**** **ISO/IEC JTC 1/SC 29/WG 3 N0** **722**

**ISO/IEC JTC 1/SC 29/WG 3**

**MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title:** **WD ISO/IEC 23009-9** **Redundant encoding and packaging for segmented live media**

**Status:** Approved

**Date of document:** 2022-11-15

**Source:** ISO/IEC JTC 1/SC 29/WG 3

**No. of pages:** 37 (with cover page)

**Email of Convenor:** young.L@samsung.com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3>

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 3**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/WG 3 N** **00** **722**

**Mainz DE –October 2022**

**ISO 23009-9:2022**

ISO JTC 1/SC 29/WG 3

Date: 2022-11-15

Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 9: Redundant encoding and packaging for segmented live media (REaP)

WD stage

**Warning for WDs and CDs**

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

© ISO 2022

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

Contents

[Foreword ix](#_Toc119499917)

[Introduction x](#_Toc119499918)

[1 Scope 1](#_Toc119499920)

[2 Normative references 1](#_Toc119499921)

[3 Terms and definitions 1](#_Toc119499922)

[3.1 Terms 1](#_Toc119499923)

[3.2 Abbreviated terms 2](#_Toc119499930)

[4 Document organization 2](#_Toc119499931)

[5 Reference workflow for Redundant encoding and packaging (REaP) 3](#_Toc119499932)

[6 Formats for Redundant distribution Encoding 4](#_Toc119499933)

[6.1 General workflow and operation (informative) 4](#_Toc119499934)

[6.2 Constraints on the Ingest Media Presentation Description (I-MPD) 5](#_Toc119499935)

[6.3 Constraints on the redundant encoder track and segment format 6](#_Toc119499936)

[6.4 Format for exchanging information between distribution encoders (TuC) 7](#_Toc119499937)

[7 Formats for Redundant media presentation packaging 7](#_Toc119499938)

[7.1 General requirements on redundant packager output format 7](#_Toc119499939)

[7.2 General requirements on redundant packager input 8](#_Toc119499940)

[7.3 Specific requirements on redundant packager D-MPD output format 8](#_Toc119499941)

[7.4 Specific requirements on redundant packager HLS output format 9](#_Toc119499942)

[8 Asset storage and recording Format 9](#_Toc119499943)

[8.1 Track format for storage of live archives 10](#_Toc119499944)

[8.2 Storage Media Presentation Description (S-MPD) 12](#_Toc119499945)

[8.2.1 Overview 12](#_Toc119499946)

[8.2.2 Constraints on the S-MPD for 24x7 live archiving and recording 13](#_Toc119499947)

[Annex A Example Media Presentation Description 14](#_Toc119499948)

[Annex B Recommended Configurations 20](#_Toc119499949)

[B.1 Recommended segment durations for encoder synchronization 20](#_Toc119499950)

[Annex C Example Applications 21](#_Toc119499951)

[C.1 Redundant and high availability distribution: 21](#_Toc119499952)

[C.2 Distributed encoding: 21](#_Toc119499953)

[C.3 A/B watermarking: 21](#_Toc119499954)

[C.4 Metadata Sources: 21](#_Toc119499955)

[C.5 Live Archives 21](#_Toc119499956)

[Annex D Example methods of carriage of Source timing 22](#_Toc119499957)

[Annex E Example method for redundant encoder tracks generation 23](#_Toc119499958)

[Annex F Example method for generating a recording S-MPD from an I MPD 25](#_Toc119499959)

[Bibliography 27](#_Toc119499960)

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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](https://www.iso.org/directives-and-policies.html)).

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](https://www.iso.org/iso-standards-and-patents.html)).

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](https://www.iso.org/foreword-supplementary-information.html).

This document was prepared by Technical Committee ISO/TC *[or ISO/PC]* JTC1, *[name of committee]*, Subcommittee SC 29.

A list of all parts in the ISO 23009 series can be found on the ISO website.

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](https://www.iso.org/members.html).

Introduction

This standard specifies a reference architecture for the creation of segmented live media based on redundant packaging and encoding using common timing information. It also specifies media formats and constraints such that conformant redundant packagers and encoders produce interchangeable live media segments.

This standard also specifies formats for 24x7 live recording and archiving of segmented live media.

The International Organization for Standardization (ISO) draws attention to the fact that it is claimed that compliance with this document may involve the use of a patent.

ISO takes no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has assured ISO that he/she is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with ISO. Information may be obtained from the patent database available at [www.iso.org/patents](http://www.iso.org/patents).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those in the patent database. ISO shall not be held responsible for identifying any or all such patent rights.

# Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 9: Redundant encoding and packaging for segmented live media (REAP)

# Scope

This document defines media formats for redundant encoding and packaging of live segmented media.

Related and derived formats for media ingest and asset storage are also defined.

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

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

ISO/IEC 23000-19, *Information technology — Multimedia application format (MPEG-A) — Part 19: Common media application format (CMAF) for segmented media*

# Terms and definitions

For the purposes of this document, the terms and definitions given in [1], [2], [3] and the following 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/>

## Terms

### Media Presentation

A CMAF presentation or a DASH Media Presentation description or an HTTP live streaming master playlist.

### Encoder

Entity or computational unit used to encode or transcode an input to one or more ISO BMFF tracks.

### Packager

Entity or computational unit used to package an input to an ISO BMFF track format and/or media streaming format such as DASH or HLS.

### Origin

Entity or computational unit used to serve media content based on HTTP requests.

### Unix Epoch

Seconds elapsed since 00:00:00 UTC on January 1st 1970 (excluding leap seconds).

### Ceil()

Nearest integer larger than or equal to the input number.

## Abbreviated terms

The following abbreviated terms are used in this document

CMAF Common Media Application Format as defined in ISO/IEC 23000-19

DASH MPEG-DASH as defined in ISO/IEC 23009-1

DASH-IF DASH Industry Forum

HLS HTTP Live Streaming as defined in IETF RFC 8216

ID Identifier

IETF Internet Engineering Task Force

I-MPD Ingest Media Presentation Description

MPD Media Presentation Description

S-MPD Storage Media Presentation Description

UUID Universally Unique Identifier

# Document organization

This specification is organized as follows: Clause 1 to 4 provide the introduction, scope, references, definitions, functions and abbreviated terms.

