



International
Standard

ISO/IEC 23000-19

**Information technology —
Multimedia application format
(MPEG-A) —**

**Part 19:
Common media application format
(CMAF) for segmented media**

*Technologies de l'information — Format pour application
multimédia (MPEG-A) —*

*Partie 19: Format CMAF (Common Media Application Format)
pour médias segmentés*

**Third edition
2024-02**

Reference number
ISO/IEC 23000-19:2024(en)

Full document details: ISO IEC 23000-19-2024

© ISO/IEC 2024



Documents

Document Sharing Hub



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2024

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

© ISO/IEC 2024 – All rights reserved

Contents

Page

Foreword	vii
Introduction	viii
1 Scope	1
2 Normative references	1
3 Terms and definitions	3
3.1 Media objects	3
3.2 Logical structure	3
3.3 Application model	5
4 Abbreviated terms	6
5 Document organization	9
6 CMAF hypothetical application model, media object model and profiles	9
6.1 Overview of the hypothetical application model and media object model	9
6.2 CMAF content processing model	10
6.3 Late binding CMAF track synchronization	11
6.4 Adaptive switching of CMAF tracks in CMAF switching sets	12
6.5 CMAF specified objects and profiles	13
6.5.1 Object derivation and interoperability code points	13
6.5.2 Encoded media objects	13
6.5.3 Logical media object sets	13
6.5.4 Addressable media objects	13
6.5.5 CMAF profiles, brand and identifiers	14
6.6 CMAF media object model	15
6.6.1 CMAF fragments	15
6.6.2 CMAF tracks	16
6.6.3 CMAF track files	16
6.6.4 CMAF segments	17
6.6.5 CMAF chunks	17
6.6.6 CMAF switching sets and adaptive switching	18
6.6.7 CMAF selection sets and late binding	21
6.6.8 CMAF presentation timing model	22
6.6.9 Manifest information	24
6.6.10 CMAF addressable media objects, resources, and resource identifiers	24
7 CMAF track format	25
7.1 Overview	25
7.2 CMAF brands	25
7.3 CMAF media objects	27
7.3.1 CMAF boxes	27
7.3.2 CMAF track media objects	29
7.3.3 CMAF addressable media objects	34
7.3.4 CMAF switching sets	35
7.3.5 CMAF selection sets	38
7.3.6 CMAF presentations	38
7.4 Additional boxes, not defined in the ISO Base Media File Format	39
7.4.1 Track Encryption Box (' tenc ')	39
7.4.2 Sample Encryption Box (' senc ')	39
7.4.3 Protection System Specific Header Box (' pssh ')	39
7.4.4 Media profile specific boxes	40
7.4.5 Event Message Box (' emsg ')	40
7.5 Constraints on ISO Base Media File Format boxes	40
7.5.1 Movie Header Box (' mvhd ')	40
7.5.2 Metadata Boxes	40
7.5.3 Kind Box (' kind ')	41

7.5.4	Track Header Box ('tkhd')	41
7.5.5	Media Header Box ('mdhd')	41
7.5.6	Video Media Header Box ('vmhd')	42
7.5.7	Sound Media Header Box ('smhd')	42
7.5.8	Subtitle Media Header Box ('sthd')	42
7.5.9	Data Reference Box ('dref')	42
7.5.10	Sample Description Box ('stsd')	42
7.5.11	Protection Scheme Information Box ('sinf')	42
7.5.12	Track contained media sample information boxes	42
7.5.13	Edit List Box ('elst')	43
7.5.14	Track Extends Box ('trex')	43
7.5.15	Movie Fragment Header Box ('mfhd')	43
7.5.16	Track Fragment Header Box ('tfhd')	44
7.5.17	Track Run Box ('trun')	44
7.5.18	Sample Group Description Box ('sgpd')	44
7.5.19	Media Data Box ('mdat')	45
7.5.20	Sub-sample Information Box ('subs')	45
7.6	The Structural CMAF Brand 'cmfc'	45
7.7	The structural CMAF Brand 'cmf2'	45
7.7.1	General	45
7.7.2	Edit List Box ('elst')	45
7.7.3	Track Run Box ('trun')	45
8	Common encryption of CMAF tracks	46
8.1	Multiple DRM system support	46
8.2	Track encryption	46
8.2.1	General requirements	46
8.2.2	CMAF track constraints	47
8.2.3	Encryption constraints	48
8.2.4	CMAF presentation encryption	49
9	Video CMAF tracks	49
9.1	Overview	49
9.2	General video CMAF track format	50
9.2.1	General video CMAF track structure and constraints	50
9.2.2	Video Media Header ('vmhd')	50
9.2.3	Track Header Box ('tkhd')	51
9.2.4	Sample Description Box ('stsd')	51
9.2.5	Video CMAF fragment presentation time	52
9.2.6	Video media sample dependencies	52
9.2.7	Video edit lists	52
9.2.8	General video CMAF fragment random access constraints	52
9.2.9	Additional random access pictures within CMAF video fragments	52
9.2.10	Image framing and encoding constraints	53
9.2.11	General video CMAF switching set constraints	53
9.3	NAL structured video CMAF tracks	54
9.3.1	Overview	54
9.3.2	CMAF track format constraints for NAL structured video	54
9.3.3	NAL structured video access units contained in media samples	56
9.3.4	NAL structured video coding sequences corresponding to CMAF fragments	56
9.3.5	Elementary stream constraints	56
9.3.6	General CMAF switching set constraints for NAL structured video	57
9.3.7	Single initialization CMAF switching set constraints for NAL structured video tracks and media profiles	57
9.4	AVC video CMAF tracks	58
9.4.1	Storage of AVC elementary streams	58
9.4.2	Constraints on AVC elementary streams	58
9.5	AVC video Internet Media Type parameters	60
9.5.1	AVC signalling of "codecs" parameters	60
9.6	Video media profiles	60

10	Audio CMAF tracks	61
10.1	Overview	61
10.2	General audio CMAF track format	61
10.2.1	Derivation	61
10.2.2	Track Header Box (' <i>tkhd</i> ')	61
10.2.3	Sound Media Header Box (' <i>smhd</i> ')	62
10.2.4	Sample Description Box (' <i>stsd</i> ')	62
10.2.5	AudioSampleEntry	62
10.2.6	Audio offset edit list	62
10.3	AAC audio CMAF tracks	62
10.3.1	Overview	62
10.3.2	"codecs" parameter signalling	62
10.3.3	Considerations for AAC audio encoding	63
10.3.4	AAC track constraints	64
10.3.5	AAC elementary stream constraints	65
10.4	AAC core audio CMAF media profile	66
10.5	AAC adaptive switching audio CMAF media profile	67
10.5.1	General constraints	67
10.5.2	CMAF fragment encoding constraints	67
10.5.3	General considerations and requirements	67
10.5.4	Constraints for AAC-LC	68
10.5.5	Constraints for HE-AAC	68
10.5.6	Constraints for HE-AACv2	70
10.6	Audio media profiles	70
11	Subtitles and captions	70
11.1	Overview	70
11.2	WebVTT	71
11.3	IMSC text and image tracks	71
11.3.1	General	71
11.3.2	Common constraints	71
11.3.3	IMSC1 text track constraints	72
11.3.4	IMSC1 image track constraints	72
11.4	CTA-608 and CTA-708	73
11.5	Metadata for subtitles	73
11.6	Sparsity in Subtitle Tracks	73
11.7	11.7	
	Subtitle media profiles	73
12	CMAF media profiles and CMAF presentation profiles	73
12.1	CMAF media profiles	73
12.1.1	General guidelines for specifying CMAF media profiles	73
12.1.2	Guidelines for audio CMAF media profiles	74
12.1.3	Guidelines for video CMAF media profiles	75
12.2	CMAF presentation profiles	76
12.2.1	General	76
12.2.2	CMAF profile conformance	76
13	Timed metadata tracks	77
Annex A (normative)	CMAF presentation profiles, media profiles and supplemental data	79
Annex B (normative)	HEVC video CMAF track format and CMAF media profiles	83
Annex C (informative)	Source formats	91
Annex D (informative)	Hypothetical player model	102
Annex E (informative)	Event messages	105
Annex F (informative)	Error handling for missing media	106
Annex G (informative)	Recommendations for AAC CMAF switching set encoding	107

Annex H (normative) **Scalable HEVC media profile and track format**..... **110**

Annex I (normative) **AAC multichannel CMAF media profiles and track format**..... **116**

Annex J (normative) **MPEG-H 3D audio track format and CMAF media profile**..... **119**

Annex K (normative) **MPEG-D USAC track format and CMAF media profile**..... **124**

Annex L (normative) **IMSC 1.1 media profiles**..... **126**

Annex M (normative) **CMAF track and media profiles for VVC**..... **128**

Annex N (normative) **CMAF track and media profiles for EVC**..... **138**

Bibliography **142**

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and <https://patents.iec.ch>. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This third edition cancels and replaces the second edition (ISO/IEC 23000-19:2020), which has been technically revised. It also incorporates the Amendment ISO/IEC 23000-19:2020/Amd.1:2021.

The main changes are as follows:

- addition of [subclauses 9.6](#), [10.6](#) and [11.7](#),
- addition of [Annexes M](#) and [N](#).

A list of all parts in the ISO/IEC 23000 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

Common media application format (CMAF) combines and constrains several MPEG specifications to define a multimedia format that is optimized for delivery of a single adaptive multimedia presentation to a variety of devices, using a variety of adaptive streaming, broadcast, download and storage methods.

Several MPEG specifications have been adopted for much of the video delivered over the internet and other IP networks (cellular, cable, broadcast, etc.). Various organizations have taken MPEG's core coding, file format and system standards and combined them into their own specifications for their specific application. While these specifications are similar, their differences result in unnecessary duplication of engineering effort and duplication of identical content in slightly different formats, which results in increased storage and delivery costs.

CMAF provides a common media specification that application specifications, such as MPEG dynamic adaptive streaming over HTTP (DASH), can reference and a common media format that allows a single encoded multimedia presentation to be used by many applications.

[Clause 6](#) provides a description of the objects and terminology specified, the CMAF object model, and the hypothetical application model, which defines how these objects can be combined to form adaptive multimedia presentations.

The specifications in [Clauses 7](#) through [Clause 12](#) are terse to facilitate development and testing and assume an understanding of [Clause 6](#). [Clause 7](#) specifies ISO Base Media File Format boxes and structures such as movie fragments and tracks that are used to construct all CMAF media objects. [Clause 8](#) through [Clause 11](#) contain details specific to encryption, audio, video, and subtitle tracks. [Clause 12](#) specifies the combination of CMAF tracks and media profiles into CMAF presentations. It also recommends how to specify additional CMAF media profiles and presentation profiles, which can be specified by other documents and organizations.

CMAF presentation profiles and CMAF media profiles are specified in annexes to allow the addition of new profiles without changing the core document. Additional informative annexes have been added to provide explanations and recommendations on specific topics.

Information technology — Multimedia application format (MPEG-A) —

Part 19: Common media application format (CMAF) for segmented media

1 Scope

This document specifies the CMAF multimedia format, which contains segmented media objects optimized for streaming delivery and decoding on end user devices in adaptive multimedia presentations.

CMAF specifies a track format derived from the ISO base media file format, then derives addressable media objects from CMAF tracks that can be used for storage and delivery.

CMAF specifies sets of tracks that share encoding and packaging constraints that enable the selection of multiple tracks to form a multimedia presentation and allow seamless switching of alternative encodings of the same content at different bit rates, frame rates, resolution, etc.

CMAF specifies a hypothetical application model that determines how tracks in a CMAF presentation are intended to be combined and synchronized to form a multimedia presentation. The model abstracts delivery to allow any delivery method. The hypothetical application model assumes a manifest and player, but CMAF does not specify a manifest, player, or delivery protocol, with the intent that any that support the hypothetical application model can be used.

CMAF specifies media profiles and brands that constrain media encoding and packaging of CMAF tracks to enable seamless adaptive switching of tracks and allow devices to identify compatible content by its brand.

CMAF specifies presentation profiles that conditionally require sets of CMAF tracks conforming to specified media profiles and allow content creators and devices to identify compatible multimedia presentations.

CMAF enables extensibility by specifying how new media profiles and presentation profiles can be specified and identified and includes guidelines for those specifications.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 14496-1, *Information technology — Coding of audio-visual objects — Part 1: Systems*

ISO/IEC 14496-3, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-10, *Information technology — Coding of audio-visual objects — Part 10: Advanced video coding*

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 14496-14, *Information technology — Coding of audio-visual objects — Part 14: MP4 file format*

ISO/IEC 14496-15, *Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format*

ISO/IEC 14496-30, *Information technology — Coding of audio-visual objects — Part 30: Timed text and other visual overlays in ISO base media file format*

ISO/IEC 23001-7, *Information technology — MPEG systems technologies — Part 7: Common encryption in ISO base media file format files*

ISO/IEC 23003-4:2020, *Information technology — MPEG audio technologies — Part 4: Dynamic range control*

ISO/IEC 23003-3, *Information technology — MPEG audio technologies — Part 3: Unified speech and audio coding*

ISO/IEC 23008-2, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

ISO/IEC 23008-3:2022, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio*

ISO/IEC 23009-1, *Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats*

ISO/IEC 23091-3, *Information technology — Coding-independent code points — Part 3: Audio*

ISO/IEC 23090-3, *Information technology — Coded representation of immersive media — Part 3: Versatile video coding*

ISO/IEC 23094-1, *Information technology — General video coding — Part 1: Essential video coding*

IETF RFC 5234, *Augmented BNF for Syntax Specifications: ABNF*¹⁾

IETF RFC 6381, *The ‘Codecs’ and ‘Profiles’ Parameters for “Bucket” Media Types*²⁾

ITU-R Recommendation BT.709, *Parameter values for the HDTV standards for production and international programme exchange*

ITU-R Recommendation BT.1886, *Reference electro-optical transfer function for flat panel displays used in HDTV studio production*

ITU-R Recommendation BT.2035, *A reference viewing environment for evaluation of HDTV program material or completed programmes*

ITU-T Recommendation X.667:2014, *Information technology — Open Systems Interconnection — Procedures for the operation of OSI Registration Authorities: Generation and registration of Universally Unique Identifiers (UUIDs) and their use as ASN.1 object identifier components*³⁾

ANSI/SCTE 214-1, *MPEG DASH for IP-Based Cable Services Part 1: MPD Constraints and Extensions*, <https://www.scte.org/standards/library/catalog/scte-214-1-mpeg-dash-for-ip-based-cable-services-part1-mpd-constraints-and-extensions/>

W3C IMSC1, *TTML Profiles for Internet Media Subtitles and Captions 1.0.1*, <https://www.w3.org/TR/ttml-imscl>

W3C IMSC1.1, *TTML Profiles for Internet Media Subtitles and Captions 1.1*, <https://www.w3.org/TR/ttml-imscl1.1>

W3C WebVTT, *The Web Video Text Tracks Format, Candidate Recommendation 4 April 2019* <https://www.w3.org/TR/webvtt1/>

W3C TTML Media Type Definition and Profile Registry, *W3C Working Group Note*, <https://www.w3.org/TR/ttml-profile-registry>

1) <https://tools.ietf.org/html/rfc5234>

2) <https://tools.ietf.org/html/rfc6381>

3) <https://www.itu.int/rec/T-REC-X.667>

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <https://www.electropedia.org/>

3.1 Media objects

3.1.1

CMAF fragment

encoded ISO BMFF media segment conforming to CMAF constraints

3.1.2

CMAF header

sequence of CMAF constrained ISO BMFF boxes that do not reference any *media samples* (3.3.15), but are associated with a *CMAF track* (3.2.1) and necessary for the decoding of its *CMAF fragments* (3.1.1)

3.1.3

CMAF addressable media object

CMAF media object packaged for storage or delivery

Note 1 to entry: Examples include a *CMAF track file* (3.1.6) containing a *CMAF header* (3.1.2) and *CMAF fragments* (3.1.1), or a *CMAF segment* (3.1.5) containing one or more CMAF fragments, or a *CMAF chunk* (3.1.4) containing a partial sequence of the *media samples* (3.3.15) of a CMAF fragment.

