2019*, Community Group Report, 4 December 2019, https://w3c.github.io/webmediaapi/.

[WAVEWMD] *Web Media Application Developer Guidelines*, Community Group Report, 13 December 2018, https://w3c.github.io/webmediaguidelines/.

2.3 Document Notation and Conventions

The following terms are used to specify conformance elements of this specification. These are adopted from the ISO/IEC Directives, Part 2, Annex H [ISO-P2H ISO-P2H]. For more information, please refer to those directives.

- SHALL and SHALL NOT indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.
- 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.
- MAY and NEED NOT indicate a course of action permissible within the limits of the document.

Terms defined to have a specific meaning within this specification will be capitalized, e.g. "Track", and should be interpreted with their general meaning if not capitalized.

NOTE *The descriptions and definitions of Player behavior in gray-box italics are high-level expressions of expected playback functioning, providing both background and motivation for portions of the Content Specification. Those terms and behavior are defined in detail in the WAVE Device Playback Capabilities specification [WAVEDPC] including the exact playback model as well as the requirements for the device in terms of expected observations.*

2.4 Definitions

For the purposes of this document the following terms and definitions apply, also including those defined in [CMAF], Section 3.

Term	Definition
Common Header	Optional encoding using a single CMAF Header associated with all CMAF Tracks in a sequence of Sequential Switching Sets in a WAVE Program.
Continuous Rendering	*Perceptually continuous rendering of Sequential Switching Sets in sequenced CMAF Presentations.*
Event Stream	A sequence of related events. Note that when events are signaled using the Event Message box (‘emsg’), Event Message boxes in the same Event Stream will contain the same 'schemeIdUri' and 'value' field values.
Hypothetical Render Model	*Assumed Player capabilities and constraints used to define the intended result (semantics) of content constraints and combinations, e.g., encoding expected to result in seamless switching and splicing by that render model.*
Media Pipeline	*System that performs parsing, decryption, decoding, synchronization, and rendering of WAVE Programs, independent of the delivery protocol used.*
Player	*Combination of hardware and software, including a Media Pipeline and possibly a Web application, that can stream and render a WAVE Program.*
Sequential Switching Sets	CMAF Switching Sets in a WAVE Program intended for sequential presentation with *Continuous Rendering* and user experience.
Splice	Transition between Sequential Switching Sets in sequenced CMAF Presentations in a WAVE Program.
Splice Point	WAVE Program presentation start time at the start of the first presented video media sample in the second of two adjacent Sequential Switching Sets in a WAVE Program.
WAVE Baseline Splice Constraint Profile	WAVE Splice constraints intended to allow *Continuous Rendering* of Sequential Switching Sets in WAVE Programs on most adaptive streaming *Players*.
WAVE CMFHD Baseline Program	WAVE Program conforming to allowed combinations of CMAF CMFHD Presentation Profiles and the WAVE Baseline Splice Constraint Profile.
WAVE Program	Sequence of one or more CMAF Presentations, if more than one, then all audio and video contained in Sequential Switching Sets, and optionally conforming to a WAVE Splice Constraint Profile.
WAVE Splice Constraint Profile	Constraints on Splices between Sequential Switching Sets in a WAVE Program intended to enable *Continuous Rendering* of sequenced CMAF Presentations.

3 WAVE Content (Informative)

3.1 Introduction

WAVE content that is intended for synchronized *Continuous Rendering* is called a WAVE Program. A WAVE Program contains a sequence of one or more CMAF Presentations specified in [CMAF]. Each CMAF Presentation contains CMAF Switching Sets that each contain one or more synchronized CMAF Tracks that conform to a CMAF Media Profile and are encoded and packaged to enable *Seamless Adaptive Switching*. WAVE Programs can optionally contain CMAF Presentations that conform to CMAF Presentation Profiles, which conditionally require inclusion of CMAF Switching Sets conforming to specified CMAF Media Profiles and encryption. CMAF Presentations conforming to a Presentation Profile can contain additional CMAF Switching Sets that encode the same content in other CMAF Media Profiles.

This specification includes WAVE-approved CMAF Media Profiles and WAVE Programs containing a sequence of CMAF Presentations conforming to WAVE Splice Constraint Profiles.

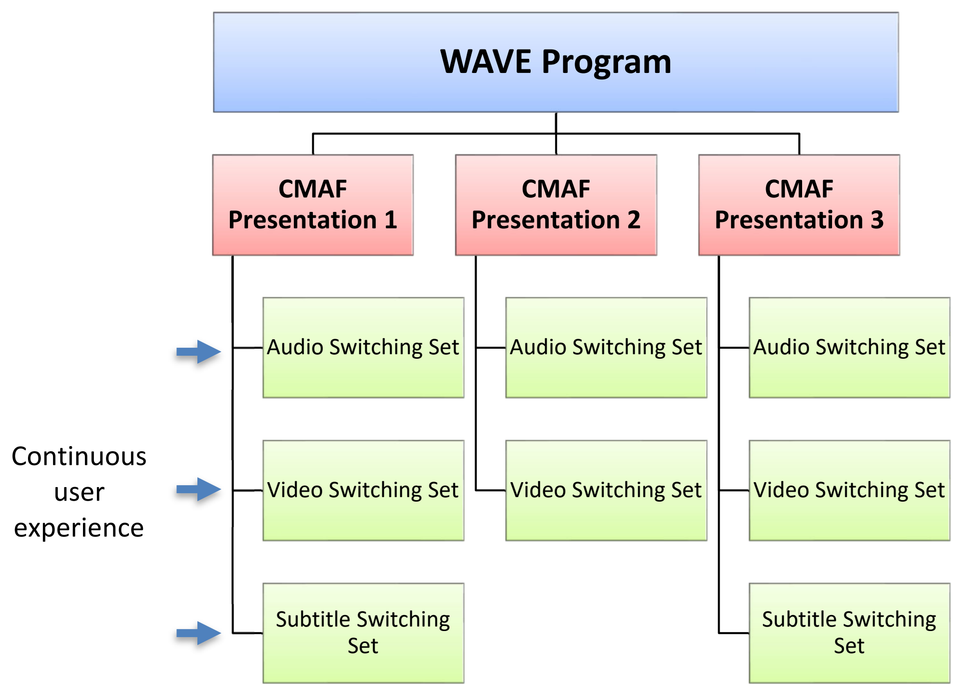

Figure 1 – WAVE Program containing a sequence of multiple CMAF Presentations

CMAF specifies CMAF Switching Sets that contain one or more CMAF Tracks with time aligned CMAF Fragments, typically encoded at different bitrates and qualities. CMAF Fragments with the same decode timestamp are switchable alternatives.

Figure 2 illustrates time aligned CMAF Fragments in a CMAF Switching Set. Fragment decode and presentation times are "continuous", meaning there are no gaps or overlaps in the decode time and duration of CMAF Fragments or the presentation time of the media samples in each CMAF Track.

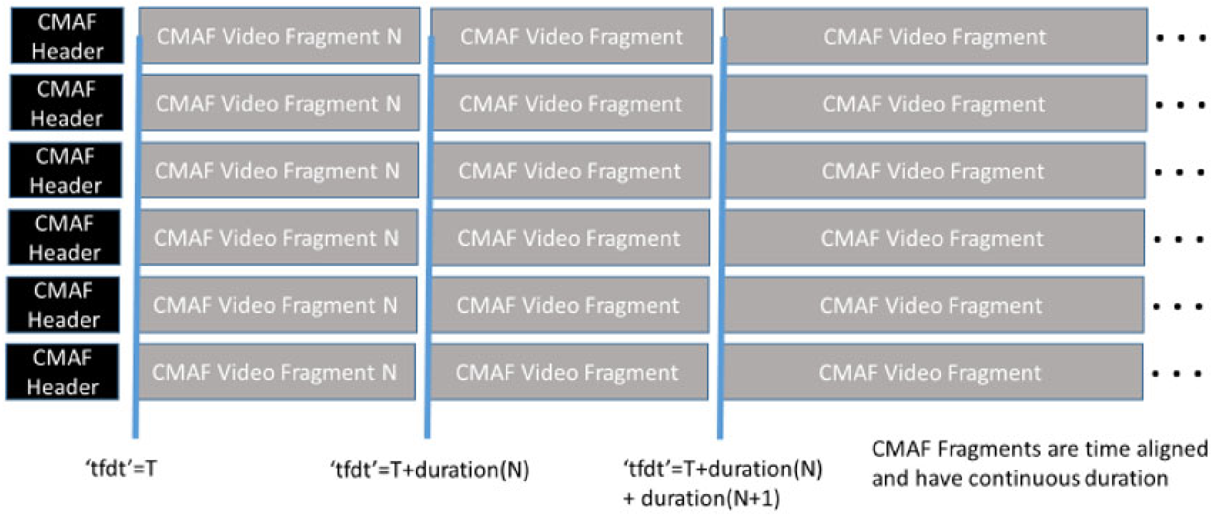

Figure 2 – Video CMAF Switching Set showing alternative Tracks & Fragment time alignment

> **NOTE** *A WAVE Player is expected to be able to adaptively switch alternative CMAF Fragments, i.e., downloaded and sequenced into the same media source buffer and Media Pipeline supporting that CMAF Media Profile for parsing, decoding, decryption, and rendering of perceptually continuous content.*

Video is usually offered with multiple CMAF Tracks in one Switching Set because it has the highest bitrate, but it is also possible to provide multiple audio CMAF Tracks for adaptive bitrate switching. Audio and subtitle Switching Sets often contain only one CMAF Track because adaptive switching is not offered.