Clause 5 describes a reference workflow/architecture for redundant encoding and packaging of live segmented media.

Clause 6 specifies media formats for redundant encoder synchronization.

Clause 7 specifies media formats used for redundant packager synchronization.

Clause 8 specifies the track format and MPD constraints for asset storage.

Annex A provides example media presentation descriptions for the ingest and storage applications.

Annex B provides informative recommendations for media segment durations.

Annex C provides example applications and use cases.

Annex D provides ways a contribution signal may carry timing information.

Annex E provides an example method for redundant encoder tracks generation.

Annex F provides an example method for generating recording S-MPD from I-MPD.

# Reference workflow for Redundant encoding and packaging (REaP)

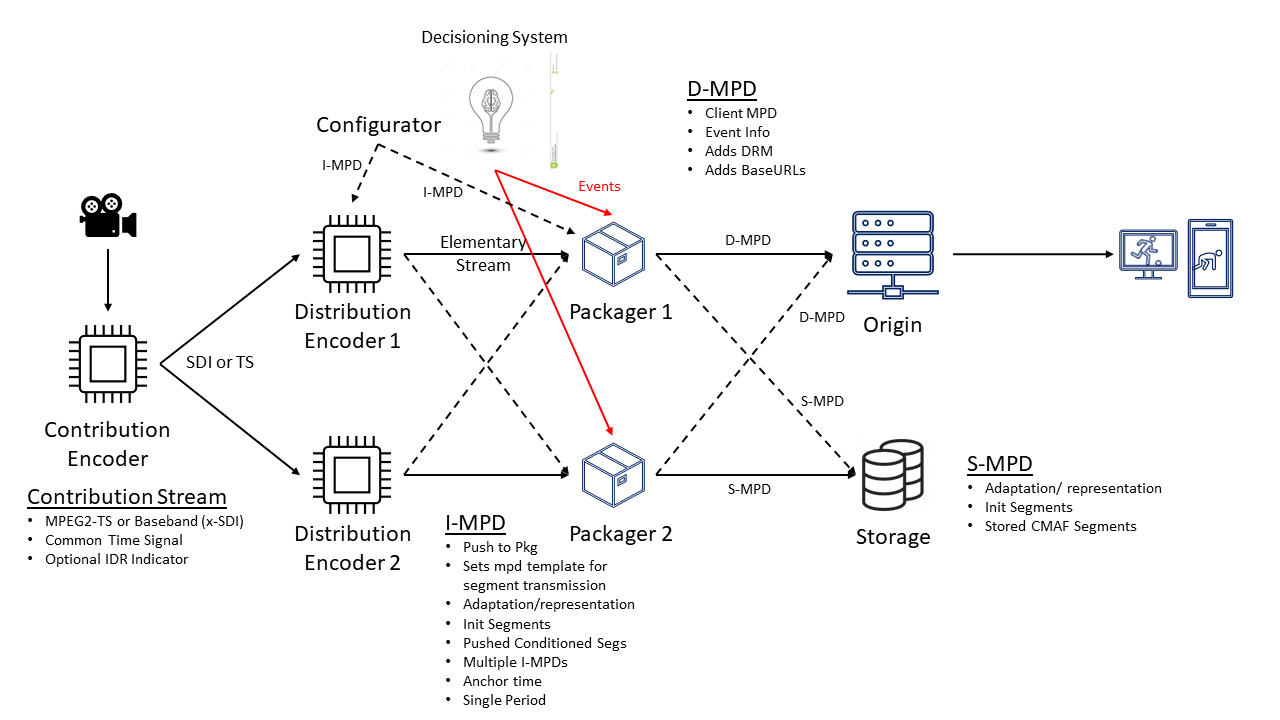


Figure 1 Reference workflow for redundant encoding and packaging of live segmented media

Figure 1 illustrates the reference workflow for redundant encoding and packaging (REaP). The assumption is that a contribution encoder or source transmits a source signal to two (or more) distribution encoders generating live segmented media. The two (or more) distribution encoders transmit live segmented media to two (or more) packagers that generate media presentations and transmit those through one or more origins to players or devices. In addition, streams are recorded or stored in a storage unit.

Figure 1 details the assumptions and intended use of the defined media formats in the rest of this document as follows.

The definition of formats produced by the contribution encoder are out of scope of this document, however, it is assumed that such formats contain a common time signal. Example formats from the contribution encoder and relevant common timing formats are given in annex D.

The redundant distribution encoders 1 and 2 generate multiple representations of segmented media. Clause 6 defines the ingest media presentation description (I-MPD) for announcing these representations. In addition, it defines the segment format constraints for the transmission of live segmented media.

The packagers 1 and 2 are responsible for generating the Delivery Media presentation Description (D-MPD) for delivery to 1 or more clients through 1 or more origins. Constraints on the D-MPD based and live media segments are defined in clause 7. Decisioning systems may be present to insert timed events by a packager or encoder. The interface to the decisioning system is out of scope for this document, but for this reference architecture to be valid it is expected that redundant packagers/encoders will receive functionally equivalent input from the decisioning system.

The functional block for storage is used to record and store media. Clause 8 defines the S-MPD (Storage MPD) to support recording and storage of redundant live segmented media.

Annex C describes some of the intended use cases.

# Formats for Redundant distribution Encoding

## General workflow and operation (informative)



Figure 2 An example workflow for redundant distribution encoder and packager operation.

Figure 2 illustrates an example of using redundant distribution encoders that follows the general reference workflow of figure 1, combining the origin and packager functional entities. The example workflow shows the output of the (distribution) encoders: media segments carrying the same label S1, S2 that are interchangeable. The segments are transmitted to packagers/origins and the output of these entities in Figure 2 is a media stream such as based on DASH or HLS. To achieve redundancy, the segment duration D may be set as part of the encoder configuration and the STS to compute the offset relative to epoch of input timestamps. Encoders that are intended to have synchronized output have the same values set for these parameters and interchangeable segments have the same duration and the same earliest presentation time relative to Unix Epoch.