3.1.4

CMAF chunk

CMAF media object that contains a consecutive subset of the *media samples* (3.3.15) of a *CMAF fragment* (3.1.1), where only the first CMAF chunk of a CMAF fragment is constrained to be an *adaptive switching* (3.3.9) point

3.1.5

CMAF segment

CMAF addressable media object (3.1.3) consisting of one or more consecutive *CMAF fragments* (3.1.1) from the same *CMAF track* (3.2.1)

Note 1 to entry: A “CMAF segment” is conformant to an “ISO BMFF segment” and a “DASH segment”.

3.1.6

CMAF track file

one *CMAF track* (3.2.1) stored consecutively in a single ISO BMFF file with the earliest *CMAF fragment* (3.1.1) constrained to start at decode time zero

3.1.7

CMAF principal header

CMAF header (3.1.2) that includes sufficient information to initialize the media decoder and decryption for all *CMAF tracks* (3.2.1) of the *CMAF switching set* (3.2.2)

3.2 Logical structure

3.2.1

CMAF track

sequence of *CMAF fragments* (3.1.1) that are consecutive in presentation time, contain one media stream, conform to at least one structural CMAF brand, together with an associated *CMAF header* (3.1.2) that can initialize playback

3.2.2**CMAF switching set**

set of one or more *CMAF tracks* ([3.2.1](#)), where each track is an alternative encoding of the same source content, and are constrained to enable seamless track *switching* ([3.3.9](#))

3.2.3**aligned CMAF switching set**

set of *CMAF switching sets* ([3.2.2](#)), the *CMAF tracks* ([3.2.1](#)) of which all contain alternative encodings of the same source content in time-aligned *CMAF fragments* ([3.1.1](#)), but all CMAF tracks do not conform to a single CMAF switching set

3.2.4**CMAF selection set**

set of one or more *CMAF switching sets* ([3.2.2](#)), where each CMAF switching set encodes an alternative aspect of the same presentation over the same time period, only one of which is intended to be played at a time, e.g. an alternative language or codec

3.2.5**CMAF presentation**

set of one or more *CMAF selection sets* ([3.2.4](#)) that can be simultaneously decoded to produce a multimedia user experience, potentially including synchronized audio, video, and subtitles

3.2.6**CMAF media profile**

encoding constraint on a *CMAF track* ([3.2.1](#)) and its contained *media samples* ([3.3.15](#)) associated with a CMAF structural brand

3.2.7**CMAF presentation profile**

requirement on the *CMAF media profiles* ([3.2.6](#)) contained in a *CMAF presentation* ([3.2.5](#))

3.2.8**required media profile**

CMAF media profile ([3.2.6](#)) conditionally required by a *CMAF presentation profile* ([3.2.7](#))

3.2.9**manifest**

document describing one or more *CMAF presentations* ([3.2.5](#))

Note 1 to entry: Manifest formats are not specified in this document.

3.2.10**audio programme**

complete collection of all audio programme components and, if present, a set of accompanying presets

3.2.11**audio programme component**

smallest addressable unit of an audio programme

3.2.12**CMAF supplemental data**

data that can be present in a *CMAF track* ([3.2.1](#)) and its contained *media samples* ([3.3.15](#)) conformant to a set of requirements identified by a brand

3.2.13**CMAF structural brand**

four-character code used in brand-signaling boxes to indicate compliance to box-level constraints as opposed to media-level constraints

3.3 Application model

3.3.1

CMAF hypothetical application model

CMAF presentation (3.2.5) application model based on *late binding* (3.3.3) and synchronization of *CMAF tracks* (3.2.1) that partly determines the CMAF track encoding constraints necessary for an intended CMAF presentation

3.3.2

player

component of the *CMAF hypothetical application model* (3.3.1) responsible for interpreting a *manifest* (3.2.9), requesting resources, and rendering a *CMAF presentation* (3.2.5)

3.3.3

late binding

selection (3.3.8) and synchronization of separately stored *CMAF tracks* (3.2.1) by a *player* (3.3.2) resulting in a synchronized multimedia presentation

3.3.4

CMAF presentation timeline

timeline shared by all *CMAF tracks* (3.2.1) in a *CMAF presentation* (3.2.5), starting at CMAF presentation time zero, which is coincident with the earliest *media samples* (3.3.15) intended for presentation

3.3.5

presentation time offset

earliest presentation time of each *CMAF track* (3.2.1) at the start of a *CMAF presentation* (3.2.5)

Note 1 to entry: Presentation time offset is an encoded property of tracks in a presentation, but it can also refer to that value stored in a *manifest* (3.2.9).

3.3.6

CMAF fragment duration

sum of the *media sample* (3.3.15) durations documented in the `TrackFragmentRunBox` of all `MovieFragmentHeaderBoxes` in the *CMAF fragment* (3.1.1)

3.3.7

CMAF presentation duration

sum of the *CMAF fragment durations* (3.3.6) of the longest *CMAF track* (3.2.1) in a *CMAF presentation* (3.2.5), starting from its earliest presentation time on the *CMAF presentation timeline* (3.3.4)

3.3.8

selection

choice of a *CMAF track* (3.2.1) from alternatives in a selection set (e.g. selecting an audio track by language), possibly by user action or stored user preference

3.3.9

switching

changing to a different *CMAF track* (3.2.1) during presentation, including adaptively switching between *CMAF fragments* (3.1.1) in a *CMAF switching set* (3.2.2)

3.3.10

seamless switching

switching (3.3.9) between *CMAF tracks* (3.2.1) without interrupting presentation of the media content, i.e., decoding *media samples* (3.3.15), at the same time and quality as though their containing CMAF track was decoded without switching

3.3.11

CMAF switching set constraints

CMAF media profile (3.2.6) constraints that enable seamless *switching* (3.3.9) between *CMAF tracks* (3.2.1) in a *CMAF switching set* (3.2.2) conforming to that media profile

3.3.12**single initialization CMAF switching set constraints**

additional *CMAF switching set constraints* (3.3.11) so *CMAF fragments* (3.1.1) do not depend on a different *CMAF header* (3.1.2) when *switching* (3.3.9)

3.3.13**resource identifier**

externally specified identifier that identifies a *CMAF addressable media object* (3.1.3)

Note 1 to entry: An example is a URI or other object identifier specified by a delivery protocol and *manifest* (3.2.9).

3.3.14**stream access point**

media sample (3.3.15) random access property

3.3.15**media sample**

media data in a *CMAF fragment* (3.1.1) associated with a single decode start time and duration

Note 1 to entry: The term “sample” is often used in the context of video to refer to the spatial samples of an image and in the context of audio to refer to PCM waveform samples. In this document, each type of sample is identified by a defined term. A media sample defined by ISO BMFF is always identified by the term “media sample”. The word “sample” is frequently used in ISO BMFF to refer to objects and parameters such as a “sample entry”, “sample size”, etc., and those terms are used without modification in this document.

3.3.16**audio PCM sample**

digital sample quantizing the amplitude of an audio waveform at regular and frequent intervals, e.g. 48 kHz

3.3.17**video spatial sample**

quantized values representing the colour and brightness of an area of an image corresponding to a two-dimensional spatial tessellation of the image

3.3.18**subsampling**

video encoding using a smaller number of *video spatial samples* (3.3.17) than the source video, that number being an integer submultiple that can be scaled to the source video size based on video stream parameters without position shift or picture aspect ratio distortion

4 Abbreviated terms

AAC	advanced audio coding
ABNF	augmented backus-naur form
ADIF	audio data interchange format
ADTS	audio data transport stream
AOT	audio object type
ASC	audio specific configuration
AU	access unit
AVC	advanced video coding
CCE	coupling channel element

CDN	content delivery network
CMAF	common media application format
CPE	channel pair element
CVS	coded video sequence [A sequence of media samples (coded video frames), starting with a SAP type 1 or 2, and including all media samples prior to the next SAP type 1 or 2 in decoding order.]
DASH	dynamic adaptive streaming over HTTP
DRC	dynamic range control
DRM	digital rights management
DSE	data stream element
DTV	digital television
EBU	European Broadcast Union
EOTF	electro-optical transfer function
EVC	essential video coding
GOP	group of pictures
HDMI	high-definition multimedia interface
HDR	high dynamic range
HDTV	high definition television
HEVC	high efficiency video coding
HLG	hybrid log-gamma
HRD	hypothetical reference decoder
IDR	instantaneous decoding refresh
IMDCT	inverse modified discrete cosine transform
IPF	immediate playout frames
ISO BMFF	ISO base media file format, defined in ISO/IEC 14496-12
KID	key identifier, defined in ISO/IEC 23001-7
LFE	low frequency enhancement
LKFS	loudness, K-weighted, relative to nominal full scale
MAE	MPEG-H audio metadata information
MHAS	MPEG-H audio stream
MIME	multipurpose internet mail extensions
MPD	media presentation description

MPEG	moving picture experts group
MSE	media source extension
NAL	network adaptation layer
NTSC	National Television System Committee
OETF	opto-electronic-transfer-funktion
PCM	pulse code modulation
PNG	portable network graphic
PPS	picture parameter set
QMF	quadrature mirror filter
RAP	random access point
RGB	red blue green
SAP	stream access point, defined in ISO/IEC 14496-12
SAR	sample aspect ratio
SBR	spectral band replication
SCE	single channel element
SEI	supplemental enhancement information
SHVC	scalable high efficiency video coding
SPS	sequence parameter set
TTML	timed text markup language
UHD	ultra high definition
URI	uniform resource identifier
URL	uniform resource locator
URN	uniform resource name
USAC	unified speech and audio coding
UTC	coordinated universal time
UUID	universally unique identifier
VOD	video-on-demand
VCL	video coding layer
VPS	video parameter set
VUI	video usability information
VVC	versatile video coding
XML	eXtensible Mark-up Language

5 Document organization

For document organization, refer to the Introduction.

6 CMAF hypothetical application model, media object model and profiles

6.1 Overview of the hypothetical application model and media object model

CMAF defines a hypothetical application model so that encoding to that model results in consistent CMAF track encoding, representation in manifests, track selection, late binding, synchronization, decoding, and rendering of CMAF presentations.

Decoding requirements can be inferred from encoding constraints and the hypothetical application model but are not directly specified by CMAF. CMAF does not specify manifest formats or associated resource identification and transport. However, CMAF does specify CMAF addressable media objects derived from encoded CMAF fragments, which can be referenced as resources by a manifest. External specifications can define how a manifest describes a CMAF presentation, including identifying CMAF addressable media objects as resources and representing their logical relationships determined by the CMAF tracks, CMAF switching sets, CMAF selection sets, and CMAF presentations they are derived from.

[Figure 1](#) illustrates the media objects that are specified by CMAF, starting with the encoded CMAF fragments that form CMAF tracks, then logical CMAF track sets determined by CMAF track encoding constraints, then derived CMAF addressable media objects that can package encoded CMAF fragments or their media samples for storage and delivery.

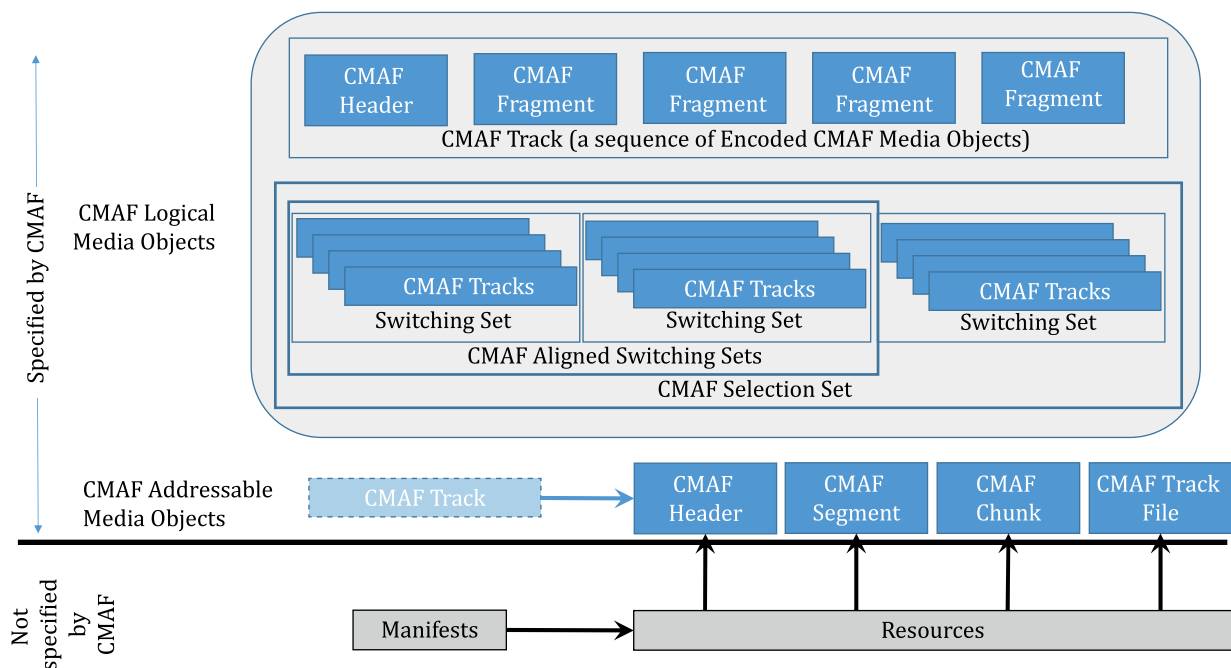


Figure 1 — Media objects specified in CMAF and presented by externally specified applications, such as adaptive streaming

[Figure 1](#) illustrates the mapping between CMAF specified CMAF presentations, and externally specified manifests and resources. Multiple manifests may reference the same CMAF presentation and CMAF addressable media objects. Specification of manifests and resource delivery is outside the scope of this document.

To accurately represent a CMAF presentation, a manifest will describe CMAF track relationships determined by each track's source content and CMAF track encoding constraints, e.g. that CMAF tracks belong to the same CMAF switching set, which belongs to a CMAF selection set. CMAF groups CMAF tracks based on their

encoding constraints in logical media objects called CMAF selection sets and CMAF switching sets that also determine intended use in late binding, track selection, seamless switching, and synchronization. Additional CMAF track metadata such as CMAF media profile brands, “codecs” parameters, language fields, etc. can be included in manifests to enable adaptive track selection and playback, optimized for each user and device.

Manifests can reference CMAF addressable media objects by resource identifiers used by manifests and servers to select the identified CMAF addressable media objects for delivery and playback. Multiple CMAF addressable media object types are specified for different delivery use cases. Use cases include prerecorded content that is downloaded or streamed as files, and live and on demand adaptive streaming over the Internet. The size of the CMAF addressable media objects can be optimized for efficient download and CDN caching, or fast bit rate switching and low latency, depending on the application.

[Figure 2](#) illustrates the relationship between CMAF and streaming specifications that can define a mapping between their manifest and resource formats, and CMAF presentations and the CMAF addressable media objects they include.