In addition, different video, audio and subtitle Switching Sets may be included for selection of different codecs, languages, functionality, etc.

CMAF Switching Sets in a CMAF Presentation with alterative content or encoding are logically grouped into a CMAF Selection Set, for example by referencing them as Adaptation Sets in the same Group in a DASH manifest.

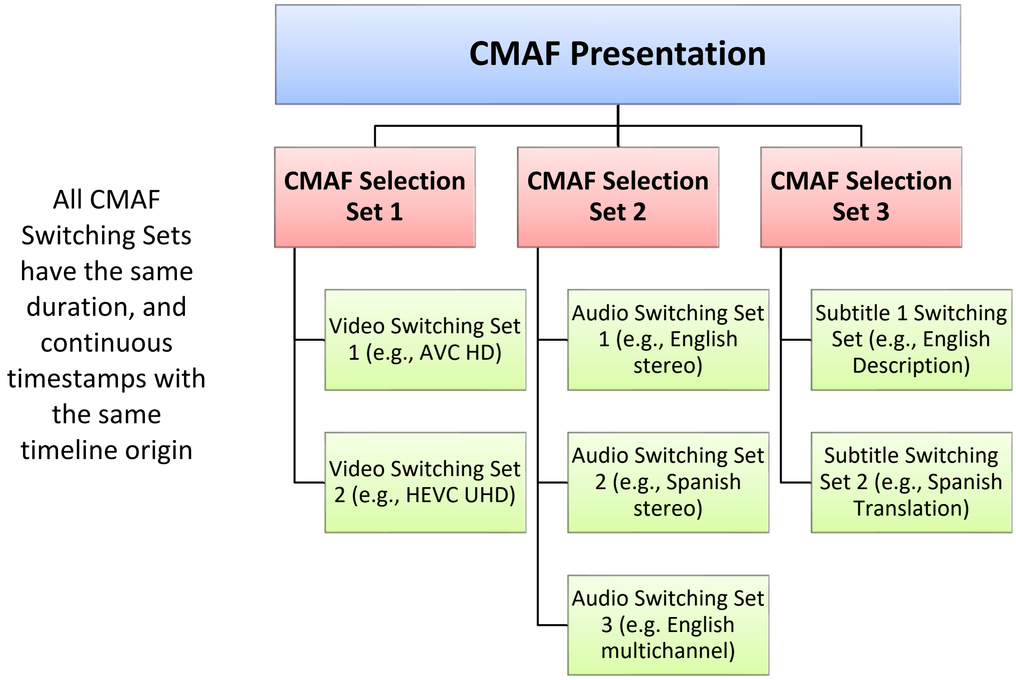

Figure 3 – CMAF Presentation structure

CMAF also specifies CMAF Addressable Media Objects for storage and delivery of CMAF Fragments, including CMAF Track Files, Segments, and Chunks. But CMAF does not constrain the manifest formats or delivery protocols that use them. The media objects that are encoded and decoded are CMAF Headers and CMAF Fragments, regardless of the CMAF Addressable Media Objects used to deliver the CMAF Fragments. This makes encoding and decoding interoperable, regardless of the manifest, the delivery protocol, object addressing, and segmentation used by a streaming *Player*.

The CMAF *Hypothetical Render Model* supports late binding of CMAF Tracks and Switching Sets into synchronized multimedia presentations described by a manifest that identifies CMAF Addressable Media Objects, CMAF Tracks, and their relationships to enable Track selection, synchronization, and adaptive switching.

Figure 4 illustrates the structure of a CMAF video Fragment, and how it can be packaged either as a single CMAF Segment or multiple CMAF Chunks. CMAF Chunks allow short sequences of media samples in a CMAF Fragment to be live encoded, packaged, delivered, and decoded just-in-time, before the entire Fragment has been encoded. The figure illustrates that only the first media sample of a video Fragment (shown in black) is required to be a stream access point for decoding, splicing, etc. – e.g. an IDR picture in AVC and HEVC. Other Addressable Media Objects include a CMAF Segment, containing one or more sequential CMAF Fragments; and a CMAF Track File, containing a CMAF Header and all Fragments of a Track stored as an ISOBMFF file.

CMAF only specifies individual CMAF Presentations. CMAF does not specify encoding or rendering of sequences of CMAF Presentations. WAVE Programs are defined in this document to describe a sequence of CMAF Presentations, and Sequential Switching Sets with Splice Constraints that are intended to permit *Continuous Rendering*.

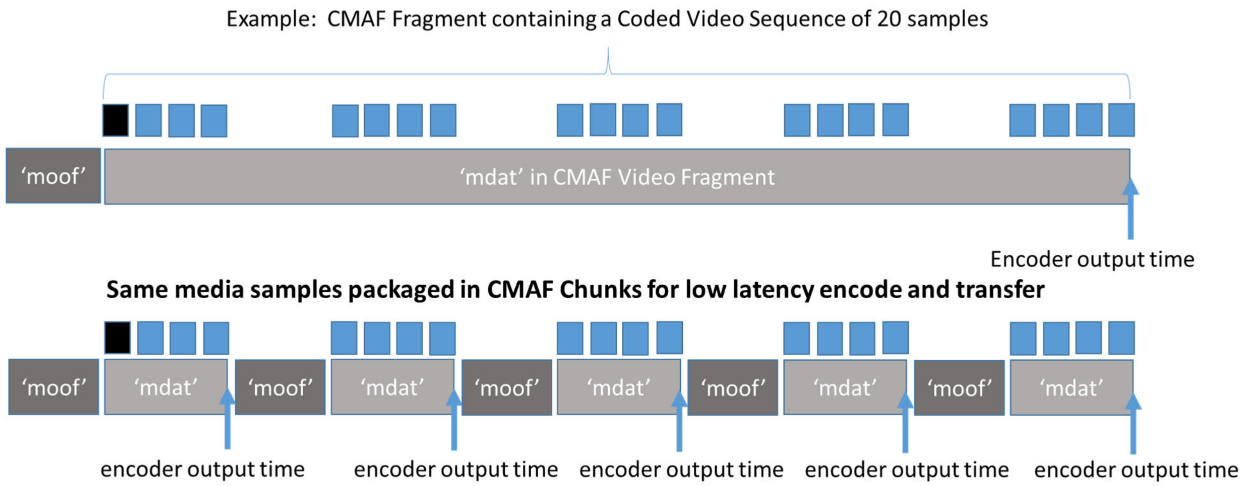

Figure 4 – CMAF Fragment Structure, Packaged in CMAF Chunks for Low Latency Delivery

> **NOTE** *The WAVE Device Playback Specification [WAVEDPC] defines the requirements and conformance metrics for Continuous Rendering of a CMAF Presentation conforming to a specific WAVE Splice Constraint Profile. Clause 5.3 of [WAVEDPC] provides an extensive CMAF content model, detailing addressable objects, presentation timing, track and switching set models.*

3.2 Event Messages and other Prepended Boxes

CMAF, ISOBMFF and DASH specify several boxes that may be included in CMAF Chunks, Fragments, Segments and CMAF Track Files.

These include:

- ISOBMFF Segment Type Box ('styp')
- DASH Event Message Box ('emsg')
- ISOBMFF Producer Reference Time Box ('prft')
- ISOBMFF/DASH Segment Index Box ('sidx')

The details regarding the use of these boxes, such as permitted sequences, are specified in [CMAF]. The above boxes can be prepended to any ISOBMFF segment, i.e. precede any Movie Fragment Box ('moof') followed by a Media Data Box ('mdat').

> **NOTE** *Event Messages Boxes can be repeated over time and over any or all audio and/or video CMAF Tracks. Each event within one Event Stream has a unique ID. Subsequent repetitions of an Event Message Box, indicated by the same ID within one Event Stream, are intended to be ignored. Repetition over time allows the event to be signaled well in advance of its presentation time to give a Player ample time to respond, but also be signaled just before the event's presentation time so the Event Message Box will be detected when a Program is random accessed close to the event's presentation time.*

An Event Message Box should be repeated in all CMAF Tracks within one CMAF Switching Sets so that whichever CMAF Track is downloaded, the event messages will be received.

> **NOTE** *According to CMAF these prepended boxes are only expected to be detected at the start of a CMAF Addressable Media Object. For instance, when these boxes are contained at the start of a CMAF Chunk that isn't first in a Fragment, or a Fragment that isn't first in a Segment (containing multiple Fragments), a Player application isn't expected to parse the entire object to find them. This allows a Player application to look for these specific boxes at the start of each downloaded CMAF Chunk or Segment before it is appended to a source buffer for complete parsing, decryption, and decoding. This limited parsing determines where boxes need to be encoded for detection. When both Chunk addressing (for low latency live) and Segment addressing will be used (for CDN cached Segments or progressive chunk transfer coding of a Segment), prepending a message on a Chunk and a Segment supports both levels of object addressing and parsing.*
>
> *When an application finds an 'emsg' box with a new ID at the start of a downloaded Addressable Media Object, it is expected to pass the box to a handler registered for that Event Message schemeIdUri, which identifies the message syntax and semantics.*

3.3 WAVE Media Profiles

WAVE content includes WAVE-referenced CMAF Media Profiles published in [CMAF] and elsewhere. CMAF defines general requirements for specifying additional CMAF compliant Media Profiles. This makes CMAF extensible and enables support for non-MPEG codecs. Because of this extensibility, some CMAF media profiles are specified by organizations other than ISO/IEC, such as the European Telecommunications Standards Institute (ETSI).