Specific protocol details of the transmission from encoder to packager/origin are out of scope of this document. The general workflow operation process is assumed to comprise the following steps:

1. The receiver packager sets up a URL endpoint that the encoder can issue requests to.
2. The encoder authenticates, connects, and sends an ingest media presentation description (I-MPD) and/or initialization segments for each segmented live media representation. The I-MPD defines the grouping of Representations using AdaptationSets and the segment URL naming via SegmentTemplate elements.
3. The receiver reads the I-MPD. When the CMAF profile for DASH is used, or when content is formatted according to the CMAF track format, representations map to CMAF tracks, and the adaptation sets map to CMAF switching sets.
4. The sender sends each media segment using a separate request. The URL is derived from URL endpoint combined with the segment name derived from the SegmentTemplate Element.
5. The receiver appends the media segments to the corresponding initialization segment and stores these in memory.
6. Each representation/track uses a separate connection, e.g. a separate TCP connection.
7. Video, audio, text/subtitle and metadata tracks are all supported.
8. Retransmission of media segments in a track/representation may happen in case of failures. In such cases it is also common to retransmit the initialization segment.
9. The media segment's earliest presentation time is used as a “key” of segments in a track, the track name identified by Representation@id is used as a key for the track. Using these keys, interchangeable segments for different tracks from different senders can be identified by the receiver.

## Constraints on the Ingest Media Presentation Description (I-MPD)

The I-MPD is used to establish the naming convention of the segment URLs and grouping of tracks in adaptation sets. In case the content is conforming to CMAF, each representation in the I-MPD represents a CMAF track, each AdaptationSet in the I-MPD represents a CMAF switching set.

The I-MPD conforms to ISO/IEC 23009-1. Additional constraints on the I-MPD are the following:

1. The I-MPD should follow the CMAF profile and/or iso live profiles, the MPD@profiles should include *urn:mpeg:dash:profile:cmaf:2019* and the ISO live profile of DASH *urn:mpeg:dash:profile:isoff-live:2011*
2. The Period@availabilityStartTime should be set to “1970-01-01T00:00:00Z” and the Period @start should be set to PT0S or a semantically equivalent value.
3. In case an I-MPD is updated, identical naming conventions shall apply.
4. The I-MPD shall contain a single Period element.
5. The Period@BaseURL element shall be absent.
6. An EventStream element shall not be used to carry time varying timed metadata.
7. The I-MPD shall contain a SegmentTemplate element in each AdaptationSet element but none shall be present in a Representation Element.
8. The AdaptationSet@BaseURL element shall be absent.
9. In the case an AdaptationSet signals encrypted tracks, the ContentProtection element shall be present with the following constraints:
   * 1. The dashif:Laurl element shall be present in the ContentProtection element and contain a URL of a license server. The authentication server shall not be used.
     2. The URL of the license server shall start with https:// and should use mutually authenticated TLS for the purpose. In this case the value of dashif:Laurl.licenseType shall include the string “mtls”
     3. A response of the license server should be a valid CPIX document.
     4. In case both decryption and encryption are needed, the value of dashif:Laurl.licenseType used for decryption shall include the string “decrypt”.
10. The SegmentTemplate@initialization in the I-MPD shall contain a single substring $RepresentationID$ and the SegmentTemplate@media shall contain a single substring $RepresentationID$ and a single substring $Number$ or $Time$ (not both). A separator character (- or \_) shall be between $Representation$ and $Number$ or $Time$ if one or more Representation@id ends with a number.
11. SegmentTemplate@media shall be identical for SegmentTemplate element
12. SegmentTemplate@initialization shall be identical for each SegmentTemplate element
13. Each SegmentTemplate element should contain a SegmentTimeline element
    1. A SegmentTimeline element may be empty.

## Constraints on the redundant encoder track and segment format

Entities conforming to this specification and CMAF shall produce media segments in conformance with ISO/IEC 23000-19 (CMAF) clause 7 (CMAF Track). Entities conforming to this specification and DASH shall produce media segments as described in ISO/IEC 23009-1 (DASH).

In the case of CMAF, conforming entities apply the track synchronization model specified in clause 6 as indicated in ISO/IEC 23000-19. Conforming encoders shall use the Unix epoch as the reference, therefore media presentation times shall be relative to 00:00:00 on January 1st 1970 excluding leap seconds.

Annex E provides an example method for redundant encoder track generation.

Encoder entities conforming to this specification that produce CMAF or DASH media segments shall generate boxes of the type MovieFragmentBox (moof) and optionally of the type SegmentTypeBox with the following constraints:

1. Media segments should be of a constant segment duration **D,** where the duration is the sum of all media sample presentation durations as indicated in the TrackFragmentRunBox.
2. In case of splicing, when a segment is created with duration **A** instead, the next media segment shall be of duration **2 x D - A** to keep the numbering and the number of segments since epoch. Alternatively, splicing can be achieved by inserting an IDR within the media segment and not creating a segment boundary. In this case the segment duration is D.
3. The SegmentType box may contain a ‘slat’ brand in case the input of the encoder was missing frames and the encoder filled the gap with filler frames. In case the SegmentTypeBox contains a ‘slat’ brand, that shall be interpreted to mean that one or more frames in the segment were replaced with filler content to avoid discontinuities.
4. In case the segment is the last segment, the SegmentTypeBox should contain the ‘lmsg’ brand. If the SegmentTypeBox contains the `lmsg` brand it is the last segment in the track.
5. The MovieFragmentHeaderBox shall contain the segment sequence number (**K**).
6. The TrackFragmentDecodeTime box shall contain a baseMediaDecodeTime that is equal to K x D x track\_timescale.
7. An edit list shall not be used when using epoch relative media presentation timing.
8. The 'roll' sample group may be used to indicate requirements for audio playout based pre- roll samples.

NOTE: Inserted/replaced periods that use period relative media presentation times (as opposed to epoch relative media presentation timing) may have an edit list inserted to meet requirements of the audio playout process.

1. Samples that overlap a leap second, may be adjusted to account for the leap second (reducing the duration of samples). Otherwise, media frames occurring during the leap second may be discarded by the encoder and not included in track.

## Format for exchanging information between distribution encoders (TuC)

At MPEG 140 it was proposed to define a format to exchange information between distribution encoders, the message should contain the following information:

* In-media timestamp for frame PTS(*T)*
* Offset from Epoch time corresponding to PTS(T), in units of microseconds
* Offset from Epoch time to the time the frame with the in-media timestamp of PTS(T) entered the encoder buffer, in units of microseconds.
* Input error notification
* Synchronization state (acquisition period, synchronized, not synchronized)
* Frame type for frame *T*
* Input URI
* Input state
* Image descriptors (e.g., color and edge histograms)

Details of this messaging will be developed and further input is welcome.