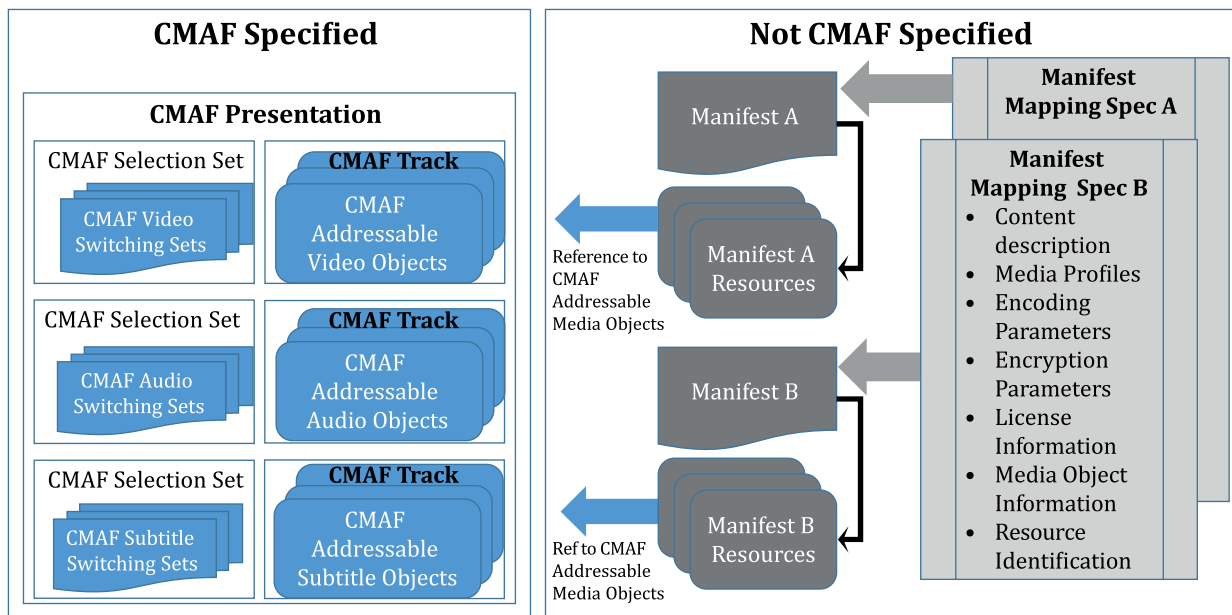


Figure 2 — CMAF hypothetical application model using externally defined manifests that describe the CMAF presentations and media objects

6.2 CMAF content processing model

The CMAF content processing model is shown in [Figure 3](#). Each CMAF presentation is composed of one or more media content components from different source material, for example, audio components in different languages, video components with different views of the same subject, subtitles with different languages or functions, etc. A typical presentation includes audio, video, and subtitles. Presentations are also possible that only include audio components, or include multiple audio components, or include multiple video sources, e.g. side-by-side video, picture in picture, sign language overlay, etc.

Synchronized CMAF tracks can be created and encoded at different times and synchronized on playback if they share a common presentation timeline. Content packaged with the appropriate CMAF fragment decode and presentation times can form a multimedia presentation conforming to the hypothetical application model. Each CMAF track can be encoded, encrypted and packaged by an independent encoder if each encoder constrains timing, encoding and encryption parameters to the constraints of the intended CMAF presentation, CMAF selection set, and CMAF switching set. CMAF tracks in a CMAF switching set can be independently encoded by multiple encoders with the necessary CMAF fragment time alignment by deriving CMAF fragment timing, spatial and temporal subsampling, etc., from the parameters of the shared media source they encode from.

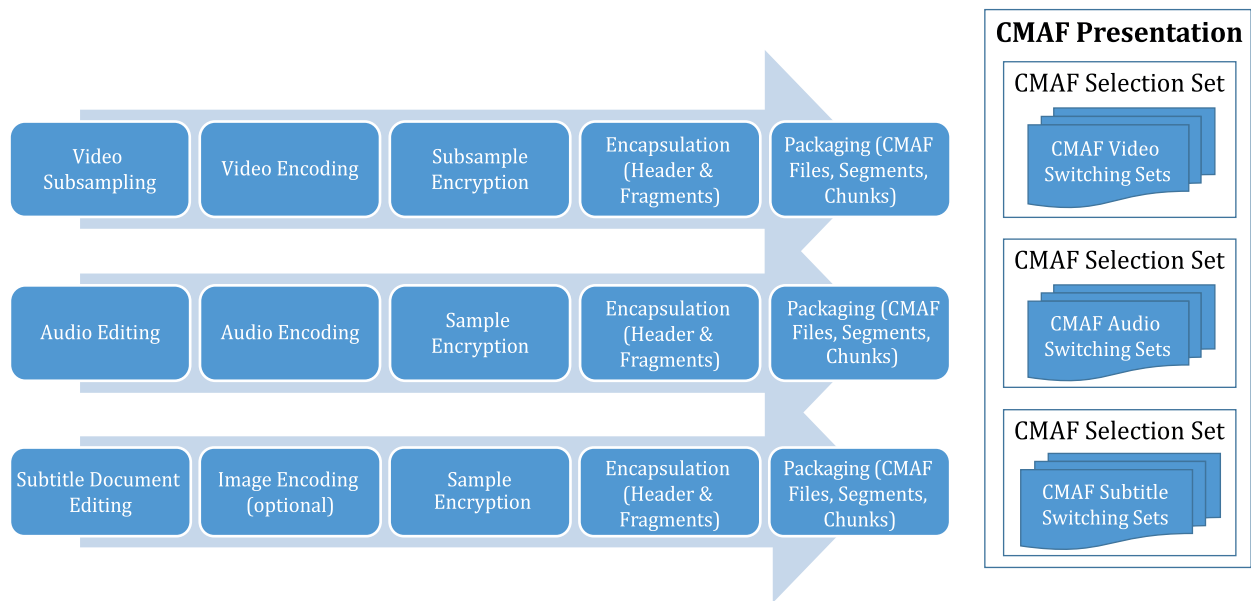


Figure 3 — CMAF presentation content generation

As illustrated in [Figure 3](#), video media components are often preprocessed (e.g. subsampled) and encoded in multiple media streams, typically different in bit rate and encoded resolution. Media streams can be encrypted and encapsulated into CMAF fragments, which can then be packaged as CMAF addressable media objects appropriate for one or more delivery methods. CMAF addressable media objects can be made available as resources and identified as defined by other media application specifications.

CMAF addressable media objects are derived from CMAF fragments and CMAF chunks as specified in [Clause 7](#), and inherit CMAF fragment constraints such as alignment of media sample sequences within a CMAF fragment, CMAF fragment time alignment within a CMAF switching set, CMAF track and CMAF media profile constraints, etc. This makes CMAF fragment encoding and decoding independent of the CMAF addressable media objects used to store and transport the CMAF fragments. But CMAF addressable media objects need not be constructed by packaging complete CMAF fragments as long as the resulting CMAF addressable media objects conform to CMAF fragment and CMAF track constraints.

For live streaming, a CMAF fragment in each CMAF track can be encoded simultaneously from each media content source component, and each CMAF fragment packaged and made available as soon as possible as a CMAF addressable media object, such as a CMAF segment or a CMAF chunk.

6.3 Late binding CMAF track synchronization

The feature of synchronizing separately stored CMAF tracks during playback is referred to as *late binding*. The late binding timing model and accurate recording of CMAF fragment presentation times during encoding are necessary to synchronize late bound CMAF tracks during delivery and presentation.

In the CMAF hypothetical application model, all CMAF tracks in a CMAF presentation share the same CMAF presentation timeline, which has a value of zero at the earliest media sample intended for presentation. All CMAF tracks in a CMAF presentation also share a common track decode timeline origin. A media sample of a CMAF track T1 that has a CMAF track presentation time of X is presented at the same time as the media sample of any other CMAF track T2 that also has the CMAF track presentation time X (the duration X having been divided by each track's timescale).

As defined in ISO BMFF, the presentation time of a media sample in a file is obtained from the decode time (derived from the `baseMediaDecodeTime` in the `TrackFragmentBaseMediaDecodeTimeBox` and the duration of previous samples) and possibly from composition time offsets and edit lists that offset each track's presentation time on the file's movie presentation timeline, which starts at zero. Equivalent decode and presentation times in different tracks can have different integer values if the tracks have different timescales.

CMAF tracks are not contained in a single multitrack ISO BMFF file with a shared movie timeline, but the CMAF presentation timeline provides an equivalent shared timeline and start point for all CMAF tracks in a CMAF presentation. If a CMAF presentation contains only CMAF track files, the presentation time of the first media sample of each CMAF track file is 0. However, in CMAF presentations using CMAF tracks, the presentation time of the earliest presented media sample can be a non-zero value, but is limited to an equivalent presentation time across all CMAF tracks. All CMAF tracks in a CMAF presentation are required to share a common timeline origin, but they may use different timescales, such as 90 kHz for video and 48 kHz for audio, so an equivalent decode or presentation time can have a different integer value stored in `baseMediaDecodeTime`.

A shared CMAF presentation timeline permits synchronization of different CMAF tracks at the CMAF player by, for example, starting presentation with the earliest media sample of the earliest video CMAF fragment and simultaneously presenting selected audio and subtitle CMAF tracks at their presentation time equivalent to the presentation time of the earliest video media sample. If an audio or subtitle CMAF fragment contains media samples with earlier presentation times than the earliest video media sample, the leading media samples are not expected to be presented. If a CMAF presentation contains no video, then the starting CMAF presentation time can be the earliest audio media sample.

A CMAF track's presentation time at the start of the CMAF presentation can be specified in a manifest presentation time offset so long as the equivalent offset is specified for all CMAF tracks in a CMAF presentation. For instance, each period in a DASH manifest is considered a CMAF presentation. The requirement that CMAF presentations share a common presentation timeline origin means that all presentation time offsets in a DASH Period will be equivalent values.

6.4 Adaptive switching of CMAF tracks in CMAF switching sets

CMAF tracks are contained in CMAF switching sets. Only CMAF tracks encoded from the same media content component can belong to the same switching set, but CMAF tracks encoded from the same media source may also be grouped into different switching sets that each have different encoding constraints necessary for seamless adaptive switching.

CMAF tracks in a CMAF switching set have the following characteristics.

- The CMAF tracks are alternative encodings of a single media content component and media type, e.g. the same audio or video source.
- The CMAF tracks are perceptually equivalent, e.g. the same aspect ratio, colour space, duration, etc., typically resulting from encoding the same media source, i.e., master file or input stream.
- The CMAF tracks conform to a common CMAF media profile.
- The CMAF tracks are seamlessly switchable based on the general constraints on CMAF tracks and encoding constraints defined by each CMAF media profile in the CMAF switching set.

Track switching refers to the presentation of decoded media samples of one CMAF track up to a presentation time t , and presentation of decoded media samples of another CMAF track from time t onwards. For instance, this could be accomplished using two decoders, downloading overlapping portions of any two streams, decoding both streams, and switching on any decoded media sample. Track switching could be user initiated or programmatic.

Track switching within a CMAF switching set conforming to a CMAF media profile allows a player to download a sequence of non-overlapping CMAF fragments from different CMAF tracks and feed them to a single decoder for seamless playback. When a player performs CMAF track switching automatically during playback in response to available bandwidth, video quality, decoding capacity, etc., that is called “adaptive switching”.

6.5 CMAF specified objects and profiles

6.5.1 Object derivation and interoperability code points

Common media application format specifies track formats and media formats for encoding multimedia and metadata derived from the fragmented ISO base media file format and various audio, video, and subtitle media formats.

- CMAF specifies encoded media objects in a CMAF track format that applies to all media types and determines the objects that are encoded and decoded.
- CMAF specifies logical sets of encoded media objects and their general encoding constraints.
- CMAF specifies profiles that determine the specific encoding constraints on media objects that conform to identified interoperability points called profiles.
- CMAF addressable media objects are derived by packaging CMAF headers, CMAF chunks, and CMAF fragments as ISO BMFF data structures available for storage and delivery.

CMAF addressable media objects inherit the properties of encoded media objects, their profiles, and logical media object and set constraints, and make the process of encoding and decoding of CMAF fragments independent of the delivery method and packaging. Different delivery specifications can map the same CMAF addressable media objects to their different manifests to enable interoperability. CMAF addressable media objects, such as CMAF chunks and CMAF segments, are logically derived from CMAF fragments, but it is up to implementations whether to physically derive them from CMAF fragments by post processing or encode and package them directly.

6.5.2 Encoded media objects

- CMAF specifies **CMAF headers** and **CMAF fragments** that form **CMAF tracks** that conform to **CMAF media profiles**. These are the encoded objects from which all other objects are derived.

6.5.3 Logical media object sets

- CMAF specifies **CMAF switching sets** that logically group CMAF tracks that are alternative encodings of the same content constrained to enable seamless switching. A CMAF switching set is constrained to simplify seamless switching between tracks by requiring time-aligned CMAF fragments, constrained encoding parameters, and shared encryption keys.
- CMAF specifies **aligned CMAF switching sets** that logically group multiple CMAF switching sets that are alternative encodings of the same content with time-aligned CMAF fragments. Some players can seamlessly switch between time-aligned CMAF fragments in different CMAF switching sets that are encoded with different characteristics, such as different codecs or encryption keys.
- CMAF specifies **CMAF selection sets** that logically group alternative CMAF switching sets based on their content and encoding constraints, e.g. alternative languages or codecs. At most, one CMAF track at a time is intended to be selected from each CMAF selection set in a CMAF presentation.
- CMAF specifies **CMAF presentations** that logically group CMAF tracks containing related synchronized content that can be selected, streamed, adaptively switched, decoded, decrypted, and synchronized on playback to render a multimedia presentation.

6.5.4 Addressable media objects

- CMAF specifies **CMAF addressable media objects** that are derived from CMAF tracks and fragments, and can be used for storage and delivery of a CMAF presentation. CMAF addressable media objects include **CMAF headers**, **CMAF track files**, **CMAF segments**, and **CMAF chunks**. CMAF fragment encoding and decoding are independent of the addressable media objects used to transport those CMAF fragments.

6.5.5 CMAF profiles, brand and identifiers

6.5.5.1 Overview

CMAF specifies the following brands and identifiers to support the creation and identification of interoperable content. They provide standardized conformance points for consistent content creation and validation. Players and devices can use these brands and identifiers to signal or recognize playback compatibility.

- CMAF specifies CMAF media profiles, CMAF supplemental data, and ISO BMFF brands for widely used audio, video, and subtitle formats. Each CMAF media profile specifies constraints on codecs, media samples, CMAF fragments, and CMAF tracks to identify encoder/decoder interoperability and optional functionality including random access and seamless adaptive switching specified as CMAF switching set constraints. CMAF also specifies guidelines for the specification of CMAF media profiles and matching brands to enable other specifications to define CMAF media profiles that conform to general CMAF requirements. See [Table A.1](#), [Table B.1](#), [Table A.2](#), and [Table A.3](#).
- CMAF specifies CMAF presentation profiles and identifiers that conditionally require inclusion of CMAF tracks conforming to specific CMAF media profiles. CMAF presentations conforming to the CMAF presentation profiles in [A.1](#) are intended to be compatible with most media playback devices.

6.5.5.2 CMAF presentation profile identifiers

CMAF presentation profile identifiers identify CMAF presentations containing CMAF tracks conforming to CMAF media profiles that are conditionally required by the presentation profile. Players can rely on the availability of required CMAF tracks based on the presentation profile identifier. Other CMAF switching sets and media profiles can also be included in the CMAF presentation to enable additional features that might not be supported by all players conforming to the CMAF presentation profile. A CMAF presentation may conform to and signal multiple presentation profiles if all the required CMAF tracks for each presentation profile are available. Presentation profile conformance includes start alignment and synchronization of all CMAF tracks.

6.5.5.3 CMAF media profiles and ISO BMFF brands

CMAF file and media profile brands identify a CMAF track's file and media conformance to CMAF track and media profile constraints. Each media profile defines codec-specific properties in addition to those required by all CMAF tracks, such as the CMAF header sample entry, media sample format, CMAF fragment constraints, media sample presentation synchronization, scaling, mixing, layering, random access and encoding constraints such as codec profiles and levels. CMAF media profiles can specify constraints between multiple CMAF tracks in a CMAF switching set, and thereby imply player requirements necessary to seamlessly switch between CMAF fragments in those CMAF switching sets.