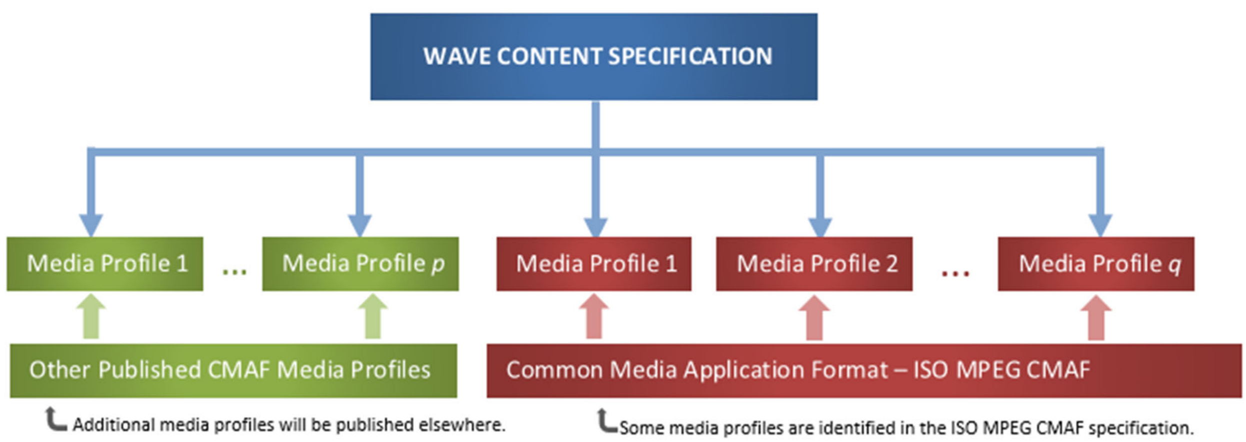

Figure 5 – The WAVE Content spec references published CMAF Media Profiles

3.4 CMAF Presentation Profiles

CMAF specifies three Presentation Profiles for widely supported Media Profiles and Common Encryption schemes [CENC]. CMAF also specifies a general method for defining additional Presentation Profiles that can conditionally require other CMAF Media Profiles. By requiring Media Profiles that must be included when audio, video, and subtitle Tracks are included in a CMAF Presentation Profile, the minimum decoder requirements for interoperability in *Players* and their *Media Pipelines* can be inferred.

CMAF Presentation Profiles provide a content requirement such that *Players* conforming to the Presentation Profile are ensured that they can playback the content. For this purpose, a Presentation Profile defines Media Profiles that are conditionally required in a CMAF Presentation conforming to this profile. Additional Switching Sets conforming to other CMAF Media Profiles may be included in a CMAF Selection Set for selection and playback by devices that support those additional Media Profiles. For instance, an HD resolution Switching Set may be required, but a UHD resolution Switching Set can also be included in the CMAF Selection Set to provide potentially improved experiences for devices and displays capable of UHD resolution.

> ***NOTE*** *Aligned CMAF Switching Sets enable a Player that supports both Media Profiles and their DRM requirements to seamlessly switch CMAF Tracks from both Switching Sets.*

Extensibility and interoperability are accomplished by content and *Players* using the same CMAF Media Profile as signaled by the compatibility brand. The brand’s four-character code can be used in a manifest to identify that CMAF Tracks and CMAF Switching Sets conform to the requirements of that CMAF Media Profile.

3.5 WAVE Programs

By definition, WAVE Programs are a sequence of one or more CMAF Presentations intended for *Continuous Rendering*.

A WAVE Program constrained to a WAVE Splice Constraint Profile allows Players conformant to that WAVE Splice Constraint Profile to play the entire WAVE Program without interruption at Splices. For instance, a live “channel” or a manifest that is updated can consist of a sequence of CMAF Presentations added over time that may initially be unknown but will conform to the WAVE Splice Constraint Profile.

> ***NOTE*** *The constraints in a WAVE Splice Constraint Profile relate to the ability of certain Players or a Hypothetical Render Model conforming to a WAVE Splice Constraint Profile to present content “seamlessly” at Splice points. Content that does not conform to a WAVE Splice Constraint Profile should still be playable by any Player that supports the relevant WAVE Media Profiles but there may be some disruption at Splice Points for some Players.*

4 Media Profiles

4.1 Introduction

Media offered in Switching Sets in WAVE SHALL conform to CMAF Media Profiles as defined by [CMAF] and specified there or in other specifications.

WAVE content SHALL include one or more CMAF Track(s) and Switching Set(s) conforming to at least one WAVE approved CMAF Media Profile for each Selection Set in each CMAF Presentation. CMAF Selection Sets MAY include additional encodings and Switching Sets of the same content components conforming to Media Profiles that are not referenced by this specification.

WAVE content conforming to CMAF Presentation Profiles SHALL include conditionally required CMAF Media Profiles specified by each CMAF Presentation Profile in each CMAF Selection Set for each content component.

> ***NOTE*** *CMAF Media Profiles summarized in Table 1, Table 2, Table 3 and Table 4 include the maximum encoding parameters allowed in conforming content and can allow alternative encoding options (e.g. alternative video sample formats, parameter storage, etc.). Conformant CMAF Tracks often use lower values than the maximum (lower bitrate, resolution, frame rate, etc.), and a subset of the encoding options. Devices that claim conformance to a CMAF Media profile are expected to render at least the maximum parameter values and all encoding options.*

4.2 WAVE Video Media Profiles

4.2.1 General

Each WAVE Video Media Profile SHALL conform to the normative reference listed in Table 1.

Table 1 summarizes the WAVE Video Media Profiles, their brand, example 'codecs' subparameter string(s), important parameter constraints and the Media Profile's normative reference.

Where the Media Profile specifies more than one option for a particular characteristic (e.g., Color primaries, sample entry, etc.), all listed options are valid in content.

Table 1 – WAVE Video Profiles

Media Profile Name	INFORMATIVE Codec	INFORMATIVE Profile	INFORMATIVE Level	INFORMATIVE Color primaries & matrix coefficients	INFORMATIVE Transfer Characteristics	INFORMATIVE 'codecs' MIME subparameters	NORMATIVE CMAF Brand	NORMATIVE Normative Reference
HD	AVC	High	4.0	1 (BT.709)	1 (BT.709 OETF)	avc1.640028 avc3.640028	'cfhd'	[CMAF] Table A.1
HHD10	HEVC	Main10 MainTier	4.1	1 (BT.709)	1 (BT.709)	hev1.2.4.L123.B0 hvc1.2.4.L123.B0	'chh1'	[CMAF] Table B.1

Media Profile Name	INFORMATIVE Codec	INFORMATIVE Profile	INFORMATIVE Level	INFORMATIVE Color primaries & matrix coefficients	INFORMATIVE Transfer Characteristics	INFORMATIVE 'codecs' MIME subparameters	NORMATIVE CMAF Brand	NORMATIVE Normative Reference
UHD10	HEVC	Main10 MainTier 10-bit	5.1	1 (BT.709) 9 (BT.2020)	1 (BT.709 OETF) 14 (BT.2020 OETF)	hev1.2.4.L153.B0 hvc1.2.4.L153.B0	'cud1'	[CMAF] Table B.1
HLG10	HEVC	Main10 MainTier 10-bit	5.1	9 (BT-2020)	18 (BT.2100 Table 5 HLG OETF) 14 (BT.2020 OETF)	hev1.2.4.L153.B0 hvc1.2.4.L153.B0	'clg1'	[CMAF] Table B.1
HDR10	HEVC	Main10 MainTier 10-bit	5.1	9 (BT.2020)	16 (BT.2100 Table 4 PQ EOTF)	hev1.2.4.L153.B0 hvc1.2.4.L153.B0	'chd1'	[CMAF] Table B.1

4.2.2 Video Media Profile MIME types

MIME types and subparameters assist with the task of describing the format of content and the capabilities required to decode it. This section describes the MIME types used for WAVE video media profiles.

The MIME type and subtype of a video CMAF Track are defined by RFC 6381 and are "`video/mp4`".

An example 'codecs' MIME subparameter for each WAVE video Media Profile is listed in Table 1.

The 'codecs' string is expected to equal the video encoding parameters actually used to encode each CMAF Track and describe profiles, levels and constraints that are broad enough to cover the values in the sample entry in each applicable CMAF Track but not exceeding the maximum values permitted for the media profile. The 'codecs' string includes the codec, sample format (e.g., 'avc3'), profile, and level.

> **NOTE** *The majority of Players in use do not recognize the 'profiles' subparameter or CMAF Media Profiles. They will often determine if a Track might be decodable based on the 'codecs' subparameter. A Player that cannot decode every CMAF Track in a Switching Set might be able to decode some of the Tracks identified by the 'codecs' subparameter to be encoded at lower codec profile and levels.*

RFC 6381 specifies the values in the `'profiles'` subparameter to equal a quoted list of comma delimited compatibility brands starting with the major brand in the File Type Box (`'ftyp'`).

For example, an AVC HD CMAF Track encoded at the highest allowed profile and level could have a MIME type of:

```
video/mp4; codecs="avc3.640028"; profiles="cfhd,cmfc,iso9,ccea"
```

In this example, the first brand listed in the `profiles` subparameter indicates conformance to the `'cfhd'` CMAF Media Profile brand, and the other brands indicate conformance to `'cmfc'` CMAF Track constraints and boxes (including additional video and conditionally required encryption boxes), `'iso9'` ISO base media file format, and 'ccea' the presence of CEA closed captions in the video elementary stream.

