
Information technology — Dynamic adaptive streaming over HTTP (DASH) —

Part 3: Implementation guidelines

*Technologies de l'information — Diffusion en flux adaptatif
dynamique sur HTTP (DASH) —*

Partie 3: Lignes directrices de mise en oeuvre



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2015, Published in Switzerland

All rights reserved. Unless otherwise specified, 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
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword	v
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms, definitions, and abbreviated terms	1
4 General	2
4.1 System overview	2
4.2 Normative parts	2
4.3 Main design principles	3
4.3.1 Common timeline	3
4.3.2 Data model	3
4.3.3 Segments	4
4.3.4 Segment types	5
4.3.5 Segment addressing schemes	5
4.3.6 Stream access points	6
4.3.7 Remote elements	6
4.3.8 Events	7
4.3.9 General-purpose descriptors	7
4.4 Background on DASH profile concept	8
4.5 Dynamic aspects	8
5 Guidelines for content generation	10
5.1 General guidelines	10
5.1.1 Video content generation	10
5.1.2 Audio content generation	12
5.1.3 Content preparation for live streaming	14
5.1.4 Guidelines for generation of segment file names	14
5.2 Guidelines for ISO-BMFF content generation	17
5.2.1 On-demand streaming	17
5.2.2 Live streaming	21
5.2.3 Enabling trick modes	23
5.2.4 Support for SubRepresentations	24
5.2.5 Enabling delivery format to storage file format conversion	26
5.3 Guidelines for MPEG-2 TS content generation	30
5.3.1 General recommendations	30
5.3.2 Live streaming	31
5.3.3 On demand streaming	32
5.4 Guidelines for Advertisement Insertion	33
5.4.1 Use cases	33
5.4.2 Architectures and workflows	34
5.4.3 App-driven ad insertion	36
5.5 DASH MPD and Segment-based Live Service Offering	37
5.5.1 Preliminaries	37
5.5.2 Service Offering Requirements and Guidelines	38
5.5.3 Client requirements and guidelines	41
5.6 Guidelines for low latency live service	43
5.6.1 Use case	43
5.6.2 General Approach: Chunked transfer	43
5.6.3 MPD generation	43
6 Client implementation guidelines	44
6.1 General	44
6.2 Client architecture overview	44
6.3 Example of client operation	45

6.4	Timing model for live streaming	45
6.4.1	General	45
6.4.2	MPD information	45
6.4.3	MPD times	46
6.4.4	Context derivation	46
6.4.5	Derivation of MPD times	47
6.4.6	Addressing methods	47
6.4.7	Scheduling playout	48
6.4.8	Validity of MPD	48
6.5	MPD retrieval	48
6.6	Segment list generation	49
6.6.1	General	49
6.6.2	Template-based generation of segment list	50
6.6.3	Playlist-based generation of segment list	51
6.6.4	Media segment list restrictions	51
6.7	Rate adaptation	52
6.8	Seeking	53
6.9	Support for trick modes	53
6.10	Stream switching	54
6.11	Client support for dependent representations	54
6.11.1	General	54
6.11.2	Client trick-mode support using SubRepresentations	55
6.12	Events	55
6.12.1	General processing	56
6.12.2	Inband events	56
7	Extending DASH	56
7.1	Extension of MPD Schema in external namespace	56
7.1.1	General	56
7.1.2	Example	56
	Bibliography	58

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 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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC 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 meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: [Foreword - Supplementary information](#)

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This second edition cancels and replaces the first edition (ISO/IEC TR 23009-3:2015), which has been technically revised.

It also incorporates new features from ISO/IEC 23009-1:2014/Amd 1:2015 and ISO/IEC 23009-1:2014/Cor 1:2015.

ISO/IEC TR 23009 consists of the following parts, under the general title *Information technology — Dynamic adaptive streaming over HTTP (DASH)*:

- *Part 1: Media presentation description and segment formats*
- *Part 2: Conformance and reference software*
- *Part 3: Implementation guidelines*
- *Part 4: Segment encryption and authentication*

Introduction

This part of ISO/IEC TR 23009 provides guidelines for implementation and deployment of streaming media delivery systems based on the ISO/IEC 23009 series. These guidelines include the following:

- guidelines for streaming content generation;
- guidelines for implementation of streaming clients;
- guidelines for deployment of systems designed based on the ISO/IEC 23009 series.