# Formats for Redundant media presentation packaging

This clause defines media formats and related constraints for input and output to redundant packagers.

## General requirements on redundant packager output format

The Packager entity is responsible for converting the I-MPD and live media segments to a valid delivery MPD (D-MPD) or media playlist based on RFC 8216. It is assumed that the workflow from Figure 1 or Figure 2 is used where 2 (or more) distribution encoders transmit to 2 (or more) packagers.

This clause defines recommended output formats to enable synchronized and interchangeable generation of media playlists from distributed packagers.

The following general requirements apply:

1. The D-MPD shall conform to ISO/IEC 23009-1.

2. The D-MPD should be generated according to the mapping described in the CMAF profile for DASH 23009-1 clause 8.12. Further constraints on generation of the D-MPD may be implementation dependent and application specific.

3. Media segment formatting should follow the constraints defined in clause 6.3.

NOTE: Inserted/replaced periods that use period relative media presentation times (as opposed to epoch relative media presentation timing) may have an edit list inserted to meet requirements for the audio playout process, and also may not comply with segment formats defined in clause 6.3.

4. Packagers shall, upon an HTTP request or response containing a media playlist or D-MPD in the body, set the Last-Modified HTTP Header: that has a syntax of: <day-name>, <day> <month> <year> <hour>:<minute>:<second> GMT to a time corresponding to the earliest presentation plus segment duration of the newest segment.

NOTE: This avoids some of the race conditions of MPD’s generated by different redundant packagers.

5. To enable debugging, the packagers should write ProducerReferenceTimeBoxes to indicate when a media segment was written to disk. This may result in more than one ProducerReferenceTimeBox per segment

6. If a segment that is expected to be available or received is not yet available to a playlist generator, it shall return an error response when a media playlist is requested.

7. In addition, a media playlist generator may include the ETag header to an HTTP response containing an MPD. In case the ETag header is included it shall be included for every playlist or segment requests.

## General requirements on redundant packager input

Additional requirements on media formats used as input to redundant packagers are the following:

1. As shown in Figure 1, encoders cross-transmit segments to the different available redundant packagers in the workflow. Each encoder 1……N transmits all output media segments to all packagers 1…..M. Interchangeable segments of each Representation shall have identical earliest media presentation time K x D x track\_timescale.

NOTE: In some practical cases cross transmission maybe costly, optimizations in the transmission protocol and delivery are encouraged, but these are out of the scope of this document.

2. Encoders shall transmit segments at roughly the same time configurable within bounds using a configurable fixed encoder delay **Dc**. Therefore, a segment with earliest presentation **T** shall be transmitted at T + segment\_duration + Dc by each of the encoders within a configurable bound (e.g. 500 ms). Therefore, redundant packagers should receive interchangeable input media segments within time limited bounded time differences.

NOTE: In practical systems this requires wall clock synchronization of encoders, that would usually be within 100 milliseconds bounds (+- 100 ms) in practical systems using NTP servers or other synchronization methods.

3. Each encoder should write a ProducerReferenceTimeBox for each segment, that shall contain the time a segment entered the encoder (flags field set to 0) and the time it was encoded (flags field set to 1). Encoders should write ProducerReferenceTimeBoxes reflecting the time a segment entered the encoder (flags field set to 0) and shall write the time it was encoded (flags field set to 1) in a ProducerReferenceTimeBox to a segment. This may result in segments with multiple ProducerReferenceTimeBoxes as input to the redundant packager

4. Splicing information and metadata shall be available at each of the packagers at least 8 seconds prior to required use.

5. Timed metadata shall be available to the packager during its active period. This implies that boxes of the type DASHEventMessageBox shall be repeated in segments as long as they are active. Otherwise, a separate segmented timed metadata track for carriage of event message metadata is used such as based on ISO/IEC 14496 clause 12.3 or ISO/IEC 23001-18. Each segment then contains all metadata events that overlap the media presentation internal of the segment. 6.

## Specific requirements on redundant packager D-MPD output format

Apart from clause 7.1 the following additional requirements for generating a D-MPD based on ISO/IEC 23009-1 apply:

1. The D-MPD shall conform to ISO/IEC 23009-1. The D-MPD should be generated according to the mapping described in the CMAF profile for DASH 23009-1 clause 8.12. Further constraints on generation of the D-MPD may be implementation dependent and application specific.

2. The media presentation description, when returned, shall set the MPD@publishTime to the time corresponding to the earliest presentation time plus segment duration of the newest segment.

3. If a segment that is supposed to be available or received and published in the MPD but it is not yet available at the packager, the packager shall return an error response when an MPD is requested.

4. Packagers Generators shall write the D-MPD with aligned representations. The segments in each representation and the durations of representations shall be aligned (within 100 milli seconds). The only exception to this is at the beginning or end of a live stream presentation. This implies that an MPD is only modified when updated segments for each representation are available to the packager.

NOTE: Redundant packagers using a shared storage is one way of enabling such consistent manifest generation between redundant packagers.

## Specific requirements on redundant packager HLS output format

The following requirements apply for the generation of media playlists based on RFC 8216.

1. The segment URL’s indicated under an #EXTINF tag shall follow a naming structure that can be expressed using a SegmentTemplate@media string using $Number$ or $Time$.

2. Playlist shall include one or more #EXT-X-PROGRAM-DATE-TIME tags to link the wall clock time to the media segment time.

3. Media presentation timestamps of the live media segments shall be relative to Unix epoch and segment duration shall be near constant, or durations shall be compensated between subsequent segments.

4. In case the segment format is based on transport stream i.e. MPEG-2 TS, the presentation time stamps shall correspond to the media presentation time stamps from the media segments received from the encoder, but wrapped in 33 bits and using a 90 Khz scale.

5. In case MPEG-2 Transport stream is used, the #X-TIMESTAMP-MAP=MPEGTS:<MPEG-2 time>, LOCAL=YYYY-MM-DDTHH:MM:SS.mmmZ shall be used to map the MPEG-2 transport stream presentation timestamps to the corresponding local time.