The inclusion of the CMAF brands conveys that the content conforms to CMAF constraints, e.g., the ISOBMFF tracks are consistently fragmented and timestamped, and are constrained relative to other tracks that conform to a Switching Set and are encoded and packaged to enable *Seamless Adaptive Switching*.

4.3 WAVE Audio Media Profiles

4.3.1 General

Each Audio Media Profile SHALL conform to the normative reference listed in Table 2.

Table 2 summarizes the WAVE Audio Media Profiles, their brand, 'codecs' subparameter string(s), important parameter constraints, and the Media Profile's normative reference.

Where the Media Profile specifies more than one option for a particular characteristic (e.g., profiles, channel configuration, etc.), all listed options are valid in content.

Table 2 – WAVE Audio Media Profiles

Media Profile Name	INFORMATIVE Codec Family	INFORMATIVE Allowed Codecs or Profiles	INFORMATIVE Level	INFORMATIVE 'codecs' MIME subparameter	NORMATIVE CMAF Brand	NORMATIVE Normative Reference
AAC Core	AAC	AAC-LC, HE-AAC or HE-AAC v2	2	mp4a.40.2 mp4a.40.5 mp4a.40.29	'caac'	[CMAF] Table A.2
Adaptive AAC Core	AAC	AAC-LC, HE-AAC or HE-AAC v2	2	mp4a.40.2 mp4a.40.5 mp4a.40.29	'caaa'	[CMAF] Table A.2
AAC Multichannel	AAC	AAC-LC, HE-AAC	6	mp4a.40.2 mp4a.40.5 mp4a.40.29	'camc'	[CMAF] Table i.2
Enhanced AC-3, including AC-3	AC-3 EAC-3	AC-3 EAC-3	n.a.	ec-3	'ceac'	[EAC3]
AC-4, Single Stream	AC-4	AC-4	3	ac-4.02.01.03	'ca4s'	[AC4]
MPEG-H, Single Stream	MPEG-H	Low Complexity (LC)	3	mhm1.0x0B mhm1.0x0C mhm1.0x0D	'cmhs'	[CMAF] Table j.2
DTS-HD Audio	DTS-HD	DTS, DTS-HD	n.a.	dtsc dtsh dtse	'dts1'	[DTS1]
USAC Stereo	MPEG-D USAC	Baseline USAC	2	mp4a.40.42	'casu'	[CMAF] Table K.2

4.3.2 Audio Media Profile MIME types

The MIME type and subtype of a subtitle CMAF track are defined by RFC 6381 and are "`audio/mp4`".

An example '`codecs`' MIME subparameter for each audio WAVE Media Profile is listed in Table 2.

The `'codecs'` string is expected to equal the actual profile and level encoded rather than a superset it might also conform to.

RFC 6381 specifies the string in the `'profiles'` subparameter to equal a quoted comma delimited list of the compatibility brands in the File Type Box (`'ftyp'`), starting with the major brand.

Example audio MIME type strings:

`audio/mp4; codecs="mp4a.40.5"`

`audio/mp4; codecs="mp4a.40.5"; profiles="caac,cmfc,iso9"`

In the second example, the first brand listed in the `profiles` subparameter indicates conformance to the `'caac'` CMAF Media Profile brand, and the other brands indicate conformance to `'cmfc'` CMAF Track constraints and boxes (including additional audio and encryption boxes), and `'iso9'` ISO base media file format.

> ***NOTE*** *Players that do not recognize the 'profiles' subparameter or CMAF Media Profiles will often determine if a Track might be decodable based only on the 'codecs' subparameter.*

4.4 WAVE Subtitle Media Profiles

4.4.1 General

Each Subtitle Media Profile SHALL conform to the normative reference listed in Table 3.

Table 3 describes the WAVE Subtitle Media Profiles and brands specified in CMAF in the normative reference indicated, as well as their `'codecs'` subparameter strings.

Table 3 – WAVE Subtitle Profiles

Media Profile Name	*INFORMATIVE* Description	*INFORMATIVE* 'codecs' MIME subparameter	NORMATIVE CMAF Brand	NORMATIVE Normative Reference
TTML IMSC1 Text	IMSC1 Text Profile	stpp.ttml.im1t	'im1t'	Specified in [CMAF], 11.3.3
TTML IMSC1 Image	IMSC1 Image Profile	stpp.ttml.im1i	'im1i'	Specified in [CMAF], 11.3.4
TTML IMSC1.1 Text	IMSC1.1 Text Profile	stpp.ttml.im2t	'im2t'	Specified in [CMAF] L.1.2
TTML IMSC1.1 Image	IMSC1.1 Image Profile	sttp.ttml.im2i	'im2i'	Specified in [CMAF] L.1.3

4.4.2 Subtitle MIME Media Types

The MIME type and subtype of a subtitle CMAF track are defined by RFC 6381 and [CMAF] and are `"application/mp4"`.

The `'codecs'` subparameter is listed in Table 3.

RFC 6381 specifies the values in the `'profiles'` subparameter to equal a quoted comma delimited list of the compatibility brands in the File Type Box (`'ftyp'`), starting with the major brand.

For example, the MIME type of an IMSC1 text subtitle CMAF Track with subparameters could be:

`application/mp4; codecs=”stpp.ttml.im1t”; profiles=”im1t,cmfc,iso9"`

If a subtitle CMAF Track can be processed by a TTML renderer conforming to additional TTML profiles, they can be additionally listed in the `‘codecs’` subparameter. For example, a `‘codecs’` subparameter of "`stpp.ttml.im1t|etd1`" would signal that the Track can also be processed by a renderer conforming to [EBU3380]. For additional TTML profiles, see [TTMLREG].

For CMAF conformant subtitle Tracks, the `content_type` field of the MIME box within the `XMLSubtitleSampleEntry` box is required to contain a codecs parameter that contains the substring `‘im1t’, ‘im1i’, ‘im2t’ or ‘im2i’`, according to whether it conforms to text or image Media Profile. For example, for an IMSC1 text subtitle track, “`stpp.ttml.im1t|etd1`” is a valid codecs parameter.

The MIME type and subtype for a TTML document (not a CMAF Track) is "`application/ttml+xml`".

4.5 WAVE Media Profile Encryption

4.5.1 General

The encryption of WAVE Media Profiles SHALL be compliant with [CMAF] clause 8, using either the ‘cenc’ AES-CTR scheme or the ‘cbcs’ AES-CBC subsample pattern encryption scheme, as specified in [CENC], clause 10.1 and 10.4, respectively.

Note some [CMAF] encryption constraints which are recommendations in [CENC] are requirements in [CMAF]. See [CMAF], clause 8.2.3.

For additional information, see section 7.2.2, WAVE Baseline Splice Profile Constraints and Recommendations.

4.5.2 Application of the ‘cbcs’ mode of Common Encryption

The ‘cbcs’ mode of Common Encryption uses pattern encryption as defined in clause 9.6 of [CENC]. This pattern consists of the number of encrypted cipher blocks indicated by the field `default_crypt_byte_block` or `crypt_byte_block` (if present in a sample group description) followed by the number of unencrypted sample data blocks indicated by the field `default_skip_byte_block` or `skip_byte_block` (if present in a sample group description).

The Pattern Block length = crypt_byte_block + skip_byte_block and the encrypt:skip pattern is crypt_byte_block:skip_byte_block.

- Clause 10.4.1 of [CENC] recommends a Pattern Block length of 10.
- Clause 10.4.2 of [CENC] recommends a encrypt:skip pattern of 1:9 (i.e. 10% encryption).

To ensure interoperability with platforms which are limited to supporting only one encrypt:skip pattern, application of ‘cbcs’ mode encryption of WAVE media profiles SHALL follow the following rules:

- Packaging WAVE Video Media Profiles using the ‘cbcs’ mode of Common Encryption [CENC] SHALL use a Pattern Block length of 10 and an encrypt:skip pattern of 1:9.

5 CMAF Presentation Profiles

The CMAF Presentations Profiles specified in the 2018 edition of the [CMAF] standard are CMFHD, CMFHDc, and CMFHDs. The conditionally required Media Profiles for these Presentation Profiles are specified in [CMAF] and summarized in Table 4. Additional Media Profiles may be present in conforming CMAF Presentations in additional Switching Sets.

The intent of these Presentation Profiles is to provide playback of at least the required Media Profiles on most deployed devices, which typically include decoders and decryptors for those Media Profiles and at least one of the specified decryption schemes. Future editions of this specification might include additional Presentation Profiles by reference.

As specified in CMAF, for the CMFHD Presentation Profiles:

- If a video Track is included, then a conforming Presentation will at least include that video in a CMAF Switching Set conforming to the required AVC video Media Profile.
- If an audio Track is included, then a conforming Presentation will at least include that audio in a CMAF Switching Set conforming to the required AAC audio Media Profile.
- If a subtitle Track is included, then a conforming Presentation will at least include that subtitle in a Switching Set conforming to the TTML text Media Profile. For avoidance of doubt, CEA-608/708 closed captions embedded in video SEI messages are not a CMAF subtitle Track.

Each CMAF Presentation Profile contains either all unencrypted samples, or some samples encrypted with Common Encryption [CENC] using the 'cenc' scheme, or the 'cbcs' scheme, but not both schemes in one Presentation.