6.5.5.4 Signalling of supplemental data in CMAF tracks

CMAF defines brands, called "supplemental data brands" to identify the presence of additional information in a CMAF track that is not required by the CMAF media profile of that track. For example, NAL units, SEI messages, ISO BMFF Boxes can be added to an otherwise conformant CMAF track file, where this supplemental data does not make the CMAF track file non-conformant. Signalling its presence may be useful during production workflow, and to decoders for track selection and decoder initialization.

6.5.5.5 Single initialization CMAF switching set constraints identifier

Single initialization constraints specify that a CMAF switching set conforms to general CMAF switching set constraints and to additional CMAF media profile defined constraints that enable seamless switching and decoding without dependence on the CMAF header for each CMAF track. CMAF switching set single initialization constraints can be signalled in a manifest, such as the `bitstreamSwitching` attribute in a DASH manifest. Players can take advantage of these constraints by processing a single CMAF header only once prior to sequencing CMAF fragments from the CMAF switching set.

6.5.5.6 ISO BMFF segment brands for CMAF media objects

Segment type brands are specified in [7.2](#) that can be stored in a `SegmentTypeBox` that start-delimits and identifies the type of CMAF media object that follows.

6.5.5.7 Specification of CMAF presentation and media profiles

The general requirements for CMAF media profiles and CMAF presentation profiles are specified in [Clause 12](#).

This document contains the following CMAF profiles.

- CMAF presentation profiles are specified in [Clause A.1](#).
- CMAF media profiles and CMAF supplemental data are specified in [Clauses A.2, A.3, A.4](#), and [B.5](#).

Additional CMAF media profiles can be specified if they conform to the general CMAF track requirements, specify and register ISO BMFF brands, and define codec-specific CMAF track formats, codec constraints, and optional CMAF switching set constraints for seamless switching following the guidelines in [12.1](#).

Additional CMAF presentation profiles can be specified if they conform to the general CMAF requirements and specify a URL that identifies the conditionally required media profiles, as described in [12.2](#).

6.6 CMAF media object model

6.6.1 CMAF fragments

In the CMAF hypothetical application model, CMAF fragments are the media objects that are encoded and decoded.

Each CMAF media profile further constrains CMAF fragments and media sample data.

The CMAF media profiles in [Annex A](#) and their referenced track formats constrain CMAF fragments to be decodable independently of each other; for example, NAL structured video CMAF fragments contain one or more complete coded video sequences to make them randomly accessible and independently decodable. Audio CMAF fragments contain a sequence of audio access units, and subtitle CMAF fragments contain a single subtitle document in a media sample, constrained so that each CMAF fragment is randomly accessible and decodable. Common Encryption can also signal the necessary decryption parameters in each CMAF fragment to make it independently decryptable, in combination with an associated CMAF header and keys.

A CMAF fragment typically consists of one `MovieFragmentBox` and `MediaDataBox` pair, but can contain more than one of these pairs. When a fragment contains multiple pairs like this, each pair is called a CMAF chunk, and each CMAF chunk contains a consecutive subset of the CMAF fragment's media samples.

This is illustrated in [Figure 4](#).

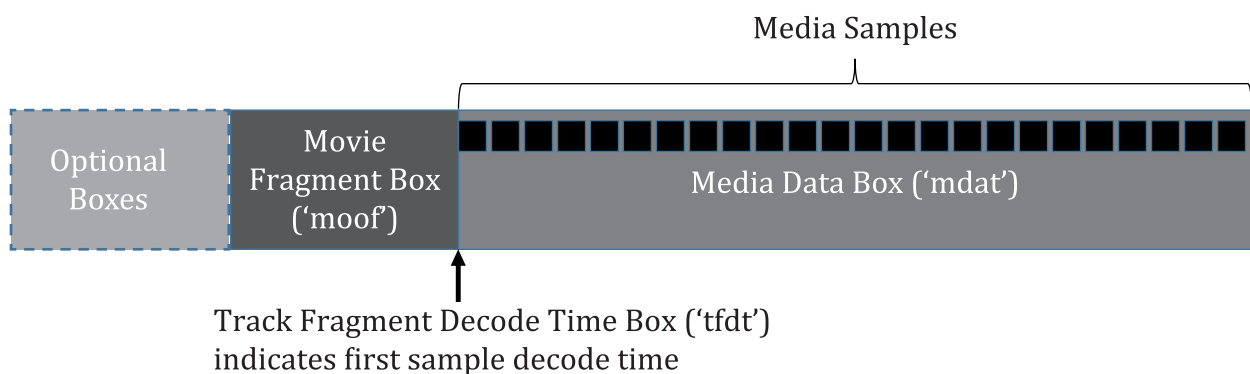


Figure 4 — Example CMAF fragment data structure containing a single CMAF chunk

6.6.2 CMAF tracks

A CMAF track is a continuous sequence of one or more CMAF fragments in presentation order conforming to a CMAF media profile and an associated CMAF header. The CMAF header contains a `MovieBox` sufficient to process and present all CMAF fragments in the CMAF track. A CMAF track can be produced by an encoder and ISO BMFF file packager, but it is made accessible in the form of CMAF addressable media objects that can be referenced as resources defined by an external media application specification.

This is illustrated in [Figure 5](#).

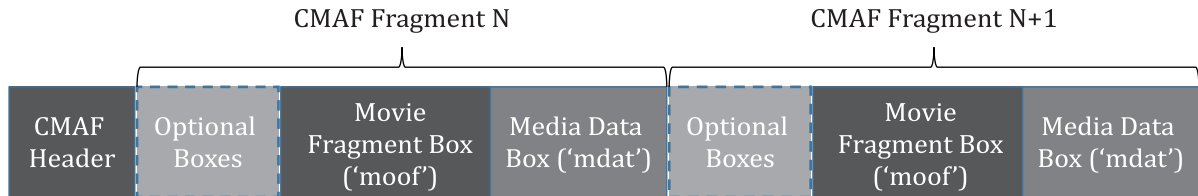


Figure 5 — CMAF track data structure

See [Clauses 7](#) through [11](#) for additional details on the construction of CMAF fragments and CMAF tracks of different media types.

6.6.3 CMAF track files

A CMAF track file is a CMAF addressable media object consisting of a CMAF track stored in a single ISO BMFF file. A CMAF track file is a stored CMAF track with the following constraints.

- It starts with a CMAF header.
- The CMAF header is followed by a continuous sequence of one or more CMAF fragments stored in presentation order.
- The first CMAF fragment has a `baseMediaDecodeTime` of zero.

Additional boxes, such as `SegmentIndexBoxes`, can be present between the CMAF header and the first CMAF fragment.

This is illustrated in [Figure 6](#).

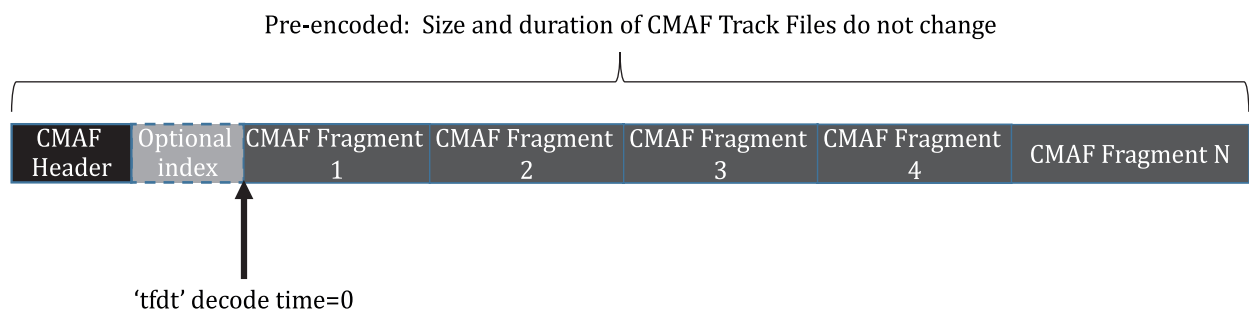


Figure 6 — CMAF track file, addressable media object

A sequence of one or more CMAF fragments can be requested from a CMAF track file resource, for example, using a resource identifier based on HTTP 1.1 consisting of the file URL and a byte range. CMAF resource identifiers can be listed in a manifest or determined by some other delivery format defined method, such as downloading a `SegmentIndexBox` to determine the byte ranges of CMAF fragments. The manifest and request method are out of scope of CMAF. A CMAF track file can also be downloaded or progressively downloaded.

6.6.4 CMAF segments

A CMAF segment is a CMAF addressable media object containing one or more consecutive CMAF fragments from a CMAF track. External application specifications can define how to reference a CMAF segment with a resource identifier that can be used by servers and manifests to reference and deliver each CMAF segment in a CMAF presentation.

This is illustrated in [Figure 7](#).

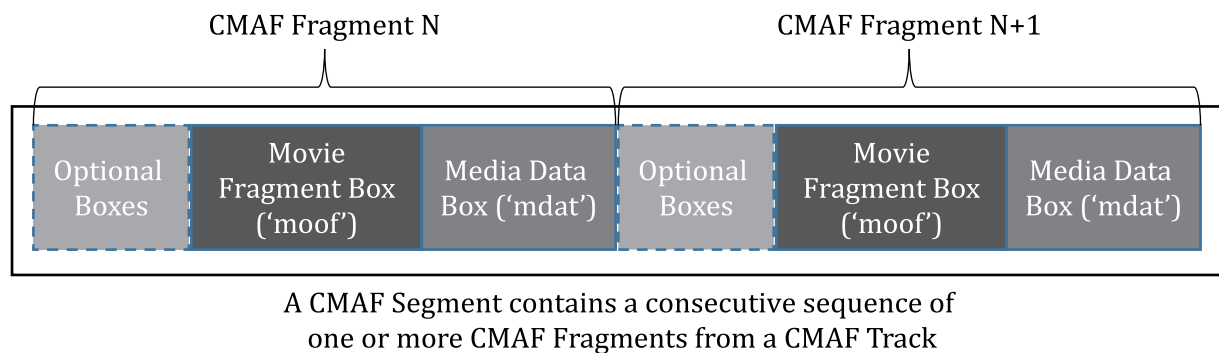


Figure 7 — Example CMAF segment, addressable media object

CMAF video fragment durations are typically 2 seconds to 6 seconds for coding efficiency, and CMAF segment durations are typically not greater than 10 seconds to 12 seconds to limit delivery latency in live streaming and limit bitrate rate adaptation response time.

Subtitle segment durations are usually close to video segment durations in live presentations to avoid increasing presentation delay. In a prerecorded CMAF presentation, a single subtitle CMAF segment can have a duration up to the duration of the CMAF track that contains it.

6.6.5 CMAF chunks

CMAF chunks are CMAF addressable media objects that contain a consecutive subset of the media samples in a CMAF fragment. CMAF chunks can be used by a delivery protocol to deliver media samples as soon as possible during live encoding and streaming, i.e., typically less than a second. The same media samples delivered in a CMAF segment containing one or more CMAF fragments would be delayed for the duration of the CMAF segment, i.e., typically several seconds. CMAF chunks enable the progressive encoding, delivery, and decoding of each CMAF fragment. A `MovieFragmentBox` at the start of each CMAF chunk provides access to the contained media samples. Broadcast, unicast, and multicast protocols can deliver and identify CMAF chunks through various methods, including resource IDs that include both a CMAF segment number corresponding to a CMAF fragment and the number of the CMAF chunk within the fragment to allow both fragment and chunk identification.

[Figure 8](#) illustrates that CMAF chunks inherit the media sample constraints of the CMAF fragment they are derived from. In this example of a video CMAF fragment containing a coded video sequence, the first media sample of the first CMAF chunk is SAP type 1 or 2, e.g. an IDR picture in an AVC CMAF fragment. The switching and splicing constraints are determined by the CMAF fragment. Boxes that lead the `MovieFragmentBox` in a CMAF fragment can also lead the `MovieFragmentBox` of the first CMAF chunk, although they are not shown in this example.

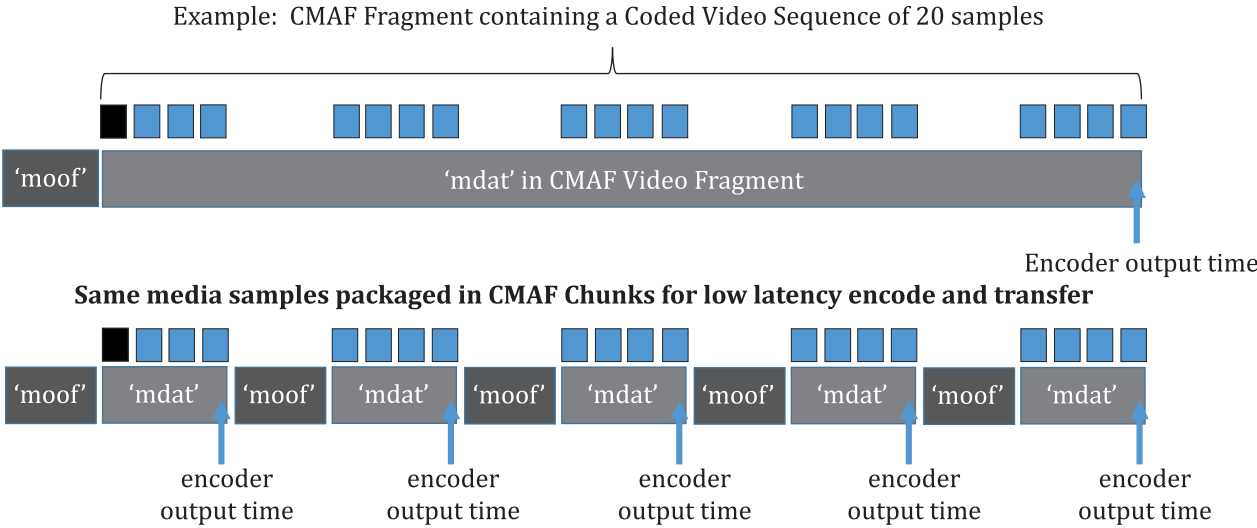


Figure 8 — CMAF chunk, addressable media objects

6.6.6 CMAF switching sets and adaptive switching

6.6.6.1 Overview

As illustrated in [Figure 9](#), CMAF switching sets contain time-aligned CMAF fragments that start with stream access points (SAP type 1 or 2) and timestamps to simplify switching between tracks by sequencing CMAF fragments from different CMAF tracks during playback.

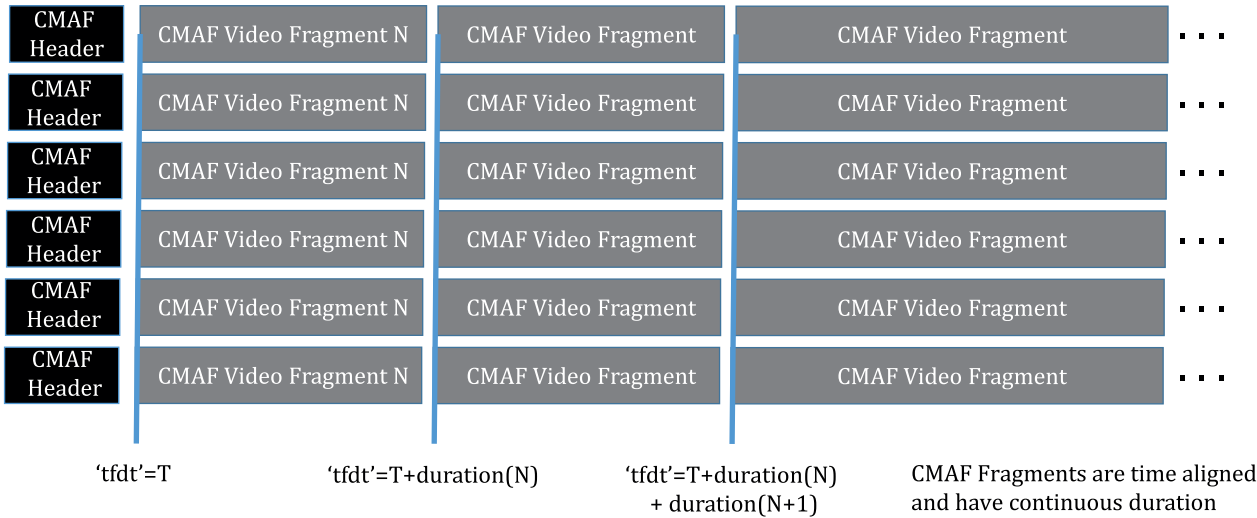


Figure 9 — A CMAF switching set

The manifest description of CMAF switching sets and the CMAF tracks they contain is out of scope of the CMAF specification, but manifests are expected to contain enough CMAF track and addressable media object information to enable automatic CMAF track selection and adaptive switching by players.