6. In case of multiple media playlists, media segments shall be aligned between media playlists within at least 100 milli seconds. Therefore, the earliest presentation time plus duration of newest segment in different playlist shall not differ by more than 100 milli-seconds between media playlists.

# Asset storage and recording Format

Asset storage and recording uses the ISO Base Media File format ISO/IEC 14496-12 and a storage Media Presentation Description (S-MPD) using Media presentation description defined in ISO/IEC 23009-1. Further restrictions to the track storage format are defined in clause 8.1. Further constraints to the Storage Media Presentation Description (S-MPD) constraints are defined in clause 8.2.



Figure 3 An example workflow for asset recording and archiving. In addition to the synchronized and redundant encoding, as in Figure 1, the workflow includes storing and archiving content in a cloud storage. The cloud storage may be accessed from the encoder, origin/packager or a client retrieving the media stream.

Figure 3 illustrates an example of using the specified method including cloud storage. Media presentations can be written to the cloud storge from the encoder, from the origin/packager or from a client receiving the media content. The format for asset storage and recording enables efficient and interoperable writing of content to a cloud storage, and efficient retrieval later on.

## Track format for storage of live archives

A storage track file is defined as follows.

**Storage Track file:** A stored ISOBMFF track that follows the constraints of aCMAF Track file (CMAF clause 7), without the restriction that the track starts at presentation time zero. In addition, it may also contain interleaved SegmentIndexBoxes after the first MovieFragmentBox that may occur at the start of a segment to enable random access on portions of the track or segment.

The CMAF Track format structure is used to store media content as defined in clause 7 of ISO/IEC 23000-19 [2].

In addition, this clause defines urn’s and identifiers to code properties in each track to allow recreating the manifest if needed. Both identifiers corresponding to DASH and CMAF constructs are introduced and linked. Figure 4 links the DASH identifiers and the CMAF identifier constructs. Table 2 defines the schemeURI for each identifier when used according to this proposal when signalling in a `kind` box. The schemeURI and value attribute in kind shall be set to the identifier scheme and UUID/ID respectively.

Track identifiers may be optionally present in tracks. If present, they shall be signalled using a schemeURI in the kind box in udta as and the value as a UTF-8 character encoded and null terminated UUID or id value as string. The defined schemeURI from Table 2 may be included in a storage track file.



Figure 4 Relationship between the manifest and its segments.



Figure 5 Orphaned content stream or asset with information to do reverse lookup.

1. Example CMAF identifiers and mapping to DASH identifiers (informative?).

|  |  |
| --- | --- |
| **DASH construct** | **CMAF construct** |
| MPD ID | CMAF Presentation ID |
| Period ID | CMAF Presentation ID |
| AdaptationSet  Group ID | Aligned Switching Set ID |
| AdaptationSet ID | SwitchingSet ID |
| Representation ID | CMAF track id (not to confuse with track\_id) |

1. Link between constructs in DASH and HLS and defined urn’s (informative?)

|  |  |
| --- | --- |
| Content ID (optional) | urn:mpeg:asset-storage-format:cmaf-content-id |
| CMAF Presentation ID (MPD) | urn:mpeg:asset-storage-format:cmaf-mpd-id |
| CMAF PresentationID (Period) | urn:mpeg:asset-storage-format:cmaf-period-id |
| Aligned Switching Set ID | urn:mpeg:asset-storage-format:cmaf-aligned-switching-id |
| SwitchingSet ID | urn:mpeg:asset-storage-format:cmaf-switching-id |
| CMAF track id (not to confuse with track\_id) | urn:mpeg:asset-storage-format:cmaf-track-id |

## Storage Media Presentation Description (S-MPD)

### Overview

A manifest presentation description (MPD) can be used to reference stored track files.

Table 3 summarizes the profiles and constraints for the different storage, recording and archiving options.

1. Options and profiles for Storage MPD (S-MPD)

|  |  |  |
| --- | --- | --- |
|  | @profiles | constraints |
| 24x7 live | urn:mpeg:dash:profile:isoff-live:2011 | As in clause 8.2, may use CMAF profile for DASH |
| simple live | urn:mpeg:dash:profile:isoff-live:2011 | may include SegmentTimeline, may use CMAF profile for DASH |
| VoD | urn:mpeg:dash:profile:isoff-on-demand:2011 | may use CMAF profile for DASH |

### Constraints on the S-MPD for 24x7 live archiving and recording

An S-MPD should be used to reference tracks used in an archived or stored/recorded media presentation.

Fixed duration track files with duration M are stored and signalled as long duration media segments in the S-MPD. These media segments contain all data from the archive track file except the initialization segment or CMAF Header. They may also include a SegmentIndexBox.

The receiver/archiver is configured to generate and store track files using a fixed duration M. The S-MPD uses the iso live profile, and SegmentTemplate elements with @duration set to the fixed duration M accounted for the timescale of the tracks. Each track file is stored as a media segment in the Media Presentation Description.

Table 4 specifies the constraints on the S-MPD for recording.

Annex E provides an example mapping from the I-MPD to the S-MPD.

1. Constraint on the S-MPD for recording 24x7 live content

|  |
| --- |
| 1. The MPD follows the iso live profile (urn:mpeg:dash:profile:isoff-live:2011) 2. MPD@type shall be ‘dynamic’ 3. @availabilityStartTime is set to the earliest presentation time of the first archive track segment (or if it is not known to the time the MPD was written) 4. Period@start is set to “PT0S” or a semantically equivalent value 5. MPD@timeShiftBufferDepth and Period@timeShiftBufferDepth shall not be present 6. Each AdaptationSet shall contain a SegmentTemplate element constrained as follows:    1. Each SegmentTemplate@presentationTimeOffset is present to make the media presentation timing relative to the epoch anchor, i.e. 1970-01-01T00:00:00Z, the time is MPD@availabilityStartTime - 1970-01-01T00:00:00Z adjusted for the SegmentTemplate@timescale    2. The SegmentTemplate@duration is set to the fixed duration M adjusted for SegmentTemplate@timescale    3. SegmentTemplate@media is the same for each SegmentTemplate and contains the substring “$RepresentationID$” and “$Number$”    4. SegmentTemplate@initialization is the same for each SegmentTemplate and contains the substring “$RepresentationID$” |