Table 4 – CMAF Presentation Profiles

Presentation Profile	Required Video Media Profile	Required Audio Media Profile	Required Subtitle Media Profile	Allowed Encryption Scheme
CMFHD	'cfhd' HD AVC	'caac' AAC Core	'im1t' TTML IMSC1 Text	None
CMFHDc	'cfhd' HD AVC	'caac' AAC Core	'im1t' TTML IMSC1 Text	'cenc'
CMFHDs	'cfhd' HD AVC	'caac' AAC Core	'im1t' TTML IMSC1 Text	'cbcs'

NOTE *Other CMAF Media Profiles that are subsets of these Media Profiles and conform to them also conform to the Presentation Profile. For instance, 'cfsd' conforms to 'cfhd' but is constrained to a lower resolution, and 'caaa' AAC Adaptive Media Profile conforms to the 'caac' AAC Core Media Profile but is additionally constrained at Fragment boundaries to enable adaptive audio switching. A decoder that conforms to the superset Media Profile can also decode and render the subset Media Profile.*

6 WAVE Programs

6.1 Introduction

A WAVE Program is defined to be a sequence of one or more CMAF Presentations, and, if more than one, then all audio and video SHALL be contained in Sequential Switching Sets. Sequential Switching Sets are defined to be CMAF Switching Sets in a WAVE Program intended for sequential presentation with *Continuous Rendering* and user experience.

The transition from one Sequential Switching Set and CMAF Presentation to the next is defined as a Splice, which occurs at a WAVE Program presentation time called the Splice Point. WAVE Programs that contain more than one CMAF Presentation MAY conform to the splice constraints of a WAVE Splice Constraint Profile (see section 7). A WAVE Program containing a single CMAF Presentation need not conform to a WAVE Splice Constraint Profile because it contains no Splices.

CMAF Presentations in a WAVE Program NEED NOT conform to any Splice Constraint Profile, other than the constraints specified for any CMAF Presentation Profiles and CMAF Media Profiles included. If a WAVE Program does not conform to a Splice Constraint Profile, then encoding constraints of Splices are undefined in WAVE.

> **NOTE** *The intent of a WAVE Program is to allow Continuous Rendering of a consistent user experience for the duration of the WAVE Program. Sequential Switching Sets identify content intended to be Continuously Rendered to provide a consistent user experience, and WAVE Splice Constraints enable rendering on compatible Players and Media Pipelines without interruption, i.e. "seamlessly". For content conforming to a WAVE Splice Constraint Profile, a compatible Player or Hypothetical Render Model can render the sequence of video frames before and after a Splice Point without lost or delayed video frames and allow the next audio track to be rendered with a smooth transition at the video Splice Point with audio accurately synchronized to the video.*

WAVE Splice Constraint Profiles specify the encoding constraints across a Splice, but methods, measurement, and tolerances for seamless playback are specified in the WAVE Device and Test specifications.

7 WAVE Splice Constraint Profiles

7.1 Introduction

WAVE Splice Constraint Profiles apply to Sequential Switching Sets in WAVE Programs containing two or more CMAF Presentations. Splice Constraint Profiles are independent of the CMAF Presentation Profiles and Media Profiles they are applied to.

This 2019 edition specifies the WAVE Baseline Splice Constraint Profile, but other Splice Constraint Profiles could be specified in the future. WAVE Baseline Splice Constraints require very restrictive encoding parameters and Fragment sequencing to enable smooth playback on the majority of adaptive streaming devices deployed in the market. Less restrictive splice constraint profiles will likely be specified in the future that will provide smooth sequencing on more advanced devices.

7.2 WAVE Baseline Splice Constraint Profile

The WAVE Baseline Splice Constraint Profile is designed for *Continuous Rendering* across Splice Points in WAVE Programs on the installed base of adaptive streaming *Players.*

These *Players* are represented by a Baseline *Hypothetical Render Model* that is the intersection of many deployed adaptive streaming *Player* capabilities. The Baseline *Hypothetical Render Model* was used to determine the Baseline Splice constraints for *Continuous Rendering* of a WAVE Program with Sequential Switching Sets in a sequence of CMAF Presentations – i.e., a WAVE Program containing multiple Presentations.

> **NOTE** *Identification that content conforms to the WAVE Baseline Splice Constraint Profile can inform a Player that there will be Sequential Switching Sets with continuous content and Splices that conform to Baseline Splice constraints for the duration of the WAVE Program, so the Player will be able to continue buffering and decoding CMAF Segments and Fragments without reinitializing its MSE buffers and Media Pipeline.*

7.2.1 Baseline Hypothetical Render Model

The WAVE Baseline Splice Constraint Profile specified in Section 7.2.2 targets deployed devices and *Player* applications that rely on a single Media Source Buffer and *Media Pipeline* to process Sequential Switching Sets.

1. It is assumed reinitialization of an MSE buffer and/or *Media Pipeline* would interrupt rendering.

> **NOTE** *Changing video parameters such as horizontal and vertical sample counts and sample cropping typical in adaptive switching does not require reinitialization of the Media Pipeline or an MSE buffer. Increasing the codec profile or level, changing codecs, etc., requires reinitializing the MSE buffer and decoder, then rebuffering segments, and filling the decoded picture buffer before video rendering can resume.*

2. It is assumed that a manifest identifies Sequential Switching Sets in sequenced CMAF Presentations and splices that conform to splice conditions specified by the WAVE Baseline Splice Constraint Profile, e.g. connected adaptation sets in a DASH manifest.
3. It is assumed that a *Player* determines from the manifest or other means that reinitialization is not required at Splice Points conforming to the WAVE Baseline Splice Constraint Profile.
4. It is assumed that adaptive streaming *Players* can set a new presentation time offset to synchronize a discontinuous decode timestamp on the first Fragment of the next CMAF Switching Set to the WAVE Program's presentation time at the Splice Point.
5. It is assumed that each Fragment immediately prior to a Splice ends at the Splice Point, and each Fragment that follows a Splice begins at the Splice Point (i.e., CMAF Fragments are consecutive and start aligned at the Splice Point).
6. It is assumed that adaptive streaming *Players* can render the intended CMAF Presentation Profile and its conditionally required CMAF Media Profiles.

7. It is assumed that adaptive streaming *Players* can decrypt either 'cenc' or 'cbcs' Presentations in a WAVE Program, but not necessarily both.

> ***Note*** *Fragment time alignment at Splice Points is a limitation caused by appending spliced segments to a single MSE buffer and is also an HLS media playlist discontinuity constraint.*
>
> *A CMAF Header can be appended to an MSE buffer as the result of a Splice or an adaptive switch, but the Media Pipeline is normally not aware why a Header was appended. Some implementations assume that the only parameter changes will be those allowed within a CMAF Switching Set (such as video resolution changes), unless the Player application reinitializes the MSE buffer. During CMAF adaptive switching, changes in Header parameters other than VUI video cropping parameters in the sample entry might thus be ignored by deployed adaptive streaming Players. If a WAVE Program conforms to the Baseline Splice Constraints, then those Splice constraints avoid the need to reinitialize the Media Pipeline on Splices, as well as switches.*

7.2.2 WAVE Baseline Splice Profile Constraints and Recommendations

Table 5 specifies required constraints on Splices between Sequential Switching Sets in WAVE Programs conforming to the WAVE Baseline Splice Constraint Profile.

Table 5 – WAVE Baseline Splice Constraint Profile

Parameter	Splice Constraint between Sequential Switching Sets
Media Profile Brand	Sequential Switching Sets SHALL conform to the same CMAF Media Profile.
BaseMediaDecodeTime	Can be discontinuous at Splice Points (i.e., require change in the presentation time offset).
Encryption	Can change between unencrypted/encrypted at Splice Points. SHALL only contain one [CENC] scheme per Program ('cenc' or 'cbcs').
Fragment Overlap	CMAF Fragments SHALL NOT overlap the same WAVE Program presentation time or Splice Point. CMAF Fragments SHALL NOT have gaps in WAVE Program presentation time at the Splice Point. Note: WAVE Program presentation time is the presentation start time of each Fragment and media sample relative to the start of the WAVE Program.
Sample Entry Type	Sample entries SHALL NOT change sample type at Splice Points (e.g., 'avc1' to 'avc3').
DRM	Default_KID can change at Splice Points
Track_ID	Track_ID can change at Splice Points.
Timescale	Timescale can change at Splice Points.

Parameter	Splice Constraint between Sequential Switching Sets
Video resolution	SHALL conform to the Media Profile.
Video frame rate	SHALL conform to the Media Profile*.
Video Picture Aspect Ratio	SHALL conform to the Media Profile*.
Audio Channel Configuration	SHALL conform to the Media Profile*.
Single Initialization per Switching Set	Switching Sets MAY conform to CMAF Single Initialization Constraints to indicate that no CMAF Header needs to be appended on a track switch.
CMAF Header constraints between Sequential Switching Sets	Encoding parameters SHALL be constrained or signaled such that CMAF Fragments of the following Switching Set can be decoded by a decoder configured for the previous Switching Set without reinitialization to a higher profile, level, resolution, etc.
Common CMAF Header for Sequential Switching Sets	Sequential Switching Sets and CMAF Tracks can encode and reference the same CMAF Header to explicitly indicate that reinitialization is not required at Switch or Splice Points.

*See additional recommended Splice constraints for continuity of user experience.

Table 6 specifies additional constraints recommended for continuous user experience in a variety of WAVE Programs and rendering environments.