6.6.6.2 Decoding adaptively switched CMAF tracks

Each CMAF media profile can optionally specify one or more CMAF switching set constraints that enable players to seamlessly switch between alternative CMAF fragments in a CMAF switching set. CMAF switching set constraints enable player functions including decoder and decryptor initialization, video scaling, audio

mixing, media sample presentation timing, etc. that are intended to make adaptive switching as perceptually seamless as possible.

The media profiles in the CMAF specification reference or define CMAF switching set constraints that enable a stream of CMAF headers and CMAF fragments to decode on decoders without downloading overlapping CMAF fragments or requiring the use of multiple decoders.

An adaptively delivered CMAF switching set results in a timed sequence of CMAF headers and fragments a player selects from multiple CMAF tracks contained in a CMAF switching set. The encoding constraints of the CMAF switching set and CMAF media profile determine how the CMAF switching set can be encoded and decoded. They are independent from CMAF addressable media object packaging applied during storage or transport.

The CMAF track format specifies two types of CMAF switching set processing, determined by specific constraints on NAL structured video track formats referenced by some CMAF media profiles. If a CMAF switching set conforms to general constraints, a CMAF track's header is assumed be processed before every switch to that CMAF track. Additional constraints are specified for a CMAF switching set conforming to single initialization constraints that do not depend on processing a new CMAF header to switch to other CMAF tracks within a CMAF switching set.

[Figure 10](#) illustrates the player generated bitstreams resulting from the two types of CMAF switching set constraints and the decoding process each enables. Those two processes and bitstreams are described as “single initialization track switching” and “multiple initialization track switching”.

One method of single initialization track switching relies on a “common” CMAF header containing all referenced sample entries and decoding parameters in the CMAF switching set. It is assumed that players download CMAF headers only once per CMAF track. Consequently, new video parameter sets and indexes referenced by portions of a CMAF track encoded after the CMAF header has been downloaded would not be available to a player. Another method relies on parameter sets delivered in-band when those parameter sets are not included in the initially downloaded CMAF header or differ between CMAF tracks.

Audio CMAF switching sets typically only require single initialization because parameters that are allowed to change are contained in the audio media samples and CMAF fragments.

Subtitle CMAF tracks are typically not adaptively switched, and therefore have a single CMAF track per CMAF switching set.

See [7.3.4](#) for details on CMAF switching sets.

Single Initialization Switching Process



Multiple Initialization Switching Process

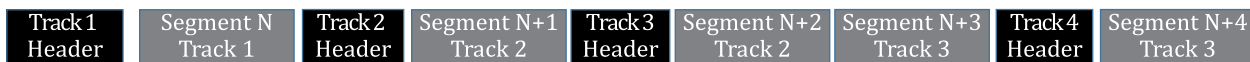


Figure 10 — Adaptively switched resource streams for different CMAF switching set constraints

6.6.6.3 Displaying adaptively switched video CMAF tracks

The normative behaviour of a video decoder processing a single track is well-specified in each video codec specification, such as AVC, but display processing is considered out of scope of those codec specifications. For instance, AVC does not specify the conversion by a display processor of YCbCr 4:2:0 video spatial samples and colour subsamples output by a decoder to 32-bit or 48-bit RGB pixels in a size, colour space, resolution, and refresh rate appropriate for a specific display.

Adaptive switching, enabled by CMAF media profile constraints, allows a player's decoder and display system to deterministically scale different video spatial sampling used to encode different CMAF fragments to the same display size, position, and refresh rate. The encoding of CMAF tracks in a CMAF switching set is constrained so that a sequence of CMAF fragments can be accurately rescaled to the same presentation height and width with precise registration and visual continuity of the video content, including its size, shape, location, colour, brightness, etc. The variations allowed between CMAF tracks conforming to CMAF switching set constraints determine the scaling behaviour needed in display processors, but display processing and other player requirements are not directly specified by CMAF.

6.6.6.4 Video parameter set storage and processing

The CMAF header of a CMAF video track needs to be processed on each CMAF track switch if decoding parameters and parameter indexes are not identical between CMAF tracks, and parameters that can change are not signalled in each CMAF fragment. This is often the case when each video CMAF track stores the video parameter set(s) for that CMAF track in its CMAF header sample table, e.g. the 'avc1' sample and track format. In that case, a decoder and display processor need to index the correct encoded video spatial sample counts and cropping parameters from a sample entry containing the correct SPS, PPS, and VUI information matching the parameter index encoded in each video NAL slice header. If CMAF tracks in a CMAF switching set use different spatial subsampling, they will have different parameter sets and sample entries, and different CMAF tracks might use the same parameter set ID value for a different parameter set. That makes it necessary to process the CMAF header and sample table of each CMAF track when switching to that CMAF track so that the correct decoding parameters in a sample entry are in effect.

Multiple initialization switching can be used to process all CMAF switching sets, but may not result in seamless presentation when a system processes the entire CMAF header rather than just the video parameters allowed to change in a CMAF switching set, such as height, width, and cropping parameters. Multiple initialization adaptive switching is necessary for a NAL structured video CMAF switching set unless it conforms to single initialization constraints specified by its CMAF media profile.

A CMAF header of a CMAF video track only needs to be processed once for a CMAF switching set conforming to single initialization constraints. There could be several ways to achieve this functionality that can be specified by each CMAF media profile.

Two methods of specifying single initialization constraints are the following.

a) A common CMAF header and sample entries for a CMAF switching set

In this case, each CMAF header and its sample table contain all parameter sets and sample entries used to encode CMAF fragments in the CMAF switching set. All the parameter indexes are unique and correspond to the index values encoded in the video NAL slice headers. There would be a different sample entry for each different subsampling used in the CMAF switching set.

b) Generic CMAF headers and parameter sets stored in each CMAF fragment

In this case, each CMAF header contains initialization information in the video configuration record and/or additional boxes as specified in [Clause 9](#) and/or a manifest, but the SPS, PPS, and VUI information used for decoding is stored and indexed within each coded video sequence and CMAF fragment. The initialization information is sufficient for the CMAF media profile identified by brand in each CMAF header, and re-initialization is only necessary if switching to a higher CMAF media profile. For example, when a CMAF switching set contains some CMAF tracks that conform to both SD and HD profile, but others that only conform to HD profile, a decoder initialized to HD could switch all CMAF tracks without reinitializing, but a decoder initialized to SD would need to limit playback to SD media profile CMAF tracks, or reinitialize to HD, if possible.

For NAL structured video, single initialization encoding can be accomplished by encoding the Sequence Parameter Set and Picture Parameter Set NALs in the IDR access unit that begins that coded video sequence (e.g. using the 'avc3' sample description and track format). Then the decoder and display processor can reference the parameter set necessary to decode and scale each CMAF fragment from the parameter set stored in the first media sample of each CMAF fragment.

Continuity between a sequence of CMAF presentations is not specified by CMAF (typically a manifest defines sequences of CMAF presentations). However, seamless playback of sequenced CMAF presentations is possible with CMAF presentations that conform to the same single initialization constraints and signal in-band parameters in each CMAF fragment.

6.6.7 CMAF selection sets and late binding

A CMAF selection set is a set of CMAF switching sets, where each CMAF switching set encodes an alternative aspect of the same presentation over the same presentation time, for example, different audio languages, video camera angles, video formats, or codecs.

CMAF players are expected to select one CMAF switching set from each CMAF selection set at the start of playback based on player compatibility and user preferences. Users or playback applications may switch between CMAF switching sets in a CMAF selection set during playback, but seamless presentation is not expected, either because the content differs (e.g. a different language or camera view) or because CMAF fragment time alignment and decoding are not constrained to decode seamlessly.

Aligned CMAF switching sets are encodings of the same source content with time-aligned CMAF fragments to enable compatible players to seamlessly switch between tracks in different but aligned CMAF switching sets that contain CMAF tracks that do not conform to all the constraints of a single CMAF switching set.

For example, in [Figure 11](#), the SD, HD, and UHD10 CMAF media profile CMAF switching sets could be aligned CMAF switching sets, and some players could seamlessly switch between them. In one case, only the encryption key might be different between the SD, HD and UHD CMAF switching sets. In another case, codecs could differ. Each CMAF switching set individually conforms to normal constraints for its media profile, and each player can determine if it can process the differences between aligned CMAF switching sets seamlessly.

Late binding allows CMAF tracks to be encoded once and used in different combinations. See [6.3](#) for more information.

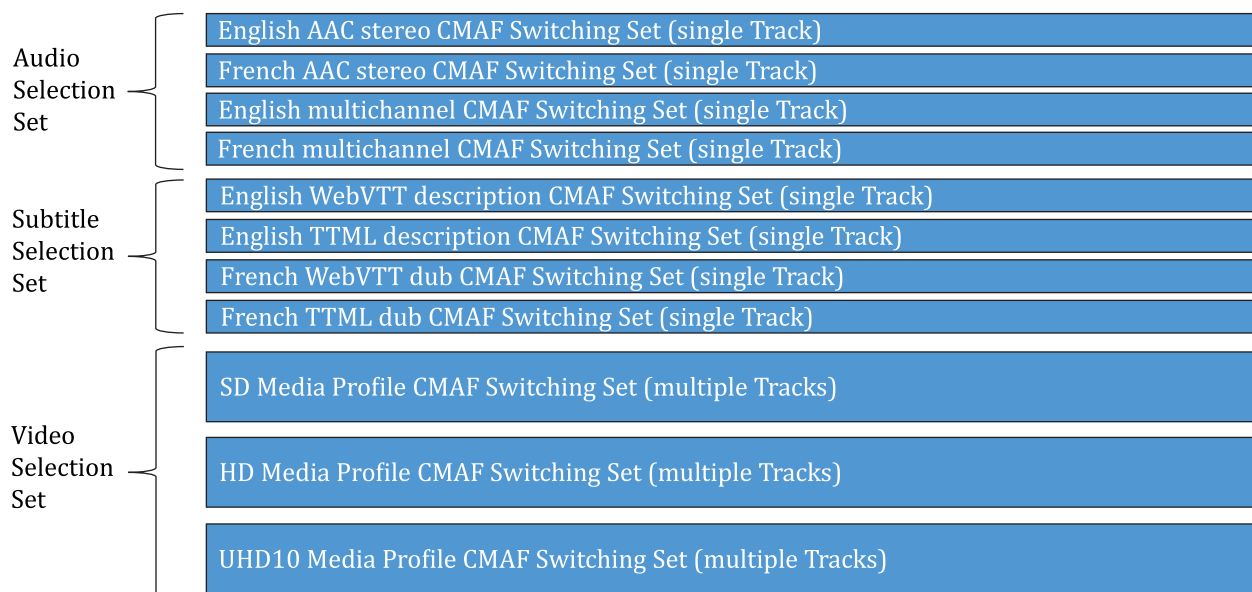


Figure 11 — Example of CMAF selection sets

The description of selection sets in a manifest is not defined by the CMAF specification; however, each manifest format is expected to provide sufficient information to enable players to select optimal CMAF switching sets from CMAF selection sets.

When there are multiple tracks in a selection set that differ by language, the language of each track needs to be identified to at least the precision needed to differentiate the tracks, using the language field and optionally the `ExtendedLanguageBox` as defined in ISO/IEC 14496-12.

When there is more than one track that is visually presented (e.g. video and subtitles), the visual layering of the tracks is indicated by the track header's layer field. CMAF recommends that subtitle CMAF tracks be positioned on layer -1 (ISO BMFF tracks with lower layer numbers overlay higher layer numbered tracks) and video CMAF tracks may be packaged with layer zero, or values relative to subtitles per the content creator's intended effect.

6.6.8 CMAF presentation timing model

There are multiple timelines involved in synchronizing a CMAF presentation. Each timeline has a timescale in units per second, increases at a continuous rate over time, and has an origin where the measure equals zero.

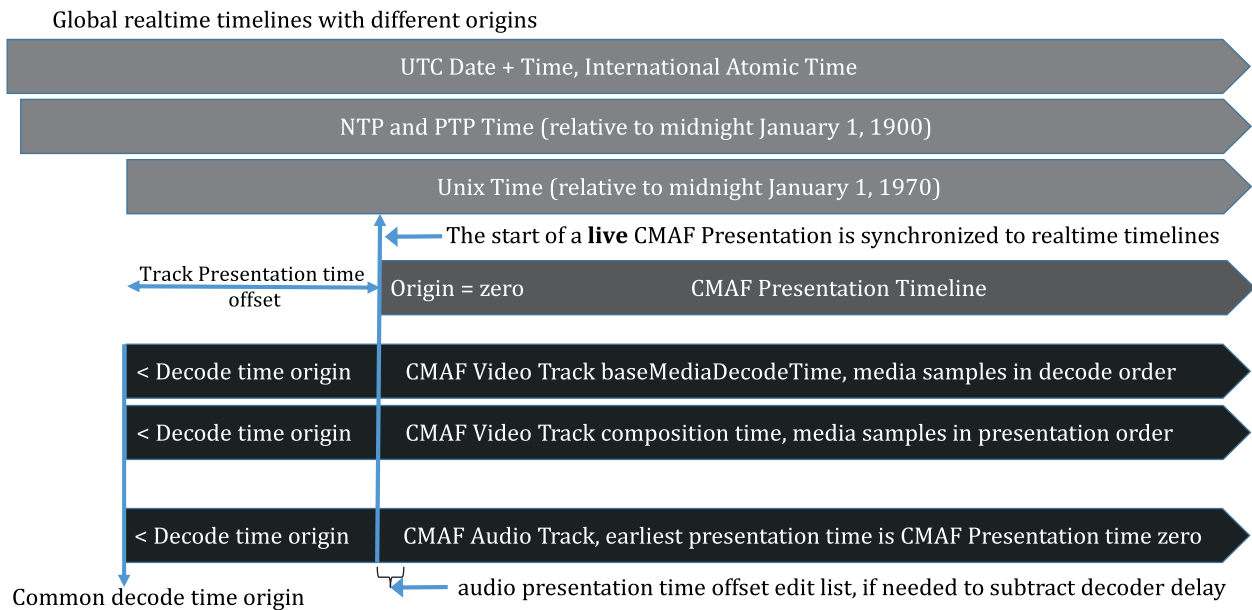


Figure 12 — CMAF timelines and synchronization model

All CMAF tracks in a CMAF presentation are specified in [Clause 7](#) to measure `baseMediaDecodeTime` in the `TrackFragmentBaseMediaDecodeTimeBox` from the same timeline origin and adjust each media sample's presentation time relative to its decode time using composition offsets and offset edit lists, when necessary. The resulting presentation timelines and decode timelines all reference the same (coincident) timeline origin. The timelines and their relationships are illustrated in [Figure 12](#).