# Example Media Presentation Description

(Informative)

* 1. **Example I-MPD (informative)**

An example I-MPD is shown below.

|  |
| --- |
| <?xml version="1.0" encoding="utf-8"?>  <MPD  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  xmlns="urn:mpeg:dash:schema:mpd:2011"  xsi:schemaLocation="urn:mpeg:dash:schema:mpd:2011 http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-DASH\_schema\_files/DASH-MPD.xsd"  type="dynamic"  availabilityStartTime="1970-01-01T00:00:00Z"  publishTime="2022-09-30T12:49:03.521051Z"  minimumUpdatePeriod="PT2S"  maxSegmentDuration="PT2S"  minBufferTime="PT10S"  profiles="urn:mpeg:dash:profile:isoff-live:2011,urn:com:dashif:dash264">  <Period  id="1"  start="PT0S">  <AdaptationSet  id="1"  group="1"  contentType="audio"  lang="en"  minBandwidth="64000"  maxBandwidth="128000"  segmentAlignment="true"  audioSamplingRate="48000"  mimeType="audio/mp4"  codecs="mp4a.40.2"  startWithSAP="1">  <AudioChannelConfiguration  schemeIdUri="urn:mpeg:dash:23003:3:audio\_channel\_configuration:2011"  value="1" />  <Role schemeIdUri="urn:mpeg:dash:role:2011" value="main" />  <SegmentTemplate  timescale="48000"  initialization="live-$RepresentationID$.dash"  media="live-$RepresentationID$-$Time$.dash">  <SegmentTimeline>  </SegmentTimeline>  </SegmentTemplate>  <Representation  id="audio\_eng=64000"  bandwidth="64000">  </Representation>  <Representation  id="audio\_eng=128000"  bandwidth="128000">  </Representation>  </AdaptationSet>  <AdaptationSet  id="2"  group="2"  contentType="video"  par="16:9"  minBandwidth="500000"  maxBandwidth="1000000"  segmentAlignment="true"  width="1280"  height="720"  sar="1:1"  frameRate="25"  mimeType="video/mp4"  codecs="avc1.42C01F"  startWithSAP="1">  <Role schemeIdUri="urn:mpeg:dash:role:2011" value="main" />  <SegmentTemplate  timescale="600"  initialization="live-$RepresentationID$.dash"  media="live-$RepresentationID$-$Time$.dash">  <SegmentTimeline>  </SegmentTimeline>  </SegmentTemplate>  <Representation  id="video=500000"  bandwidth="500000"  scanType="progressive">  </Representation>  <Representation  id="video=1000000"  bandwidth="1000000"  scanType="progressive">  </Representation>  </AdaptationSet>  </Period>  <UTCTiming  schemeIdUri="urn:mpeg:dash:utc:http-iso:2014"  value="https://time.akamai.com/?iso" />  </MPD> </Period>  <UTCTiming  schemeIdUri="urn:mpeg:dash:utc:http-iso:2014"  value="https://time.akamai.com/?iso" />  </MPD> |

* 1. **Example D-MPD (informative)**

An example D-MPD is shown below.

|  |
| --- |
| <?xml version="1.0" encoding="utf-8"?>  <MPD  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  xmlns="urn:mpeg:dash:schema:mpd:2011"  xsi:schemaLocation="urn:mpeg:dash:schema:mpd:2011 http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-DASH\_schema\_files/DASH-MPD.xsd"  type="dynamic"  availabilityStartTime="1970-01-01T00:00:00Z"  publishTime="2022-09-30T13:49:03.521051Z"  minimumUpdatePeriod="PT2S"  timeShiftBufferDepth="PT5M"  maxSegmentDuration="PT2S"  minBufferTime="PT10S"  profiles="urn:mpeg:dash:profile:isoff-live:2011,urn:com:dashif:dash264">  <Period  id="1"  start="PT0S">  <BaseURL>dash/</BaseURL>  <AdaptationSet  id="1"  group="1"  contentType="audio"  lang="en"  minBandwidth="64000"  maxBandwidth="128000"  segmentAlignment="true"  audioSamplingRate="48000"  mimeType="audio/mp4"  codecs="mp4a.40.2"  startWithSAP="1">  <AudioChannelConfiguration  schemeIdUri="urn:mpeg:dash:23003:3:audio\_channel\_configuration:2011"  value="1" />  <Role schemeIdUri="urn:mpeg:dash:role:2011" value="main" />  <SegmentTemplate  timescale="48000"  initialization="live-$RepresentationID$.dash"  media="live-$RepresentationID$-$Time$.dash">  <SegmentTimeline>  <S t="79898181120000" d="92160" r="156" />  </SegmentTimeline>  </SegmentTemplate>  <Representation  id="audio\_eng=64000"  bandwidth="64000">  </Representation>  <Representation  id="audio\_eng=128000"  bandwidth="128000">  </Representation>  </AdaptationSet>  <AdaptationSet  id="2"  group="2"  contentType="video"  par="16:9"  minBandwidth="500000"  maxBandwidth="1000000"  segmentAlignment="true"  width="1280"  height="720"  sar="1:1"  frameRate="25"  mimeType="video/mp4"  codecs="avc1.42C01F"  startWithSAP="1">  <Role schemeIdUri="urn:mpeg:dash:role:2011" value="main" />  <SegmentTemplate  timescale="600"  initialization="live-$RepresentationID$.dash"  media="live-$RepresentationID$-$Time$.dash">  <SegmentTimeline>  <S t="998727264000" d="1152" r="156" />  </SegmentTimeline>  </SegmentTemplate>  <Representation  id="video=500000"  bandwidth="500000"  scanType="progressive">  </Representation>  <Representation  id="video=1000000"  bandwidth="1000000"  scanType="progressive">  </Representation>  </AdaptationSet>  </Period>  <UTCTiming  schemeIdUri="urn:mpeg:dash:utc:http-iso:2014"  value="https://time.akamai.com/?iso" />  </MPD> |