Table 6 – WAVE Recommended Splice Constraints

Parameter	Recommended Splice Constraint between Sequential Switching Sets
Picture Aspect Ratio	SHOULD be the same for all CMAF Presentations
Frame Rate	SHOULD be the same family of multiples for all CMAF Presentations
Audio Channel Configuration	SHOULD allow the same stereo or multichannel configuration on internal and external decoders (including conversion, such as downmix and substream selection from multichannel Tracks for stereo rendering).
Video Dynamic Range and Color Space	SHOULD be consistent between Related Video Switching Sets for consistent user experience on render and display systems with limited tone mapping capabilities

See Annex A for a detailed discussion of these recommended constraints.

Annex A. System & Quality of Experience Splice Constraints (Informative)

A.1. Overview

While it may be possible for a device to decode and decrypt a sequence of CMAF Presentations conforming to WAVE Splice Constraint Profile, additional constraints are recommended for reasons of consistent user experience within a WAVE Program on a range of devices and device types.

A.2. Picture Aspect Ratio

A consistent picture aspect ratio throughout a Program allows consistent screen layout (including video framing by cropping and padding, placement of on screen controls, text, etc.), and it avoids artifacts that can result when attempting to change the screen layout at a Splice Point – e.g., adding or removing bars, zooming and cropping to fill the screen or a window, changing borders, controls, subtitles, menus, and other displayed content in response to the video shape change. Some reformatting is best accomplished using professional equipment and expertise in a studio before adaptive encoding (e.g., 9:16 portrait orientation cell phone video), rather than relying on dynamic screen adaptation algorithms in each device.

A.3. Frame Rate

A consistent video frame rate throughout a WAVE Program is recommended to avoid motion artifacts due to frame drops or frame repeats. This may not be a problem for devices with internal displays that refresh at a high frequency, but external displays connected by HDMI are expected to maintain a fixed frame rate negotiated on startup (based on EDIDs supported by both the display and device). If an HDMI frame rate of 50Hz is initialized and 60Hz content decoded, ten frames per second will be dropped resulting in uneven motion. Conversely, if HDMI is initialized at 60Hz, ten frames will be repeated for 50Hz content, also resulting in uneven motion. Displays often have much higher refresh rates but can only repeat the frames received. Some displays have the option to interpolate frames, but a service provider can't rely on that feature or its quality.

In the case of sequenced 24Hz and 30Hz CMAF Presentations, a method called "3:2 pulldown" is often carried over from analog TV systems. A 60 field per second EDID is negotiated, and 24 or 30 frames are sent as 60 fields, then converted back to 24 or 30 frames in the display ("deinterlaced"). If these are refreshed at an integer multiple (e.g. 120Hz refresh) or high enough refresh rate (e.g. 200Hz), motion representation will closely approximate the captured sample frequency and motion blur.

Switches between 25Hz and 50 Hz or between 30Hz and 60Hz should be well-tolerated, as the higher frame rate of each family can be negotiated with the display and either frame rate used without significant motion artefacts.

A.4. Audio Channels

A consistent audio channel layout may provide a better user experience, but it is content and system-dependent. For instance, the change between stereo in ads and 5.1 channel surround sound in movie or TV show may be acceptable. Systems with two speakers could mix both to stereo for consistency. A multi-speaker system could synthesize spatial audio from stereo and provide consistency between stereo and multichannel content. Decoding can be done in a device and output over HDMI as PCM, or a bitstream output that is decoded in an external AVR which is fully aware of connected speakers, their types, locations, etc. Within the constraints of an audio Media Profile, changes in channel configuration between sequenced CMAF Presentations should not cause decoding errors or time delays.

A.5. Video Color Characteristics

Consistent video rendering that utilizes a display's color gamut and dynamic range is desirable for all Presentations in a Program to create a consistent viewing experience.

Decoders in devices can "tone map" different encoded video characteristics to internal displays by applying different color lookup tables that map the decoded YCbCr subsamples to 4:4:4 RGB values tone mapped to the display's color gamut and dynamic range. External displays may also attempt to tone map all video to the display's color gamut and dynamic range.

The most consistent viewing experience on a range of displays can be accomplished by grading and encoding similar video characteristics in each Presentation, such as dynamic range, peak white, and color volume.

Metadata, such as mastering display characteristics, peak/average white level, and tone mapping metadata for different display ranges, may also be present in the bitstream and accessible to the decoder. Metadata can help tone map Presentations with different video characteristics to an internal display's known characteristics and viewing conditions. It can also be used to converge Presentations with different characteristics during decoding and color lookup for output over HDMI, assuming that decoded video will then be automatically and consistently tone mapped again in a display to match its display characteristics.

External displays that receive decoded video over HDMI or similar interfaces usually have no access to metadata in the encoded bitstream, so they apply tone mapping to the decoded video based on proprietary algorithms that analyze the video and tone map to the display's capabilities. The result favors consistency over accuracy, since TV displays are optimized for the mix of dynamic range, peak white, and color volume typical of most video sources today.

Even with newer video Media Profiles, displays, and interfaces that support tone mapping metadata, consistent video encoding and metadata for all Presentations in a Program improves viewing consistency on other systems.

A.6. Digital Rights Management Considerations

Initializing or reinitializing a DRM security context, downloading and processing a license, negotiating a secure connection over HDMI, etc. may not be seamless when attempted in realtime during a WAVE Program. However, DRMs often support "key rotation" to allow real-time encryption changes without reinitializing the DRM (including between clear and encrypted, as well as different keys). Individualized DRM licenses that are encrypted with DRM keys that are unique per device or domain can be authorized and downloaded once per WAVE Program, per channel, per subscription, per region, etc. (scope determined by business factors), then decryption keys can be delivered in CMAF Fragments and changed ("rotated") using Common Encryption to force the DRM client to check that the *Player* has a valid license on each key change, because the license is necessary to decrypt the new media key in the CMAF Fragment, which is necessary to decrypt the content.

It is possible to signal DRM license downloads far enough in advance to authorize and download them in time when a device has a reliable two-way connection and request storms are not a problem (e.g., 100 million people requesting a world cup license when returning from an ad in a live broadcast at exactly the same time).

However, setting up the new security context, negotiating secure output connections, etc. in a device may not happen fast enough to maintain seamless decoding and rendering. The time required depends on the level of security, DRM system, type of device, etc. For instance, a DRM client running as software in a browser or CPU memory may be quickly reconfigured but isn't very secure. A closed hardware system with a server controlled key download and public certs may be fast but inflexible (no DRM license with rights expression language to enforce different permissions and constraints, key rotation, robustness, breach detection, revocation, etc.). A programable DRM system with full features and high security derived from a hardware root of trust may take more time to verify the trust of all software and hardware in the secure context when one security context is destructively erased, and a new security context constructed.

Service providers should consider the type of devices and DRMs they wish to target, and their usage model to determine what additional constraints may be necessary to maintain seamless presentation of protected content.

Annex B. WAVE Program Decoding Considerations (Informative)

B.1. Overview

It is assumed that most deployed devices can't seamlessly reinitialize decoders and the DRM security context at Splices between CMAF Presentations in a WAVE Program. Significant changes in decoder configuration require re-initializing the Media Source Buffer (in HTML5 MSE), the decoder's coded picture buffer and decoded picture buffer, which deletes samples currently being decoded, then takes time to re-buffer the media source buffer and re-initialize decoding, decode and reorder frames, etc. before outputting the next picture.

CMAF specifies encoding constraints that allow seamless splicing of CMAF Fragments within a CMAF Switching Set on most devices (adaptive switching). At this time, CMAF does not specify constraints between sequenced CMAF Presentations that allow frame accurate playback of splices between CMAF Presentations. Therefore, this specification defines WAVE Baseline Splice Constraints to enable continuous (frame accurate) playback of splices between CMAF Presentations based on the following considerations.

B.2. Audio and Video Media Pipeline Initialization Considerations

During adaptive switching of a CMAF Switching Set containing NAL structured video (AVC, HEVC, etc.), the only significant decoding parameters that change after MSE buffer and Media Pipeline initialization are the encoded video width and height, and the cropping parameters signaled in Sequence Parameter Set Video Usability Information (SPS VUI). All Tracks in a Switching Set can be scaled to the same video aperture (selected by a Player to fit the video aspect ratio to a window or screen size and aspect ratio using scaling, cropping, padding, etc.). The only parameters that need to be read during a CMAF Track switch are the cropping parameters necessary for a decoder to discard portions of partially filled macroblocks. That information can come from a sample entry ('avc1' and 'hvc1' track format) or the IDR sample being decoded in the Fragment ('avc3' and 'hev1' track format). Changing the cropping parameters does not require a decoder reset, nor does decoding a lower profile, level, or resolution of subsampled Tracks in the Switching Set. Changing sample durations due to frame rate subsampling also does not require resetting the MSE buffer, decoder, or other parts of the Media Pipeline.

At a Splice Point between Switching Sets of the same video Media Profile, it is possible that a sample entry in the previous Presentation initialized a decoder profile and level that is lower than that required by the next Switching Set, unless the first Header or manifest contains initialization values sufficient for the entire WAVE Program, or convention, such as single "common Header" and URL, that guarantees the first CMAF Header used for initialization will be adequate for the entire Program.

Baseline Splice Constraints require the content following a splice to be decodable by the initialized decoder configuration; therefore, deployed devices capable of seamless adaptive switching are usually able to continuously decode a sequence of Fragments conforming to Baseline splice constraints.

B.3. CMAF Fragment Alignment Considerations

CMAF requires that each CMAF Presentation start on the first presented video sample of the first CMAF Fragment if the Presentation contains video. CMAF allows the first audio Fragment in a Presentation to overlap the start of the earliest video sample since audio frames are randomly accessible and quickly decoded, so audio rendering can be start-aligned with the video.