The timelines are:

- Track decode time** — Is determined by the storage sequence and duration of each media sample, and for any media sample, equals the sum of all prior media sample durations added to the `baseMediaDecodeTime` of the first CMAF fragment.
- Track composition time** — Is determined by each sample's composition offset in the `TrackRunBox` relative to its decode time and is used to reorder video media samples to their presentation order.
- Track presentation time** — Is determined by applying any offset edit list in the CMAF header to each media sample's composition time, as shown for a CMAF audio track in [Figure 12](#).
- CMAF presentation timeline** — Is defined in CMAF to start with a CMAF presentation time equal to zero, at the earliest video media sample presentation time, or the earliest audio media sample presentation time if there are no video CMAF tracks. The CMAF track presentation time at CMAF presentation time zero is its "presentation time offset", which can also be represented in some manifest formats, such as DASH.
- Wall clock time** — This timeline is important for live presentations, e.g. UTC time at the time of CMAF chunk availability. A wall clock timestamp may be stored in a `ProducerReferenceTimeBox` in a CMAF

chunk or fragment, and that timestamp linked to a media sample decode time. Manifests can indicate the wall clock time that coincides with the start of a CMAF presentation so a player can determine when each CMAF fragment or chunk will be available for download.

CMAF tracks in a CMAF presentation are required to use the same decode timeline origin, similar to the shared movie timeline of multiple tracks in one ISO BMFF file. But CMAF tracks can start at a non-zero decode time (`baseMediaDecodeTime`) stored in the earliest video CMAF fragment. The CMAF presentation timeline starts at zero, coincident with the earliest media sample presentation time, and is equivalent to the ISO BMFF movie timeline.

As illustrated in [Figure 13](#), CMAF tracks with audio CMAF fragments that overlap the earliest video media sample are intended to start presenting simultaneously with the earliest video media sample. Audio CMAF track presentation time is determined by the value of `baseMediaDecodeTime` recorded in each CMAF chunk or fragment, and any presentation time offset in an edit list, if an edit list is present in the CMAF track's CMAF header.

ISO BMFF defines files and requires that the decode time of each media sample is the sum of all prior media sample durations in that track in stored order. The first media sample in each track in a file therefore has a decode time of zero. Movie fragmentation is optional. In contrast, all media samples in CMAF tracks are stored in movie fragments, and the first CMAF fragment can have a non-zero `baseMediaDecodeTime` in the `TrackFragmentBaseMediaDecodeTimeBox`. The decode time of each media sample equals the sum of prior media sample durations in the CMAF track added to the `baseMediaDecodeTime` of the first CMAF fragment in the CMAF track. The decode time of each media sample also equals the sum of prior media sample durations in the CMAF fragment that contains it added to the CMAF fragment's `baseMediaDecodeTime`.

Manifests can specify a presentation time offset for each CMAF switching set to determine the CMAF track presentation time at the start of each CMAF presentation. The start of CMAF presentations is assumed to be the earliest video media sample when video is included in a CMAF presentation, otherwise the earliest audio media sample. To maintain audio, video, and subtitle synchronization encoded in the CMAF tracks, the presentation time offset of every CMAF switching set in a manifest will be equivalent, which means equal presentation time offsets, but possibly different integer values and timescales per CMAF switching set. Matching presentation time offsets plus CMAF presentation time in each CMAF track being presented maintains the encoded synchronization between audio, video, and text content on playback.

A manifest can select different start and end times with presentation time offsets and durations to present different timespans of the CMAF presentation timeline to play a portion of a CMAF presentation.

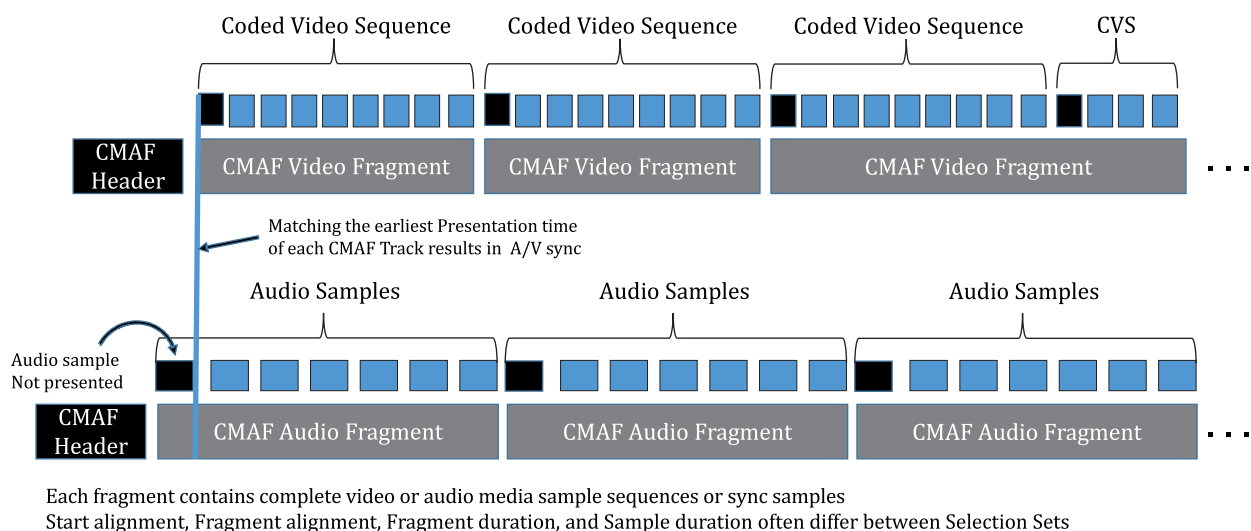


Figure 13 — Audio/video synchronization and start alignment

CMAF tracks containing video can use negative composition offsets where necessary to reorder media samples without adding composition delay, so that the earliest composition time and presentation time of each CMAF fragment equal its earliest decode time (with the optional exception of CMAF track files,

mentioned above). The `baseMediaDecodeTime` in the `TrackFragmentBaseMediaDecodeTimeBox` is the earliest media sample presentation time of each CMAF fragment, and the `baseMediaDecodeTime` of the earliest video CMAF fragment is the presentation time offset of the CMAF track at CMAF presentation time zero.

CMAF tracks, illustrated in [Figure 6](#), can also be encoded using positive composition offsets and packaged with an edit list in the CMAF header to remove composition delay.

During random access and “trick play” (fast forward, reverse, slow motion, jump, etc.), the `TrackFragmentBaseMediaDecodeTimeBox` and the `TrackRunBox` can be used to determine each media sample decode time relative to the CMAF fragment `baseMediaDecodeTime`, then any edit list present in the associated CMAF header can be used to calculate the presentation time offset relative to each media sample’s composition time.

Decode time discontinuities between CMAF fragments can result from starting a new CMAF presentation in a playlist, for example, in sequenced programs or programs interspersed with prerecorded ads. A manifest can locate the new CMAF presentation on its manifest presentation timeline and adjust CMAF switching set presentation time offsets accordingly.

However, in cases where a decode time discontinuity results from one or more missing or damaged CMAF fragments in a continuous CMAF presentation, players can use the `baseMediaDecodeTime` of the next available CMAF fragment to resume presentation synchronized to the current presentation timeline. Error concealment could include repeating video frames and skipping over missing content in the case of on demand or buffered media, but all CMAF tracks have to skip the same duration to maintain A/V synchronization.

In order to indicate timeline discontinuities caused by gaps, a manifest should signal the existence of a gap. Content with gaps is not strictly compliant to CMAF, conformance can only be guaranteed for the continuous parts between gaps.

6.6.9 Manifest information

Although CMAF does not define the form or the content of the manifest, it does define its role. A manifest is a document that describes one or more CMAF presentations, e.g. an MPEG DASH Media Presentation Description (MPD). A manifest is responsible for describing the combination and synchronization of independently packaged CMAF tracks grouped in CMAF switching sets and selection sets to form a synchronized multimedia presentation.

A manifest provides the player with information to select, initialize, start align, and synchronize the CMAF track(s) to be played and identify CMAF media objects as resources to access and possibly download them. CMAF tracks and CMAF fragments contain sufficient information to enable decryption, decoding, and presentation scheduling. A manifest can also provide information on delivery protocol, network management, authorization, license acquisition, etc., in addition to resource identification and presentation description. The manifest can also signal that tracks conform to a CMAF media profile.

In cases where there are multiple CMAF tracks in a CMAF switching set or multiple CMAF presentations in a sequence, a manifest can indicate to a player that the CMAF switching set conforms to a CMAF media profile and CMAF switching set constraints, so a player can select and initialize compatible CMAF tracks and can switch seamlessly between those CMAF tracks, reinitializing only when necessary.

6.6.10 CMAF addressable media objects, resources, and resource identifiers

CMAF headers, CMAF chunks, and CMAF fragments can be packaged and referenced as CMAF addressable media objects for storage and delivery. Each CMAF addressable media object can be referenced as a resource as specified by an external specification, e.g. MPEG DASH (ISO/IEC 23009-1).

External specifications can define resource identifiers and the CMAF addressable media objects they identify so that servers, CDNs, and manifests can access CMAF addressable media objects and sequence them for decoding and presentation. A common example is a CMAF addressable media object packaged as the body of an HTTP response and identified as a resource by a unique segment URL. A CMAF track file can be segmented during delivery by adding a byte range to each HTTP request for the file’s URL, using byte ranges matching one or more CMAF fragments or the CMAF header of the CMAF track file.

Use of the same URI per CMAF addressable media object is recommended to improve efficiency in content distribution networks and caches, even when different manifests or delivery protocols are used. The use of different URIs for the same CMAF addressable media object causes duplicate storage and delivery of the same CMAF media objects. For example, in cases where the same CMAF segments can be delivered over broadcast, multicast, and unicast, a CMAF player can request a CMAF segment with a single resource identifier and retrieve it from the player's cache regardless of which network path and encapsulation delivered it.

7 CMAF track format

7.1 Overview

The CMAF track format is derived from the ISO base media file format in this clause and structural brands are specified. At this point, the 'cmfc' and the 'cmf2' CMAF structural brands are defined. The 'cmf2' brand further restricts the 'cmfc' brand.

Several CMAF media objects are derived from the CMAF track format.

7.2 CMAF brands

[Clause 7](#) defines the requirements and constraints that apply to all CMAF tracks.

A CMAF track should include all structural CMAF brands that it conforms to, indicating conformance with [Clause 7](#) and [Clause 8](#) (specifying file format and optional encryption, but not media encoding). The file brands also declare that boxes specified in common encryption (ISO/IEC 23001-7) and DASH (ISO/IEC 23009-1) may be present.

The common constraints of all structural file type brands are described in [subclause 7.3.2.2](#).

Media profiles and their brands may specify additional box requirements and CMAF track constraints, such as the CMAF media profiles that reference the NAL structured video CMAF track format defined in [Clause 9](#), incorporating boxes specified in ISO/IEC 14496-15, and AAC audio CMAF track format specified in [Clause 10](#), incorporating boxes specified in ISO/IEC 14496-3.

CMAF also specifies addressable media objects in [subclause 7.3.3](#) that are derived from CMAF tracks and have segment type brands that may be included in a `SegmentTypeBox` prepended to an addressable media object to identify its type (see [Table 1](#)).

Table 1 — CMAF brands

Brand	Location	Conformance requirements
cmfc	FileTypeBox and SegmentTypeBox	7.6
cmf2	FileTypeBox and SegmentTypeBox	7.7
cmfs	SegmentTypeBox	7.3.3.1
cmfl	SegmentTypeBox	7.3.3.2
cmff	SegmentTypeBox	7.3.2.3
cmfr	SegmentTypeBox	7.3.2.4

An ISO brand, such as 'iso9', that indicates the ISO BMFF boxes actually present in each CMAF track should be listed in `compatible_brands` to improve interoperability with file readers. The 'isoX' brand (where "X" represents a number or letter) specified in each new edition of ISO/IEC 14496-12 determines what box versions are allowed in CMAF tracks, unless specifically constrained by the common constraints of CMAF structural brands as specified in this clause.

If any of the structural CMAF brands is the `major_brand`, the `minor_version` shall be set to 0, and file names should use the file extensions in [Table 2](#). Otherwise, file names should use the file extension and Internet Media Type specified to match the major brand, e.g. `*.mp4`, `*.3gp`, `*.uvv` and `*.uva`.

NOTE It is expected that file readers read possible future versions of CMAF that increment the `minor_version` number.

Table 2 — Common media application format file extensions

Track type	File extension	Internet media type (MIME type)
Video	.cmfv	video/mp4
Audio	.cmfa	audio/mp4
Text (subtitle)	.cmft	application/mp4

Each CMAF track has a CMAF header associated with it, although the CMAF header and CMAF fragments might not be stored as an ISO BMFF file, and might not be stored at all, i.e., only existing temporarily as streams.

7.3 CMAF media objects

7.3.1 CMAF boxes

CMAF tracks shall include the following ISO BMFF boxes with nesting, optionality, and cardinality specified in [Table 3](#) through [Table 5](#), with box definitions incorporated by reference to the clauses listed in the “Specification” column that reference ISO/IEC 14496-12, ISO/IEC 23001-7, or ISO/IEC 23009-1.

Some boxes are additionally constrained by CMAF, as specified by clauses in this document referenced in the “Constraints” column of [Table 3](#) through [Table 5](#). The versions and flags of boxes used in a CMAF track shall conform to the versions and flags of the same boxes as defined by the 'iso9' brand unless additionally constrained by this document.

CMAF addressable media objects are derived from the CMAF track, CMAF header, CMAF chunk, and CMAF fragment recorded media objects and specify additional box constraints in subclause [7.3.3](#).

Additional boxes and constraints are specified by CMAF track formats and CMAF media profiles for specific audio, video, and subtitle formats (see [subclause 7.4.4](#)).

In [Table 3](#), [Table 4](#), and [Table 5](#) the “Format Req.” column indicates the number of boxes that are required to be present in a CMAF track, where:

- “*” means “zero or more” may be present;
- “+” means “one or more” shall be present;
- “0/1” indicates only one box may be present, and it is optional;
- “1” indicates one box shall be present;
- “CR” indicates that a box is conditionally required as specified in the CMAF track format or a specific media profile and shall be present under those conditions (see “Constraints” column).

Table 3 — CMAF header boxes

NL 0	NL 1	NL 2	NL 3	NL 4	NL 5	Format Req.	Specification	Constraints	Description
ftyp						1	ISO/IEC 14496-12	CMAF 7.2	File type and Compatibility
moov						1	ISO/IEC 14496-12		Container for functional metadata
	mvhd					1	ISO/IEC 14496-12	CMAF 7.5.1	Movie header
	trak					1	ISO/IEC 14496-12		Container for each track
		tkhd				1	ISO/IEC 14496-12	CMAF 7.5.4	Track header
		edts				CR	ISO/IEC 14496-12	CMAF 7.5.13	Edit box
			elst			1	ISO/IEC 14496-12	CMAF 7.5.13	Edit list box

Table 3 (continued)

NL 0	NL 1	NL 2	NL 3	NL 4	NL 5	Format Req.	Specification	Constraints	Description
		mdia				1	ISO/IEC 14496-12		Track media information
			mdhd			1	ISO/IEC 14496-12	CMAF 7.5.5	Media header
			hdlr			1	ISO/IEC 14496-12		Declares the media handler type
			elng			0/1	ISO/IEC 14496-12		Extended language tag
			minf			1	ISO/IEC 14496-12		Media information container
				vmhd		CR	ISO/IEC 14496-12	CMAF 7.5.6	Video media header
				smhd		CR	ISO/IEC 14496-12	CMAF 7.5.7	Sound media header
				nmhd		CR	ISO/IEC 14496-12		Null media header, overall information (some tracks only)
				sthd		CR	ISO/IEC 14496-12	CMAF 7.5.8	Subtitle media header
				dinf		1	ISO/IEC 14496-12		Data information box
					dref	1	ISO/IEC 14496-12	CMAF 7.5.9	Data reference box, declares source of media data in track
				stbl		1	ISO/IEC 14496-12	CMAF 7.5.12	Sample table box, container for the time/space map
					stsd	1	ISO/IEC 14496-12	CMAF 7.5.10	Sample descriptions (see Table 4 for additional details)
					stts	1	ISO/IEC 14496-12	CMAF 7.5.12	Decoding, time to sample
					stsc	1	ISO/IEC 14496-12	CMAF 7.5.12	Sample-to-chunk
					stsz/ stz2	1	ISO/IEC 14496-12	CMAF 7.5.12	Sample size box
					stco	1	ISO/IEC 14496-12	CMAF 7.5.12	Chunk offset
					sgpd	CR	ISO/IEC 14496-12	CMAF 7.5.18	Sample group description box
					stss	CR	ISO/IEC 14496-12	CMAF 7.5.12	Sync sample box
		udta				0/1	ISO/IEC 14496-12		User data box
			cp rt			*	ISO/IEC 14496-12		Copyright box
			kind			*	ISO/IEC 14496-12	CMAF 7.5.3	Track kind box
	mvex					1	ISO/IEC 14496-12		Movie extends box
		mehd				0/1	ISO/IEC 14496-12	CMAF 7.5.1	Movie extends header
		trex				1	ISO/IEC 14496-12	CMAF 7.5.14	Track extends box
	pssh					*	ISO/IEC 23001-7	CMAF 7.4.3	Protection system specific header box

Table 4 — CMAF header protected sample entry boxes

NL 5	NL 6	NL 7	NL 8	Format Req.	Specification	Constraints	Description
stsd				1	ISO/IEC 14496-12		Sample description box
	sinf			CR	ISO/IEC 14496-12	CMAF 7.5.11	Protection scheme information box
		frma		1	ISO/IEC 14496-12		Original format box
		schm		1	ISO/IEC 14496-12		Scheme type box
		schi		1	ISO/IEC 14496-12		Scheme information box
			tenc	1	ISO/IEC 23001-7	CMAF 7.4.1	Track encryption box

NOTE [Table 4](#) is a continuation of [Table 3](#) showing nesting levels 5 to 8 separately to reduce table width.