* 1. **Example Media Playlist**

An example media playlist according to RFC 8216 is given below

|  |
| --- |
| #EXTM3U  #EXT-X-VERSION:4  #EXT-X-MEDIA-SEQUENCE:867397718  #EXT-X-I-FRAMES-ONLY  #EXT-X-TARGETDURATION:3  #EXT-X-TIMESTAMP-MAP:MPEGTS=557701792,LOCAL=2022-10-10T12:06:56.640000Z  #EXT-X-PROGRAM-DATE-TIME:2022-10-10T12:06:56.640000Z  #EXTINF:1.92, no desc  scte35-video=500000-867397718.ts  #EXTINF:1.92, no desc  scte35-video=500000-867397719.ts  #EXTINF:1.92, no desc  scte35-video=500000-867397720.ts  #EXTINF:1.92, no desc  scte35-video=500000-867397721.ts |

# Recommended Configurations

(Informative)

## Recommended segment durations for encoder synchronization (informative)

In some cases, it is difficult to achieve segment boundary alignment. This is because audio samples in ISOBMFF contain multiple media samples (e.g., 1024 samples in an MP4 audio ISOBMFF sample) and therefore the granularity of audio and video segments may not be the same. In practice choosing values based on common frame rates and sample rates can mitigate this problem. The table below illustrated some common options to avoid misalignment.

This annex does not address the non-integer framerate case, but practice has shown that for M=L x D, fixed duration M can be achieved enabling a pattern in the media segment durations. This could be exploited to achieve the same mechanism to achieve consistency but based on super- segment duration M.

**Table 1 Example rates, timescale and segment duration at 25 fps and 30 fps**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Media type | Frame rate / sample rate | Timescale | Frames per segment | Segment duration D |
| Video | 25 fps | 25 | L\*24 | L\*0,96 |
| Audio | 48 Khz | 48000 | L\*45 | L\*0,96 |
| Timed text | - | 1000 | 1 or more | L\*0,96 |
| Metadata | - | 1000 | 1 or more | L\*0,96 |
| Video | 30 fps | 25 | L\*48 | L\*1,6 |
| Audio | 48 Khz | 48000 | L\*75 | L\*1,6 |
| Timed text | - | 1000 | 1 or more | L\*1,6 |
| Metadata | - | 1000 | 1 or more | L\*1,6 |

# Example Applications

(Informative)

## Redundant and high availability distribution:

In this application encoder synchronization method is used to achieve setups with multiple encoders and receiver mainly to be fault tolerant, if one encoder or receiver fails switchover is possible. In this use case each distributed encoder produces the full bit-rate ladder, and similarly the redundant packagers generate the complete media presentations. In case one distribution encoder or packager fails redundant streams are available due to the cross transmission.

## Distributed encoding:

In this application different encoders generate different bitrates of different codecs of media types. In this case for example high bitrate Ultra High Definition resolutions are encoded on a different distributed entity. The main benefit of using the formats and architecture described in this document is that representations will be segment aligned and synchronized without explicit communication between encoders and/or packagers. In this case, encoder 1 and 2 post to all packagers. Each packager generates the complete media presentation.

## A/B watermarking:

In this application, different sources are marked with an A or a B watermark resulting in watermarked segments. Segment sequences can then be combined to generate user specific watermarks to avoid content leaking. In such case the processed and outputs tracks of encoders generating A and B watermarks can be generated using the encoder synchronization formats in this document. In this case the downstream entities can also easily apply the watermark by A/B sequencing.

## Metadata Sources:

Some metadata tracks such as timed text, programme metadata and/or accessibility metadata, e.g., Hard of Hearing could be generated at distributed sources, possibly using different computational infrastructures, such as based on automatic translation or speech interpretation. By applying the formats disclosed in this document it is easy to synchronize such tracks back for streaming in a manifest (late binding) and no explicit communication with these third-party data generators is required.

## Live Archives

By using the timing and segmentation formats from this document, stored media is linked to the Unix Epoch timeline. Segments can therefore always be mapped back to original programme times and aligned to electronic programme guides without explicit conversion requirements or explicit additional scheduling information.

# Example methods of carriage of Source timing (Informative)

(Informative)

The input signal to the contribution encoder in Figure 1 may carry timing information based on the following example formats:

1. ISO/IEC MPEG-2 TS usage of adaptation descriptors for timing, boundary labelling
2. SCTE 35 carriage of timing in Separate PID in a TS stream
3. Carriage of timing in VANC in baseband
4. Carriage of timing/ Markers using EBP for TS segment
5. Carriage of timing/markers in ISOBMFF
6. Carriage of timing in the timing SEI

The source signal coming into the transcoder can carry timing in several format depending on the signal format. In a baseband uncompressed format, timing can be carried in the digital timing frame or though production automation system messages. In a compressed format, the elementary stream can carry timing through a timing SEI message sent with every access unit. At a container level in a packetized streaming format, timing can be carried along with each packet or box or in synchronized packet or boxes. The origin of the signal may determine the type of format that timing is carried in the source signal stream and the transcoder may need to extract the timing information depending on the origin of the source signal.

Baseband video can be carried in a digital timing frame interpreted as a string of bits in a serial digital interface (SMPTE 259M). Timing can be carried as SMPTE timecode (SMPTE 12M) in the vertical blanking interval (VBI) as vertical interval timecode with each digital video time frame. Alternatively, the baseband signal can be injected with production automation messages and be feed into an encoder (SCTE 104). These messages can contain commands like insertion CMDs and carry along a timestamp message that is in the form of UTC or VITC time information which is needed to create splice points in the resulting compressed stream.

If a Video is encapsulated in a packetized stream structure of MPEG-2 TS. The PES packet can carry timing (and IDR boundary information collectively called EBP- encoder boundary point) in either the public adaptation field though the timeline descriptor (TEMI) using an AF descriptor tag of 0x04 (ISO/IEC 13818-1 Annex U -Carriage of timeline and external media information) which can carry either NTP, or PTP timing formats. A private AF descriptor tag can also be used to convey EBP information in a more concise format with NTP timing carried in this manner (Cablelabs EBP Specification). Alternatively, a media time can be carried with a PTS timestamp which give a unique timestamp in a media stream that has a duration of under 26 hours. These type of carriage structures can be carried potentially with every access unit.