However, if the last and next audio CMAF Fragments overlap at a Splice Point and are appended to the same MSE buffer, the next Fragment will overwrite the overlapped decode time range of the last Segment of the previous Presentation, thus starting playback of the next Presentation prior to the Splice Point. If video Segments were allowed to overlap the previous Presentation, there would be additional problems resulting from frame reordering after decoding, thus interrupting presentation order (AVC and HEVC reorder, other codecs like AV-1 have the same decode and presentation order).

Therefore, Baseline Splice Constraints require Fragment time alignment similar to CMAF Switching Sets – i.e., no overlap or gaps at a splice (but, with a gap tolerance of a sample duration).

B.4. Digital Rights Management Considerations

In the case of secure decryption and DRM, changing the current security context may require "secure stop", which deletes any decrypted content and randomizes the decryptor memory that has handled decoded video, keys, etc. To initialize a new security context, a new DRM license must be authenticated, and its permissions, user or domain binding, license expiration time checked, maximum output resolution selected based on the device security level, etc., and playback restrictions verified (e.g., a particular level of HDCP specified by DRM output controls verified by handshake over HDMI). To initialize or re-initialize hardware root of trust, signatures of all software running in the security context may need to be verified and a trusted video path identified (e.g., encrypted system busses and GPU) all the way through protected video output to an external display. Changing a secure DRM context can take around one second and delete samples intended for presentation unless a sufficient number of decoded and decrypted samples are securely buffered to maintain output during reinitialization, and the decryptor and decoder can run faster than realtime to "catch up" to presentation time after reinitialization. A native *Player* application may have control over the *Media Pipeline*, but a Web application depends on the *Media Pipeline* implemented by each *Player's* combination of Web browser, operating system, and hardware. Therefore, decoder and DRM reinitialization with a single *Media Pipeline* is typically not seamless.

For continuous playback of a WAVE Program that will contain encrypted Presentations, the Media Pipeline should be initialized with the required encryption scheme at the start.

B.5. Device-dependent Considerations

WAVE Programs that do not include a continuous sequence of Baseline constrained splices will have device-dependent behavior at Presentation splice points. Splices might result in a start delay of the next CMAF Presentation, black or repeated video frames, skipped audio and video … or seamless presentation; depending on device capabilities and the differences between the sequenced CMAF Presentations. The closer to Baseline constraints, the more devices will be able to render a presentation seamlessly, e.g. by avoiding Fragment presentation time overlap, DRM changes, change of codec, etc.

Additional constraints on allowed content changes between CMAF Presentations are recommended for consistent user experience, even though they could be decoded and rendered. For instance, the same picture aspect ratio, frame rate family, video characteristics, and audio channel configuration throughout a WAVE Program are recommended for a consistent user experience across a variety of devices and rendering systems.

The sequence of CMAF Presentations can have practical consequences, e.g., if the first Presentation initializes a particular codec, profile, level, resolution, protection level, etc.; a subsequent Presentation might require reinitialization. When Sequential Switching Sets are encoded with the same ("Common") Header for each Switching Set, identifying a single header indicates that Splices between these Switching

Sets in any order need not be reinitialized. A single Header per Switching Set or sequence of Switching Sets in a WAVE Program is made practical by storing the video parameter sets, which differ per Track, in each CMAF Fragment – i.e., using 'avc3' and 'hev1' track formats.

Alternative player architectures to MSE/EME can accomplish frame accurate presentation by rapidly reconfiguring decoding, decryption, and display processing between frames in response to changes in those parameters between Sequential Switching Sets. Assuming a single decoder, some degree of faster than real-time processing is required when a single *Media Pipeline* is reinitialized. Enough of the first coded video sequence following a Splice Point needs to be decoded before samples can be reordered and removed from the decoded picture buffer. Video sample output is needed before all pictures prior to the Splice Point have been rendered, and reinitialization cannot start until the last presented sample of the previous CMAF Presentation is removed from a frame buffer following the decryptor and decoder.

Overlap of CMAF Fragments in presentation time also requires faster than real-time decoding in a single *Media Pipeline*. Fragments and coded video sequences can only be bitstream spliced, decrypted, and decoded from the start of a Fragment. The end of a Fragment can be discarded, but not the start. A *Media Pipeline* must decode faster than real-time and ahead of real-time (buffering decoded samples) to splice decoded samples from overlapping Fragments at a Splice Point, rather than concatenate Fragments that form a valid bitstream that can continue decoding in real-time. This means that a decoder capable of decoding – e.g., level 5.1 content encoded at the maximum video drawing rate must be capable of decoding at a higher rate than level 5.1 to decode part of one Fragment, flush the decryptor and decoder, and decode a portion of the next Fragment faster than real-time to splice the decoded samples at the correct frame. Decoded samples need to be buffered to maintain output when truncating the last Fragment before a splice or reinitializing the MSE buffer and its decryptor and decoder at a Splice Point. If CMAF Presentations or WAVE Programs didn't require start alignment of the first video Fragment in a CMAF Presentation, continuous output would require decoding and buffering an entire Fragment of video samples, the last sample of which might be the first sample to be presented at the splice point.

Applications that instantiate a second media element and pipeline can display decoded samples at any time by switching display focus between media elements, rather than splicing compressed, possibly encrypted, Fragment bitstreams from different Presentations to form a single bitstream that is fed into a single decryptor/decoder. In the multiple media element case, splice conditioning is only required per media element, e.g., a sequence of ads processed by an ad insertion app that manages its own decoder can be bitstream spliced independently from a second bitstream and decoder that processes a WAVE Program such as a live TV stream.

B.6. Ad Insertion Considerations

Historically, most web ad insertion has relied on a separate decoder in a plugin, such as Flash, or a native app on a mobile or PC platform. Alternatively, ads are encoded in a single bitstream like broadcast systems. The addition of MSE to HTML5 browsers and support for CMAF on mobile platforms has created the opportunity to use the same WAVE Program content on most internet video devices with Baseline Splice Constraints that play seamlessly on nearly all embedded or browser hosted adaptive streaming Players, without a platform-specific ad insertion app.

Independent bitstreams and media elements for ad insertion allow ads to be played by an ad insertion app, while simultaneously streaming a live WAVE Program in a *Player* implemented as 1) a Web application on a browser or 2) a native application on a mobile platform. An embedded HLS or DASH *Player* can function independently from an ad insertion app that uses a different encoding or media format and delivery protocol (e.g. the ads could be downloaded and encoded as multiplexed M2TS files). In this case, ad delivery timing, delivery method, media format, and playback method are independent of the streaming WAVE Program (e.g., ad file playback in a simple <video> tag is possible).

B.7. Conclusion

Although it is theoretically possible for advanced devices to seamlessly splice unconstrained CMAF Presentations using multiple *Media Pipeline*s, or use faster than real-time decoding and large decoded sample buffers to splice decoded samples; Baseline splice conditions defined in this specification enable important classes of devices to seamlessly render a sequence of CMAF Presentations in a single *Media Pipeline* with a single media source buffer for each CMAF Selection Set rendered.

Annex C. Encryption Mode Considerations (Informative)

Requiring a multiplicity of media asset encodings adds unwelcome cost for asset creation, storage and distribution, as well as content delivery network (CDN) inefficiencies. WAVE attempts to address this problem by defining device-service interoperability points based on a handful of Media Profiles, each consistent with the new CMAF standard (ISO/IEC 23000-19).

For DRM-interoperable encryption, CMAF relies on the third edition of the CENC standard (ISO/IEC 23001-7:2016). CENC defines an interoperable mechanism for conveying DRM metadata, and it allows for not one but four distinct modes of encryption, represented by four-character codes – ‘cbc1’, ‘cenc’, ‘cbcs’ and ‘cens’. These are, respectively, AES-128 CBC or CTR mode encryption, with or without the power saving pattern encryption, where only a fraction of the data blocks within each video subsample are encrypted.

For the most part, only ‘cenc’ and ‘cbcs’ modes are in use for Web delivered media. If all DRM-capable devices supported both modes, then service providers could choose either mode and be ensured that a single encrypted asset would play on all those devices. Unfortunately, this is not the case. Until quite recently, the most commonly used DRM technologies either supported ‘cenc’ mode or ‘cbcs’ mode, but not both. And most devices only have one DRM technology tightly bound to the hardware.

To address this interoperability problem, some of the most commonly used DRM technologies have agreed to support both ‘cenc’ and ‘cbcs’ mode decryption. One very important exception to this is Apple Fairplay, which at the time of this writing only provides a public interface for ‘cbcs’ decryption.

This development would appear to provide the basis for a ‘cbcs’ recommendation to address encryption interoperability, since shortly, all the most commonly used DRM technologies will support ‘cbcs’ mode decryption. There are, however, reasons why simply encrypting content using the ‘cbcs’ mode of Common Encryption will not solve the encryption mode interoperability problem.

- **Encryption mode preference**: For a variety of reasons, more than a handful of service providers have expressed a strong preference for the original ‘cenc’ mode of Common Encryption and would prefer that all DRMs include support for that mode.
- **Legacy devices:** There are many devices in use today which only support one mode of Common Encryption and where there is no practical means for them to be updated to support the other mode. Furthermore, some categories of these difficult-to-update devices, such as television sets, continue to be in use for many years.
- **Rollout timeline:** The timeline for rolling out new encryption mode support to new devices can be very significant, especially when those devices are not developed by the DRM technology provider. This can take up to 2-3 years, particularly when that decryption support is provided by an SoC provider who then provides the SoC to a large commercial OEM with a lengthy rollout timeline of its own.