Table 5 — CMAF chunk, CMAF fragment, CMAF segment, and CMAF track file boxes

NL 0	NL 1	NL 2	NL 3	NL 4	NL 5	Format Req.	Specification	Constraints	Description
styp						0/1	ISO/IEC 14496-12		Segment type
prft						0/1	ISO/IEC 14496-12		Producer reference time
emsg						*	ISO/IEC 23009-1	CMAF 7.4.5	Event message
moof						+	ISO/IEC 14496-12		Movie fragment
	mfhd					1	ISO/IEC 14496-12		Movie fragment header
	pssh					*	ISO/IEC 23001-7	CMAF 7.4.3	Protection system specific header box
	traf					1	ISO/IEC 14496-12		Track fragment
		tfhd				1	ISO/IEC 14496-12	CMAF 7.5.16	Track fragment header
		tfdt				1	ISO/IEC 14496-12		Track fragment base media decode time
		trun				1	ISO/IEC 14496-12	CMAF 7.5.17	Track fragment run box
		senc				0/1	ISO/IEC 23001-7	CMAF 8.2.2.1	Sample encryption box
		saio				CR	ISO/IEC 14496-12	CMAF 8.2.2.1	Sample auxiliary information offsets box
		saiz				CR	ISO/IEC 14496-12	CMAF 8.2.2.1	Sample auxiliary information sizes box
		sbgp				*	ISO/IEC 14496-12		Sample to group box
		sgpd				*	ISO/IEC 14496-12	CMAF 7.5.17	Sample group description box
		subs				CR	ISO/IEC 14496-12	CMAF 7.5.19	Sub-sample information box
mdat						+	ISO/IEC 14496-12	CMAF 7.5.18	Media data container for media samples

7.3.2 CMAF track media objects

7.3.2.1 CMAF header

A CMAF header as defined in [3.1.2](#) conforms to the following constraints.

- a) A CMAF header shall contain the set of boxes in [Table 3](#) and [Table 4](#) with the conditions and optionality listed.

- b) Each CMAF header shall form a valid CMAF track, as specified in [subclause 7.3.2.2](#), when followed by a continuous sequence of associated CMAF fragments in decode order.
- c) A CMAF header shall be conformant with ISO/IEC 14496-12 and the following additional constraints and requirements.
 - 1) The CMAF header shall start with a `FileTypeBox`.
 - 2) The CMAF header shall include exactly one `MovieBox`.
 - 3) The `MovieBox` shall start with a `MovieHeaderBox`, as constrained in [subclause 7.5.1](#).
 - 4) The `MovieBox` shall contain exactly one track containing media data as specified in [subclause 7.3.2.2](#).

NOTE Timed metadata tracks can be provided as separate CMAF tracks in a separate selection set.
 - 5) The `MovieBox` shall contain a `MovieExtendsBox`, as defined in ISO/IEC 14496-12, to indicate that the file contains `MovieFragmentBoxes`.
 - 6) The `MovieExtendsBox` may contain a `MovieExtendsHeaderBox`, as defined in ISO/IEC 14496-12, and if so, shall provide the overall duration of the CMAF track. If the duration is unknown, this box shall be omitted.
- d) If the CMAF header requires sample entries with a decoder configuration record, as specified by a CMAF media profile, with decoding and rendering parameter limits, such as codec profile, level, image height, width, etc., then the first sample entry shall contain parameters equal or greater than the maximum values contained in all CMAF fragments in the track to enable a single initialization of media decoding and rendering systems to render all CMAF fragments in the CMAF track.

7.3.2.2 CMAF track

Table 6 — CMAF track structure

NL 0	Format Req.	Specification	CMAF constraints	Description
CMAF Header	1		CMAF 7.3.2.1	CMAF header
CMAF Fragment	1+		CMAF 7.3.2.4	CMAF fragment

A CMAF track as defined in [3.2.1](#) has the structure shown in [Table 6](#) and conforms to the following constraints.

- a) A CMAF track shall conform to at least one structural CMAF brand and contain the set of boxes in [Table 3](#), [Table 4](#), and [Table 5](#), with the conditions and optionality listed.
- b) The concatenation of a CMAF header and all CMAF fragments in the CMAF track in consecutive decode order shall be a valid fragmented ISO BMFF file, with the exception that the first CMAF fragment in a CMAF track may have a non-zero `baseMediaDecodeTime`.
- c) Each CMAF fragment in a CMAF track shall have `baseMediaDecodeTime` equal to the sum of all prior CMAF fragment durations added to the first fragment's `baseMediaDecodeTime`. A CMAF fragment duration is the sum of the media sample durations, documented in the `TrackFragmentRunBox` in the `MovieFragmentHeaderBox`.

NOTE Valid CMAF tracks do not have media time discontinuities resulting from missing media samples or fragments. Gaps in decode time can result in audio-video synchronization errors. For recommendations on handling missing media samples and missing CMAF fragments, see [Annex F](#).

- d) Each CMAF track contains a single ISO BMFF track and `TrackBox`, as determined by CMAF header constraints specified in [subclause 7.3.2.1](#).

Additional constraints specific for encryption, video, audio, and subtitle CMAF tracks are specified in [Clauses 8, 9, 10](#) and [11](#), respectively, and are derived from the general CMAF track format specified in this clause.

7.3.2.3 CMAF chunk

CMAF chunks as defined in [3.1.4](#) are the smallest CMAF media object that can be encoded, and they can be referenced as addressable media objects.

Table 7 — CMAF chunk structure

NL 0	Format Req.	Specification	CMAF constraints	Description
styp	0/1	ISO/IEC 14496-12	CMAF 7.2	Segment type signalling compatibility to CMAF chunk
prft	0/1	ISO/IEC 14496-12		Producer reference time
emsg	*	ISO/IEC 23009-1	CMAF 7.4.5	Event message
moof	1	ISO/IEC 14496-12	CMAF Table 5	Movie fragment box and the boxes it contains
mdat	1	ISO/IEC 14496-12	CMAF 7.5.19	Media data container for media samples

Each CMAF chunk conforms to the following constraints:

- A CMAF chunk shall contain the boxes in [Table 7](#) with the conditions and optionality indicated.
- The `MovieFragmentBox` shall conform to the constraints of the structural CMAF brand constraints specified in [Table 5](#), such as containing only one `TrackFragmentBox` that contains only one `TrackRunBox`.

NOTE 1 Since there is only one `TrackRunBox` per `MovieFragmentBox`, all media samples of a CMAF chunk are located in a single track run in one `MediaDataBox`.

- The `MediaDataBox` shall contain all media samples referenced by the `TrackRunBox` and should immediately follow the `TrackRunBox` in byte order.
- All media samples in a CMAF chunk shall be addressed by byte offsets in the `TrackRunBox` that are relative to the first byte of the `MovieFragmentBox`. See ISO/IEC 14496-12.
- A CMAF chunk shall contain a consecutive subset of the media samples of a CMAF fragment.
- A CMAF chunk shall contain a `TrackFragmentDecodeTimeBox` containing the `baseMediaDecodeTime` of the first media sample.
- A sequence of CMAF chunks that spans the decode time and duration of a CMAF fragment containing those media samples shall contain all the media samples in the CMAF fragment, without duplicates, i.e., CMAF chunks in a CMAF track shall not overlap or have gaps in decode time.
- The decode time and CMAF track presentation times of a media sample in a CMAF track shall not change whether the sample is contained in a CMAF chunk or a CMAF fragment.

NOTE 2 Because the CMAF fragment and the first CMAF chunk both start with the same media sample, they both have the same `baseMediaDecodeTime`.

- A CMAF chunk may include a `SegmentTypeBox` preceding the `MovieFragmentBox` and may be referenced as a CMAF addressable media object. The `SegmentTypeBox` should include the CMAF chunk brand '`cmf1`' and may include any valid `compatible_brands` listed in the `FileTypeBox` of the CMAF track's CMAF header.
- CMAF chunks may be referenced as CMAF addressable media objects, and optional file level boxes may be prepended when they are formatted as CMAF addressable media objects.

NOTE 3 Preceding boxes indicated in [Table 7](#) can be prepended to CMAF chunks at time of encoding or prepended to CMAF chunks and CMAF segments at time of delivery. It is expected that players will process the boxes prepended to the start of each CMAF addressable media object, but they will ignore these boxes when located within a CMAF addressable media object, i.e., not prepended to the first CMAF chunk in a CMAF segment. The `SegmentTypeBox` is specified by ISO/IEC 14496-12 to be the first box of an ISO BMFF segment and can be ignored in other locations.

NOTE 4 A `SegmentTypeBox` prepended to CMAF chunks or fragments at the time of encoding can be useful as a media object delimiter and identifier. It allows indexing and packaging operations to rapidly locate and identify CMAF chunks and fragments within a CMAF track without deep inspection of elementary streams or other clues to identify CMAF fragment boundaries and preceding boxes intended to be part of each media object.

7.3.2.4 CMAF fragment

A CMAF fragment as defined in [3.1.1](#) contains one or more CMAF chunks, therefore CMAF chunks may be present in CMAF addressable media objects, including CMAF segments and CMAF track files, and a CMAF chunk may be referenced as a CMAF addressable media object (see [subclause 7.3.2.2](#)).

Table 8 — CMAF fragment structure

NL 0	Format Req.	Specification	CMAF constraints	Description
styp	0/1	ISO/IEC 14496-12	CMAF 7.2	Segment type Signalling compatibility to CMAF fragment
prft	0/1	ISO/IEC 14496-12		Producer reference time
emsg	*	ISO/IEC 23009-1	CMAF 7.4.5	Event message
CMAF Chunk	1+	CMAF 7.3.2.3		CMAF chunk

Each CMAF fragment conforms to the following constraints:

- A CMAF fragment shall contain the boxes in [Table 8](#), with the conditions and optionality indicated.
- Each CMAF fragment, in combination with its associated CMAF header and optional decryption key(s), shall contain sufficient metadata to enable the CMAF fragment to be decoded, decrypted, and displayed when it is independently accessed.

NOTE 1 For example, if sample groups and sample group descriptions are used to signal encryption key changes, then a `SampleGroupDescriptionBox` and `SampleToGroupBox` probably need to be present in the `TrackFragmentBox` to make the CMAF fragment randomly accessible and decryptable.

- The first CMAF chunk in a CMAF fragment may be preceded by other boxes as long as the CMAF fragment remains conformant, including one `SegmentTypeBox` and/or one or more `ProducerReferenceTimeBox(es)` and/or `DASHEventMessageBox(es)`. (See [subclause 7.4.5](#) and [Annex E](#) for more information on Event Messages).
- A CMAF fragment with a `SegmentTypeBox` preceding the first `MovieFragmentBox` should include the CMAF fragment brand 'cmff' and may include any `compatible_brands` listed in the `FileTypeBox` of the CMAF track's CMAF header.
- A CMAF fragment can be referenced as a CMAF segment and may have a `SegmentTypeBox` prepended that includes the CMAF segment brand 'cmfs' to identify the start of the CMAF segment.

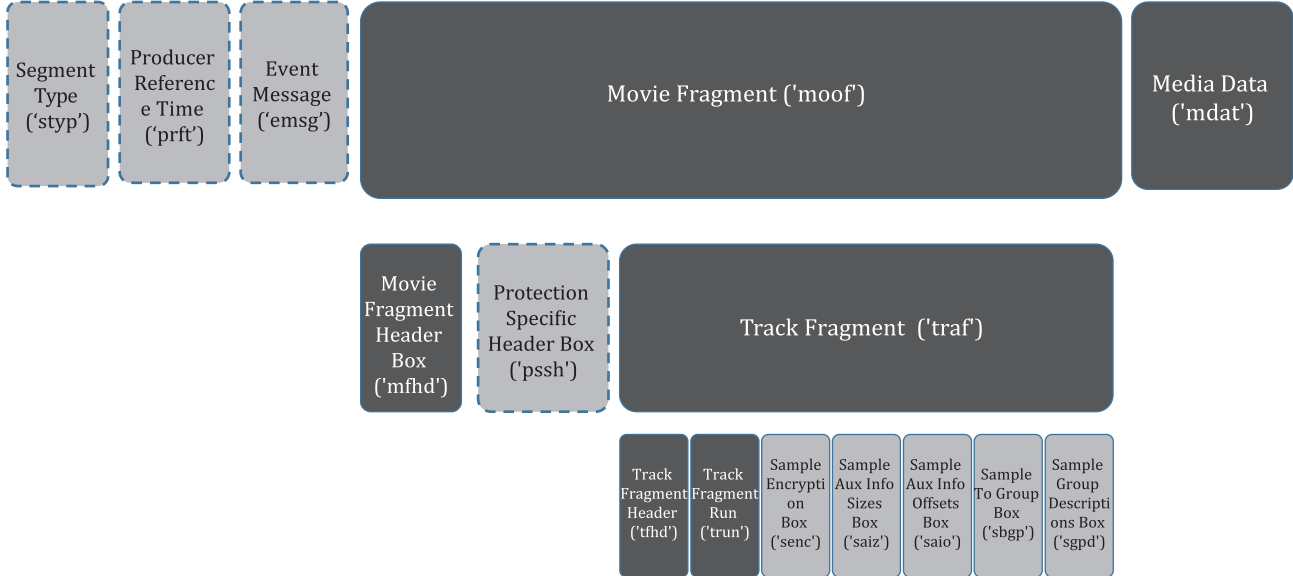
NOTE 2 It is possible that the first CMAF chunk in the first CMAF fragment in a CMAF segment would be preceded by a `SegmentTypeBox` that includes the `compatible_brands` 'cmfl', 'cmff', and 'cmfs' to indicate that it is the start of a CMAF segment, fragment, and chunk. The start of a CMAF fragment that contains multiple CMAF chunks can be identified by including both the 'cmff' and 'cmfl' brands, whereas subsequent CMAF chunks in the CMAF fragment would only include the 'cmfl' brand.

NOTE 3 As noted in [subclause 7.3.2.3](#), only a `SegmentTypeBox` at the start of a CMAF addressable media object is expected to be processed by players, whereas a `SegmentTypeBox` located within a CMAF addressable media object is expected to be ignored by players.