In the MPEG-2 TS, timing can also be carried in a separate PID of the packetized stream that can carry SCTE 35 messages which can carry splice information or program event mediapoints (see SCTE 35). In the SCTE35, splice commands can be used that can carry PTS time through its splice\_info, or time\_signal constructs. In the time\_signal descriptor field that can be added additionally to splice\_info commands, UTC or PTP time can be carried along with the descriptor. If a regular timing indicator is needed, this can be done either with regularly placed time\_descriptors or through the use of regularly placed splice\_null commands.

Video Elementary streams can also be carried in an MPEG-4 File format (ISO/IEC 14496-12) where video can be carried in ISOBMFF Boxes along with other type of boxes. One type of box in the file format stream that can carry a timing reference is the prft box that can carry an NTP time format.

# Annex E Example method for redundant encoder tracks generation

(Informative)

The annex details an example method to implement redundant distributed encoder operation using formats. The distribution encoders are expected to receive an input signal with common per frame timing information. The assumption is that the timing information can be mapped back be relative to 1-1-1970 00:00:00 UTC excluding leap seconds as a common time anchor. If internal timing is not relative to such a common anchor, then a Synchronization Time Stamp (STS) is used to map input timestamps to output timestamps relative to this anchor.

The STS (Synchronization Time Stamp) is the difference between the zero time of the input signal to 1-1-1970 00:00:00 UTC excluding leap seconds.

The output frame time is calculated using:

*(1)*

Frame times lie on a continuous timeline on the track. In some cases, the input time will need to be converted from its original form to a continuous form that also matches the output timescale to derive .

In case the *track\_timescale* and the input timescale are not the same, the input time should also be adjusted for this by a timescale conversion. In this case the original input time is also multiplied by the *track\_timescale* and later divided by the original (input) timescale. In this case to avoid rounding error and encoders should only choose integer multiples of the input timescale as the output .

The STS may not directly be obtained from the encoder configuration, but may, for example, be retrieved from timing markers or other metadata markers (e.g. SCTE-35) in the input signal.

Another conversion that the encoder is required to make up for is for timestamp wraparounds to map to a continuous timeline. For example, an MPEG-2 timestamp would need to be converted to a continuous timeline even when the timestamp wraps due to overflow. This conversion is conducted by the encoder.

In equation (1), *track\_timescale* is the timescale used by the media track, where timescale is as defined in ISO/IEC 14496-12 in the mdhd (MediaHeader box).

To calculate segment boundaries equation (2) is used:

*(2)*

The segment boundary of the segment is the earliest presentation time of that segment. To calculate the number of the next segment (i.e., the next ), equation (3).

*(3)*

In equation (3) *now* is the time in seconds relative to epoch calculated using a real-time system routine. Next K is the value of K to compute the next segment boundary.

An encoder joining a synchronized encoder session, computes Next\_K and may also compute the previous K (Next\_K - 1) and start buffering and encoding frames based on received input. Once a complete media segment is encoded with the segment boundary computed as based on (2) it can be transmitted. Joining or leaving encoders use the same segment naming, thus identical Representation@id and SegmentTemplate@media and SegmentTemplate@initialization are used in the I-MPD generated by different encoders.

The @availabilityStartTime attribute in the I-MPD is set to 1970-01-01T00:00:00Z and the Period@start should be set to PT0S. In case Period@start is not PT0s or a semantically equivalent

value, the corresponding SegmentTemplate@presentationTimeOffset attributes are used to compensate the media presentation times relative to the period start time. The media presentation times are always relative to 00:00:00 1970-01-01 UTC excluding leap seconds.

Any receiver may identify identical segments based on the baseMediaDecodeTime and the naming scheme identified from the AdaptationSet@SegmentTemplate and Representation@id. Any segment from the same Representation of track is interchangeable given that it has the same baseMediaDecodeTime. This way receivers can detect redundant/duplicate segments and/or additional representations or tracks.

# Annex F Example method for generating a recording S-MPD from an I MPD

(Informative)

This annex provides an example of how a recording S-MPD can be created from an I-MPD.

A receiver may create and write a S-MPD based on an I MPD following the steps in Table 5. A receiver may write the received segments as track files references in the S-MPD using the steps detailed in Table 6.

When the live stream stops, and no more segments are expected, the MPD@type attribute may be changed from ‘dynamic’ to ‘static’ and the SegmentTemplate@endNumber, Period@duration attributes may be set in the manifest.

1. Steps for writing segments to the S-MPD for 24x7 live content

|  |
| --- |
| 1. Initialise the S-MPD based on constraints from Table 3, read the configuration values for duration M from a memory 2. Set the MPD@availabilityStartTime and SegmentTemplate@presentationTimeOffset fields to the earliest presentation time of a received archive track file (or if this is not known to the time the MPD was written) 3. Read the I-MPD and create the corresponding AdaptationSets in the S-MPD,    1. The SegmentTimeline is not used but instead SegmentTemplate with @duration is used, @duration being set to M accounted for the SegmentTemplate@timescale    2. Representation@id matches between the I-MPD and S-MPD    3. The @startNumber is set to floor(MPD@availabilityStartTime (in seconds relative to 1970-01-01T00:00:00Z)/M)    4. SegmentTemplate@media from I-MPD but replacing $Number$ instead of $Time$ 4. Only write the S-MPD to storage if it was not already present in the storage |

1. Steps for writing the archive track files for 24x7 live content or encoder synced content (assuming the S-MPD is already written)

|  |
| --- |
| 1. Compute the next archive track boundary next\_L, the L + 1th archive track file since the established anchor 1970-01-01T00:00:00Z by next\_L = ceil(now/M), where now is the seconds based timestamp relative to Epoch or the established anchor. 2. Compute the next archive track file boundary B as next\_L \* D \* output track timescale 3. Identify from the input signal the frame time information, in this case the input is a CMAF media segment, it is the CMAF media segments earliest presentation time, potentially adjust it with a synchronisation timestamp 4. Append the incoming media segment in memory to the continue creating the archive track file 5. In case the media segment earliest presentation time plus the media segment duration equals or is greater than B goto 6 otherwise to 3 6. Upload the in memory archive track file using the SegmentTemplate URL generation substituting $RepresentationID$ for the Representation@id derived from the input track and replacing $Number$ by next\_L – 1 7. GoTo 1 |

Bibliography

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

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

[3] ISO/IEC 23000-19, Information technology — Multimedia application format (MPEG-A) — Part 19: Common media application format (CMAF) for segmented media