All of this means that to reach WAVE devices of importance to a service, that service may need to continue to support both ‘cenc’ and ‘cbcs’ mode encryption for years to come. This situation is not particularly new, since supporting old and new codecs creates the same dilemma for the service provider. However, it is important to note that the ‘cenc’ versus ‘cbcs’ issue is not a hardware driven problem, but a software engineering decision for the DRM provider exacerbated by the difficulty of providing firmware updates to legacy devices in the field.

Only time can address the legacy firmware update problem. So, the most that WAVE can do is to provide a recommendation for the future.

There are two available options for the future resolution of encryption interoperability – either 1) all devices and services support a single, specific encryption mode, or 2) all devices support both encryption modes, permitting services to choose as they will.

Given that there is strong service provider support behind both 'cenc' and 'cbcs' encryption, WAVE recommends that future devices support both 'cbcs' and 'cenc' decryption.

Annex D. CMAF Encoding for DASH-HLS Interoperability

D.1. DASH-HLS Interop Enabled by CMAF (Informative)

There are two popular adaptive streaming protocols in use today – HTTP Live Streaming [HLS] and Dynamic Adaptive Streaming over HTTP [DASH]. Both DASH and HLS are compatible with the ISO MPEG Common Media Application Format [CMAF], the encoding format adopted by the CTA WAVE project, suggesting CMAF can serve as the basis for DASH-HLS interoperability. However, the data models for DASH and HLS impose different constraints, so additional rules are needed to ensure CMAF can provide the basis for DASH-HLS interoperability.

D.2. DASH & HLS Terminology (Informative)

In addition to having different data models, DASH and HLS use different terminology. In some cases, there is a direct mapping from one term to another, but in most cases the mapping is not quite exact. In this specification, terms defined in both [DASH] and [HLS] are used. For convenience, the definitions are listed below.

Term	Src	Definition
Adaptation Set	DASH	A set of interchangeable encoded versions of one or several media content components. See also Master Playlist.
CMAF Chunk	CMAF	(from CMAF specification) CMAF media object that contains a consecutive subset of the media samples of a CMAF fragment, where only the first CMAF chunk of a CMAF fragment is constrained to be an adaptive switching point See Partial Segment (HLS). See Delivery Unit Segment (DASH).
Delivery Unit Media Segment	DASH	A Media Segment containing one or more whole self-contained movie fragments. A whole, self-contained movie fragment is a movie fragment (`'moof'`) box and a media data (`'mdat'`) box that contains all the media samples that do not use external data references referenced by the track runs in the movie fragment box.
EXT-X-DISCONTINUITY	HLS	A tag indicates a discontinuity between the Media Segment that follows it and the one that preceded it. See also Period.
Initialization Segment	DASH	Segment containing metadata that is necessary to present the media streams encapsulated in Media Segments. See also Media Initialization Segment.

Term	Src	Definition
Master Playlist	HLS	A Master Playlist provides a set of Variant Streams, each of which describes a different version of the same content. Master Playlist tags define the Variant Streams, Renditions and other global parameters of the presentation. See also Adaptation Set.
Media Content Component	DASH	One continuous component of the media content with an assigned media component type. See also Media Playlist.
Media Initialization Segment	HLS	Segment containing metadata that is necessary to present the media streams encapsulated in Media Segments. See also Initialization Segment.
Media Playlist	HLS	A Media Playlist contains a list of Media Segments, which, when played sequentially, will play the multimedia presentation. See also Media Content Component.
Media Presentation	DASH	A collection of data that establishes a bounded or unbounded presentation of media content. See also Presentation.
Media Segment	DASH	Segment that complies with media format in use and enables playback when combined with zero or more preceding segments, and an Initialization Segment (if any) See also Media Segment (HLS).
Media Segment	HLS	A Media Playlist contains a series of Media Segments that make up the overall presentation. See also Media Segment (DASH).
Partial Segment	HLS	A subset of a Media Segment that can be delivered independently to reduce publishing latency. See also CMAF Chunk and DASH Delivery Unit Media Segment.
Period	DASH	Interval of the Media Presentation, where a contiguous sequence of all Periods constitutes the Media Presentation. See also EXT-X-DISCONTINUITY.
Playlist	HLS	Either a Media Playlist or a Master Playlist. A Playlist is a media Playlist if all URI lines in the Playlist identify Media Segments. A Playlist is a Master Playlist if all URI lines in the Playlist identify Media Playlists.

Term	Src	Definition
Presentation	HLS	A Media Playlist contains a series of Media Segments that make up the overall Presentation. See also, Media Presentation.
Rendition	HLS	Renditions are alternate versions of the content, such as audio produced in different languages or video recorded from different camera angles.
Representation	DASH	Collection and encapsulation of one or more media streams in a delivery format and associated with descriptive metadata. See also, Variant Stream.
Segment	DASH	Unit of data associated with an HTTP-URL and optionally a byte range that are specified by an MPD, or with a data URL. See also, Media Segment.
Variant Streams	HLS	A Variant Stream includes a Media Playlist that specifies media encoded at a particular bitrate, in a particular format, and at a particular resolution for media containing video. It can also specify a set of Renditions. See also, Representation.

D.3. DASH & HLS Data Model Differences (Informative)

In HLS, the Master Playlist defines the number of Variant Streams and the number of Renditions for each of those Variant Streams. For each Rendition in the Master Playlist, a separate Media Playlist identifies the corresponding segments' URI for that Rendition. Since the HLS Master Playlist cannot be changed during the entire media presentation, in DASH terms, the data model is fixed for the entire Presentation at the start of media playback. While an HLS Media Playlist can mix segments of different codec/formats or from different timelines, the number of Variants Streams and Renditions cannot change within the playback of a single HLS Master Playlist.

In DASH, the Media Presentation Description (MPD) manifest defines a sequence of Periods, each of which has its own components. The type, number, codec/format, language and the number of Representations of the media components may change from one period to another during the media presentation.

A common approach for authoring heterogeneous media is to use Discontinuities in HLS and Periods in DASH. For HLS-DASH interoperability, when using the DASH Period approach, each Period must have the same set of Media Content Components so that they can be mapped to the fixed set of Variant Streams and Renditions in the HLS Master Playlist.

The DASH Industry Forum (DASH-IF) Simple and Main Live Interoperability Points [DASHIF-IOP] constrain the MPEG-DASH standard to improve interoperability. Compared to HLS, these DASH IOPs provide flexible data variation scenarios using multiple periods and, in the case of Main Live profile, event-based MPD updates.

Placing some constraints on multi-period functionality in DASH can ensure interoperability with HLS.

D.4. DASH Interop Constraints

D.4.1. Overview

To enable HLS and DASH interoperability, we define an HTTP Live Streaming Interoperability Point (HLS IOP). The HLS IOP is intended to be used for simple live DASH streaming with server-side ad insertion, using CMAF content for both HLS and DASH protocols.

HLS IOP is a more constrained subset of DASH-IF's Simple Live IOP, meaning the content generated according to HLS IOP would be also compliant to Simple Live IOP. By deploying HLS IOP, the existing HLS ecosystems can support DASH with minimal effort.

> ***NOTE*** *At the time of this publication there is a proposed amendment to [DASHDASH] which includes a DASH profile for CMAF content. If published, adherence to this profile to ensure DASH-HLS interoperability is expected to be included in a future version of this specification.*

D.4.2. Period Constraints

Updating periods with MPD update is made simpler by matching periods' @id.

For multiperiod DASH manifests, the Period@id SHALL be present and, as required by [DASH], unique within the scope of the Media Presentation.

D.4.3. Adaptation Set Constraints

The number of Adaptation Sets and their order within the Period do not change between successive Periods.

The corresponding Adaptation Sets in two Periods have the same content type, peak bandwidth and language values.

D.4.4. Representation Constraints

The number of Representations and their order within the Adaptation Set SHALL NOT change between successive Periods.

An HLS Master Playlist will contain Variant Streams with a specific collection of bitrates for the entire Presentation. To be interoperable with HLS, the corresponding DASH Adaptation Set MUST also contain Representations with a specific collection of bitrates for the entire Media Presentation. Should a new Period have an Adaptation Set missing one or more of the Representation bitrates, a Representation with a lower bitrate MAY be used in place of the missing Representation.

Take as an example a Media Presentation consisting of main program periods followed by advertising periods. Suppose the main program has encoded content at bitrates of 2,350 kbps, 3,000 kbps, 4,300 kbps and 5,800 kbps, but the advertising period content has been encoded at 2,350 kbps, 4,300 kbps and 5,800 kbps, missing the 3,000 kbps Representation. Placing at the origin advertising content encoded at 2,350 kbps but indicating it as being encoded at 3,000 kbps in the DASH manifest would provide DASH-HLS interop.

D.4.5. Content Protection Element Constraints

If the MPD or any of its updates may include any Period containing one or more Adaptation Sets with one or more content protection elements, the corresponding Adaptation Set in the first Period of the MPD SHALL also contain the corresponding content protection elements.

The player needs to acquire DRM licenses prior to playback of the protected content. Signaling such information in the first period allows the player to have DRM license prior to playback of a new period with content protection if such a period appears later in an MPD update.

Consumer Technology Association Document Improvement Proposal

If in the review or use of this document a potential change is made evident for safety, health or technical reasons, please email your reason/rationale for the recommended change to standards@CTA.tech.

Consumer Technology Association
Technology & Standards Department
1919 S Eads Street, Arlington, VA 22202
FAX: (703) 907-7693 standards@CTA.tech

Consumer
Technology
Association™