- f) Each CMAF fragment in a CMAF track should have a duration of at least one second, with the possible exception of the first and last CMAF fragments of the CMAF track.

NOTE 4 Additional constraints specific to video, audio, and subtitle media types are specified in [Clauses 9, 10, Annex A](#) and [Annex B](#).

[Figure 14](#) illustrates the box sequence (left to right) and containment (top to bottom) of a CMAF fragment.



Lower boxes indicate containment in the box above. The sequence of boxes contained in the `TrackFragmentBox` ('traf') is only recommended. Light shaded boxes are always optional. Light boxes in the bottom row are required or conditionally required when encryption is used. This example contains one CMAF chunk ('moof', 'mdat') in the CMAF fragment.

Figure 14 — CMAF fragment box sequence and containment

7.3.2.5 CMAF random access chunk

CMAF chunks are defined in [subclause 7.3.2.3](#). If chunks fulfil certain requirements, then in combination with the CMAF header, the CMAF track can be randomly accessed at CMAF Random Access Chunks, i.e. the chunk carries sufficient information to access the media in the stream, e.g. all necessary encryption in combination with the CMAF Header.

Each CMAF Random Access Chunk shall conform to the following constraints:

- A CMAF Random Access Chunk shall conform to a CMAF chunk as defined in [subclause 7.3.2.3](#).
- The chunk shall contain a position I_{SAP} such that the SAP type for this position is of type 1, 2, or 3 according to 14496-12.
- A CMAF Random Access chunk may include a `SegmentTypeBox` preceding the `MovieFragmentBox` and may be referenced as a CMAF addressable media object. The `SegmentTypeBox` should include the CMAF chunk brand 'cmfr' and may include any valid `compatible_brands` listed in the `FileTypeBox` of the CMAF track's CMAF header.

Note that any CMAF Fragment conforms to a CMAF Random Access Chunk.

7.3.3 CMAF addressable media objects

7.3.3.1 CMAF segment

Table 9 — CMAF segment structure

NL 0	Format Req.	Specification	CMAF constraints	Description
styp	0/1	ISO/IEC 14496-12		Segment type Signalling compatibility to CMAF segment
prft	0/1	ISO/IEC 14496-12		Producer reference time
emsg	*	ISO/IEC 23009-1	CMAF 7.4.5	Event message
CMAF Fragment	1+	CMAF 7.3.2.4		CMAF fragment

Each CMAF segment as defined in 3.1.5 conforms to the following constraints.

- A CMAF segment shall include the boxes in Table 9, with the conditions and optionality listed.
- A CMAF segment shall contain one or more complete and consecutive CMAF fragments in decode order.
- A CMAF segment may include a `ProducerReferenceTimeBox` and/or `DASHEventMessageBox(es)` preceding the first `MovieFragmentBox`.
- A CMAF segment may include a `SegmentTypeBox` preceding all other boxes in the CMAF segment. The `SegmentTypeBox` should include the CMAF segment brand 'cmfs' and may include any `compatible_brands` listed in the `FileTypeBox` of the CMAF track's CMAF header.

NOTE Preceding boxes can be prepended to CMAF chunks or CMAF fragments at time of encoding or prepended to CMAF segments at time of delivery. It is expected that players will process the boxes prepended to a CMAF addressable media object, but they will ignore these boxes when located within a CMAF addressable media object, i.e., not prepended to the first CMAF chunk in a CMAF segment. The `SegmentTypeBox` is specified by ISO/IEC 14496-12 to be the first box of an ISO BMFF segment and can be ignored in other locations.

7.3.3.2 CMAF chunk

A CMAF chunk is specified as a data structure within a CMAF fragment in subclause 7.3.2.3, but a CMAF chunk may also be referenced as a CMAF addressable media object, e.g. for low latency live streaming.

There is no direct restriction on the SAP type of the first media sample in a CMAF chunk, except by inheritance of CMAF fragment media sample constraints specified by a CMAF media profile. Because video CMAF media profiles typically constrain the initial media sample type to be SAP 1 or 2, the first CMAF chunk in a CMAF fragment inherits that media sample constraint.

Adaptive switching in players is normally limited to the earliest CMAF chunk in a CMAF fragment. Prepending a `SegmentTypeBox` to each CMAF chunk with appropriate brands can indicate the start of a CMAF chunk that is also the start of a CMAF fragment.

A `SegmentTypeBox`, `ProducerReferenceTimeBox`, and/or `DASHEventMessageBox(es)` may be prepended to a CMAF chunk at time of encoding or time of delivery. In a CMAF addressable media object, only when these boxes precede the first `MovieFragmentBox` are they expected to be processed by a player.

7.3.3.3 CMAF track file

A CMAF track as defined in 3.1.6 is a logical collection of CMAF fragments and a CMAF header that might never be stored or stored in that form. In contrast, a CMAF track file is a single file that contains one CMAF track, stored in sequence, and is therefore addressable as a file. Since CMAF chunks and CMAF fragments can be preceded by a `SegmentTypeBox`, CMAF media objects may be identified by `SegmentTypeBox` within a

CMAF track. CMAF media objects may be indexed in a `SegmentIndexBox` to enable addressing CMAF media objects in a CMAF track file by byte range.

Table 10 — CMAF track file structure

NL 0	Format Req.	Specification	CMAF constraints	Description
CMAF Header	1	CMAF 7.3.2.1	CMAF 7.2	CMAF header
sidx	0/1	ISO/IEC 14496-12		Segment index
CMAF Fragment	1+	CMAF 7.3.2.4	CMAF 7.3.2.4	CMAF fragments

Each CMAF track file conforms to the following constraints:

- A CMAF track file shall include the boxes in [Table 10](#), with the conditions and optionality listed.
- Additional boxes, such as `SegmentIndexBox(es)`, may be present between the CMAF header and the first CMAF fragment.
- If `SegmentIndexBoxes` exist, each subsegment referenced in the `SegmentIndexBox` shall be a single CMAF fragment contained in the CMAF track file.
- A video CMAF track may use an offset edit list as specified in [subclauses 7.5.13](#) and [10.2.6](#).
- A video CMAF track may use `v1 TrackRunBoxes` using negative composition offsets to adjust the composition time of the earliest presented video media sample in each CMAF fragment to its `baseMediaDecodeTime` and the earliest video media sample in the CMAF track file to zero, without using an offset edit list.

7.3.4 CMAF switching sets

7.3.4.1 General constraints on a CMAF switching set

These general CMAF switching set constraints, plus specific constraints defined by a CMAF media profile that the CMAF switching set conforms to, are intended to enable seamless adaptive switching by typical players.

Each CMAF switching set as defined in [3.2.2](#) conforms to the following constraints.

- A CMAF switching set shall contain one or more CMAF tracks, each of which is a synchronized encoding of the same source content.
- A CMAF switching set shall contain CMAF tracks of only one media type, i.e., audio or video or subtitles.
- All CMAF tracks in a CMAF switching set shall have the same duration (computed from the media sample durations, optionally stored in a `MovieExtendsHeaderBox`).
- All CMAF tracks in a CMAF switching set shall contain the same number of CMAF fragments.
- For any CMAF fragment in one CMAF track in a CMAF switching set, there shall be a CMAF fragment with the same decode time in all other CMAF tracks.
- All CMAF tracks in a CMAF switching set shall have the same value of `baseMediaDecodeTime` in the first CMAF fragment's `TrackFragmentBaseMediaDecodeTimeBox`, measured from the same timeline origin.
- The presentation time of the earliest media sample of the earliest CMAF fragment in each CMAF track shall be equal.
- Alternative CMAF fragments in a CMAF switching set with the same `baseMediaDecodeTime` shall contain alternative encodings of the same source media samples.

NOTE 1 These constraints do not prohibit different frame rates in different CMAF tracks in a CMAF switching set. However, these constraints restrict the combinable frame rates to those that result in identical CMAF fragment durations for alternative CMAF fragments.

- i) All CMAF tracks in a CMAF switching set shall conform to one CMAF media profile. However, CMAF tracks included in a CMAF switching set may contain different CMAF media profile brands that all conform to one CMAF media profile. See [subclause 12.1](#) for additional CMAF switching set constraints specified by CMAF media profiles.

NOTE 2 CMAF media profiles specify the maximum limits for encoding parameters, so subset or lower media profiles that can be decoded by a higher media profile decoder conform to the higher media profile.

- j) CMAF header parameters shall not differ between CMAF tracks, except as allowed in [Table 11](#). CMAF media profiles may specify additional constraints, as noted in the table.
- k) Additional CMAF header boxes specified in [Table 3](#) and [Table 4](#) that are not listed in [Table 11](#) may contain different boxes and information.

Table 11 — General constraints on CMAF header boxes in CMAF switching sets

Box	General CMAF header constraints in a CMAF switching set
ftyp	<i>Shall be identical except for media profile brands [see i) in subclause 7.3.4.1</i>
mvhd	<i>Shall be identical except for creation_time and modification_time</i>
tkhd	<i>Shall be identical except for width, height, creation_time, and modification_time. See NOTE 1.</i>
trex	<i>Identical</i>
elst	<i>Shall be identical except for video CMAF track files with a different composition offset</i>
mdhd	<i>Shall be identical except for creation_time and modification_time</i>
mehd	<i>Identical</i>
meta	<i>May contain different boxes and data</i>
udta	<i>May contain different boxes and data</i>
cprt	<i>Identical</i>
kind	<i>Identical</i>
elng	<i>Identical</i>
hdlr	<i>Identical</i>
vmhd	<i>Identical</i>
smhd	<i>Identical</i>
sthd	<i>Identical</i>
dref	<i>Identical</i>
stsd	<i>Sample entries shall have the same codingname (four-character code) and conform to other CMAF track format and media profile specified constraints. See NOTE 2.</i>
sgpd	<i>May differ</i>
pssh	<i>Identical</i>
sinf	<i>Identical</i>
schi	<i>Identical</i>
schm	<i>Identical</i>
frma	<i>Identical</i>
tenc	<i>Shall be identical except for initialization vector values</i>

NOTE 1 Track width and height can differ, but picture aspect ratio is the same for all CMAF tracks.

NOTE 2 Sample entry constraints for CMAF switching sets are defined by each CMAF media profile, its specified or referenced ISO BMFF track format, and its CMAF switching set constraints. A sample description can contain multiple sample entries.

There are two functional types of CMAF switching set initialization constraints:

- 1) Multiple initialization — CMAF switching sets conforming to the general constraints above may rely on decoding parameters stored in each CMAF track's header to decode a CMAF fragment from that CMAF track. This implies that a player needs to access and process the CMAF header for each CMAF track it switches to.
- 2) Single initialization — If CMAF headers are constrained to be equivalent within a CMAF switching set, or CMAF tracks only change parameters that are stored in each CMAF fragment, then a player only needs to access one CMAF header and each CMAF fragment to find all the necessary decoding and rendering parameters. A player only needs to initialize the CMAF switching set once, then it can continue decoding CMAF fragments from any CMAF track in the CMAF switching set.

The two types of CMAF switching set constraints and processing are illustrated in [Figure 10](#).

7.3.4.2 CMAF switching set single initialization constraints

Encoding of CMAF tracks in a CMAF switching set can be further constrained to allow one-time initialization of the CMAF switching set using a single CMAF header (see [Figure 10](#)). Each CMAF media profile or the CMAF track format it references can specify single initialization CMAF switching set constraints.

In general, single initialization constraints can be specified by:

- a “common CMAF header” in all CMAF tracks containing all decoding, decryption, and rendering parameters required by any CMAF fragment in the CMAF switching set;
- “inband parameter sets” stored and referenced in each CMAF fragment, where CMAF headers are only used to initialize decoding and display with the appropriate codec, profile, level, resolution, etc. This is typically the case for audio and can be specified for inband video parameter media sample formats such as 'avc3' sample entry;
- a combination of both CMAF header and CMAF fragment stored parameters, but with the constraint that the combination of any CMAF header in the CMAF switching set and CMAF fragment will contain the necessary decoding parameters.

NOTE Enabling changes in inband parameters over time is useful for live streaming, where content from different encodings can be sequenced in a CMAF presentation, such as programs and ads in a live “channel”, or when a live video contribution feed is received containing inband parameters that change and are not known in advance when a CMAF header is packaged and delivered to a player.

7.3.4.3 CMAF switching set single initialization constraints identifier

The identifier "urn:mpeg:cmf:ss" may be used to signal CMAF switching set conformance to single initialization constraints.

7.3.4.4 Aligned CMAF switching set constraints

Aligned CMAF switching sets are defined in [3.2.3](#) to be a “set of CMAF switching sets, the CMAF tracks of which all contain alternative encodings of the same source content in time-aligned CMAF fragments, but all CMAF tracks do not conform to a single CMAF switching set”.

Aligned CMAF switching sets conform to the following constraints:

- a) An aligned CMAF switching set shall contain two or more CMAF switching sets.
- b) Aligned CMAF switching sets shall contain CMAF switching sets that are encoded from the same media stream.
- c) Aligned CMAF switching sets shall contain CMAF switching sets of equal duration.
- d) Aligned CMAF switching sets shall contain the same number of CMAF fragments in every CMAF track.

- e) Aligned CMAF switching sets shall contain CMAF fragments in every CMAF track with matching `baseMediaDecodeTime` and duration, for each CMAF fragment in a CMAF track.

NOTE Some CMAF players can seamlessly switch between CMAF tracks in different CMAF switching sets that are aligned, such as CMAF switching sets with different encryption keys and their associated DRM license(s).

7.3.4.5 CMAF principal header

A CMAF Principal Header for a CMAF Switching Set conforms to the following constraints:

- a) It shall conform to a CMAF Header.
- b) The CMAF Header and the first sample entry shall contain parameters that are sufficient to enable a single initialization of media decoding, decryption and rendering systems to decode, decrypt and render all CMAF fragments in all CMAF Tracks of this CMAF Switching Set.
- c) If the CMAF Principal header requires sample entries with a decoder configuration record, as specified by a CMAF media profile, with decoding and rendering parameter limits, such as codec profile, level, image height, width, etc., then the first sample entry shall contain parameters equal or greater than the maximum values contained in all CMAF fragments in all CMAF track of the Switching Set to enable an initialization of media decoding and rendering systems to render all CMAF fragments in all CMAF tracks of the CMAF Switching.
- d) A media profile may define additional requirements for a CMAF Principal Header.

7.3.5 CMAF selection sets

CMAF selection sets as defined in [3.2.4](#) conform to the following constraints.

- a) A CMAF selection set shall contain one or more CMAF switching sets.
- b) All CMAF switching sets within a CMAF selection set shall be of the same media type, e.g. audio, video, or subtitles.
- c) All switching sets within a CMAF selection set shall be of the same duration, within a tolerance of the longest CMAF fragment duration of any CMAF track in the selection set.
- d) Different CMAF switching sets that encode the same content with different codecs or video formats may be contained within a CMAF selection set to enable a player to select the most compatible codec or video format and the CMAF switching set that contains it.
- e) A CMAF selection set may contain multiple aligned CMAF switching sets, which encode the same content in different CMAF switching sets that have time-aligned CMAF fragments, but with different keys, codecs, etc.
- f) One CMAF track should be presented from each selection set in a presentation, e.g. one audio, one video, and one subtitle selection set.

7.3.6 CMAF presentations

A CMAF presentation as defined in [3.2.5](#) conforms to the following constraints.

- a) All CMAF tracks in a CMAF presentation shall have the same timeline origin.
- b) Different CMAF switching sets may use different CMAF track `timescale` values. In that case, `baseMediaDecodeTime` integer values will be different even if the corresponding ISO BMFF samples would have the same presentation time.
- c) All CMAF tracks in a CMAF presentation containing a video switching set shall be start aligned with CMAF presentation time zero equal to the earliest video media sample presentation start time in the earliest CMAF fragment (see [subclause 6.6.8](#)).