INTERNATIONAL STANDARD

ISO/IEC 23000-19

First edition
2018-01

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

Reference number
ISO/IEC 23000-19:2018(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Published in Switzerland

Contents

Page

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

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 ISO documents 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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

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

For an explanation on 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 the following URL: www.iso.org/iso/foreword.html.

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

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

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.

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

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

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

IETF RFC 6381, *The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types*

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.2020[2)], *Parameter values for ultra-high definition television systems for production and international programme exchange*

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

ITU-R Recommendation BT.2100-0:2016[3)], *Image parameter values for high dynamic range television for use in production and international programme exchange*

ITU-T Recommendation X.667:2014[4)], *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*

ANSI/CTA-608-E R-2014[5)], *Line 21 Data Services*

ANSII/CTA-708-E[6)], *Digital Television (DTV) Closed Captioning*

W3C[7)], *TTML Profiles for Internet Media Subtitles and Captions 1.0 (W3C IMSC1)*

W3C[8)], *TTML Media Type Definition and Profile Registry, W3C Working Group Note (W3C TTML Registry)*

3 Terms and definitions

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

---

1) Available at https://tools.ietf.org/html/rfc5234

2) Available at http://www.itu.int/rec/R-REC-BT.2020/en

3) Available at https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2100-0-201607-I!!PDF-E.pdf

4) Available at https://www.itu.int/rec/T-REC-X.667

5) Available at http://www.techstreet.com/standards/cta-608-e-r2014?product_id=1815447

6) Available at http://www.techstreet.com/standards/cta-708-e?product_id=1860354

7) Available at http://www.w3.org/TR/ttml-imsc1

8) Available at https://www.w3.org/TR/ttml-profile-registry

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

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1 Media objects

3.1.1
CMAF fragment
encoded ISOBMFF media segment conforming to CMAF constraints

3.1.2
CMAF header
sequence of CMAF constrained ISOBMFF 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 "ISOBMFF segment" and a "DASH segment".

3.1.6
CMAF track file
one *CMAF track* (3.2.1) stored consecutively in a single ISOBMFF file with the earliest *CMAF fragment* (3.1.1) constrained to start at decode time zero

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 the `'cmfc'` brand, including 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 compatibility 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.

3.3 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, numbered as in ISO/IEC 14496-12:2015, Annex I

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 ISOBMFF is always identified by the term “media sample”. The word “sample” is frequently used in ISOBMFF 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

The following abbreviated terms are used in this document.

AU	Access Unit
CDN	Content Delivery Network
CMAF	Common Media Application Format
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
ISOBMFF	ISO Base Media File Format, defined in ISO/IEC 14496-12
KID	Key Identifier as defined in ISO/IEC 23001-7
NAL	Network Adaptation Layer
PCM	Pulse Code Modulation
PPS	Picture Parameter Set
SAP	Stream Access Point as defined in ISO/IEC 14496-12
SEI	Supplemental Enhancement Information
SPS	Sequence Parameter Set
VCL	Video Coding Layer
VPS	Video Parameter Set
VUI	Video Usability Information

5 Document organization

First-time readers of this document are advised to start with Clause 6 for 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 normative specifications in Clause 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. Clauses 8 through 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.

The following is a list of the main clauses of this document, with a brief description of each.

Clause 6 describes the segmented media encoding and playback model using the media objects defined by the CMAF.

Clause 7 describes the use of ISO Base Media File Format for the Common Media Application Format brand.

Clause 8 describes how digital rights management information and encryption is applied to the Common Media Application Format.

Clause 9 describes the general video track format, constraints for NAL structured video tracks, and the AVC video track format.

Clause 10 describes the general audio track format and specifies two AAC audio CMAF media profiles.

Clause 11 describes the subtitle track format, CMAF media profiles for WebVTT and IMSC1 TTML subtitles, and signalling of CTA 608/708 captions embedded in video streams.

Clause 12 describes the general requirements for CMAF media profiles and CMAF presentation profiles.

Annex A describes several CMAF media profiles, their compatibility brands, and a CMAF presentation profile that conditionally requires some of those media profiles.

Annex B describes packaging and codec constraints for some CMAF media profiles using the HEVC video codec.

Annex C describes framing and encoding CMAF switching sets using subsampling and scaling of video to provide seamless playback with adaptive bit rate and scaling.

Annex D describes examples of player track selection, synchronization, and adaptive switching of a CMAF presentation.

Annex E describes the use of event messages attached to media objects to deliver metadata.

Annex F describes maintaining presentation timing and delivery in the event of missing media samples and resources.

Annex G describes encoding recommendations for AAC audio CMAF tracks conforming to adaptive CMAF switching sets.

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