Information technology — Dynamic adaptive streaming over HTTP (DASH) —

Part 3: Implementation guidelines

1 Scope

This part of ISO/IEC TR 23009 provides technical guidelines for implementing and deploying systems based on ISO/IEC 23009-1.

2 Normative references

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

ISO/IEC 13818-1, *Information technology — Generic coding of moving pictures and associated audio information — Part 1: Systems*

ISO/IEC 13818-2, *Information technology — Generic coding of moving pictures and associated audio information — Part 2: Video*

ISO/IEC 14496-2, *Information technology — Coding of audio-visual objects — Part 2: Visual*

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

ISO/IEC 23003-1, *Information technology — MPEG audio technologies — Part 1: MPEG Surround*

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

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

ITU-T Rec. H.264 | ISO/IEC 14496-10, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ITU-T Rec. H.265 | ISO/IEC 23008-2, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

3 Terms, definitions, and abbreviated terms

For the purposes of this document, the terms and definitions given in ISO/IEC 23009-1; the video coding terms defined in ISO/IEC 13818-2, ISO/IEC 14496-2, ITU-T Rec. H.264 | ISO/IEC 14496-10 and ITU-T Rec. H.265 | ISO/IEC 23008-2; and the audio coding terms defined in ISO/IEC 13818-1, ISO/IEC 14496-3, ISO/IEC 23003-1 and ISO/IEC 23003-3 apply.

4 General

4.1 System overview

[Figure 1](#) shows a typical deployment scenario for Dynamic Adaptive Streaming over HTTP (DASH). The media encoding process generates segments containing different encoded versions of one or several of the media components of the media content. Each segment contains streams required for decoding and displaying a time interval of the content. The segments are then hosted on one or several media origin servers along with a Media Presentation Description (MPD) file. The media origin server may be a plain HTTP server conforming to RFC2616. The MPD information provides instructions on the location of segments as well as the timing and relation of the segments, i.e. how they form a media presentation. Based on this information in MPD, a DASH streaming client requests the segments using HTTP GET or partial GET methods. The client fully controls the streaming session, i.e. it manages the on-time request and smooth playback of the sequence of segments, potentially adjusting bitrates or other attributes, e.g. to react to changes of the device state or the user preferences.

As long as the MPD provides RESTful HTTP-URIs for the Segment locations, the HTTP-based delivery infrastructure may be kept unaware of the actual data that is delivered. This feature permits the reuse of existing HTTP caches and Content Distribution Networks (CDNs) for massively scalable Internet media distribution.

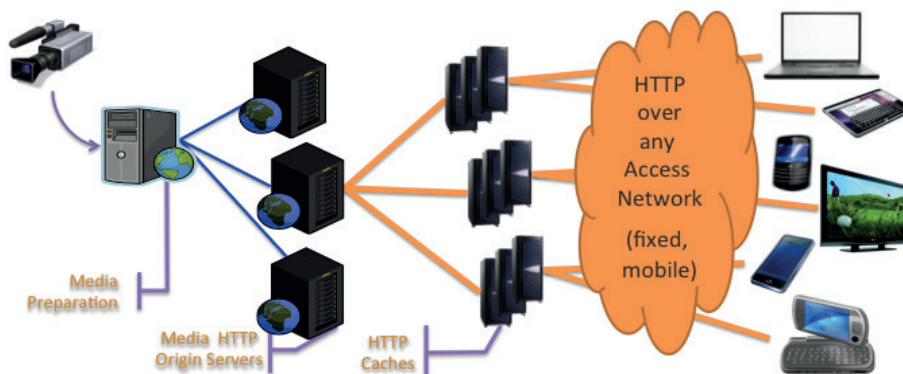


Figure 1 — Example of DASH-based media distribution architecture

4.2 Normative parts

The ISO/IEC 23009 specification serves as an enabler for Dynamic Adaptive streaming over HTTP. It does not specify a full end-to-end service, but rather base building blocks to enable efficient and high-quality streaming services over the Internet. Specifically, ISO/IEC 23009-1 defines two formats as shown in [Figure 2](#).

- The Media Presentation Description (MPD) describes a Media Presentation, i.e. a bounded or unbounded presentation of media content. In particular, it defines formats to announce resource identifiers for Segments as HTTP-URLs and to provide the context for these identified resources within a Media Presentation.
- The Segment format specifying the format of the entity body of an HTTP response to an HTTP GET request or a partial HTTP GET, with the indicated byte range through HTTP/1.1 as defined in RFC 2616, to a resource identified in the MPD.

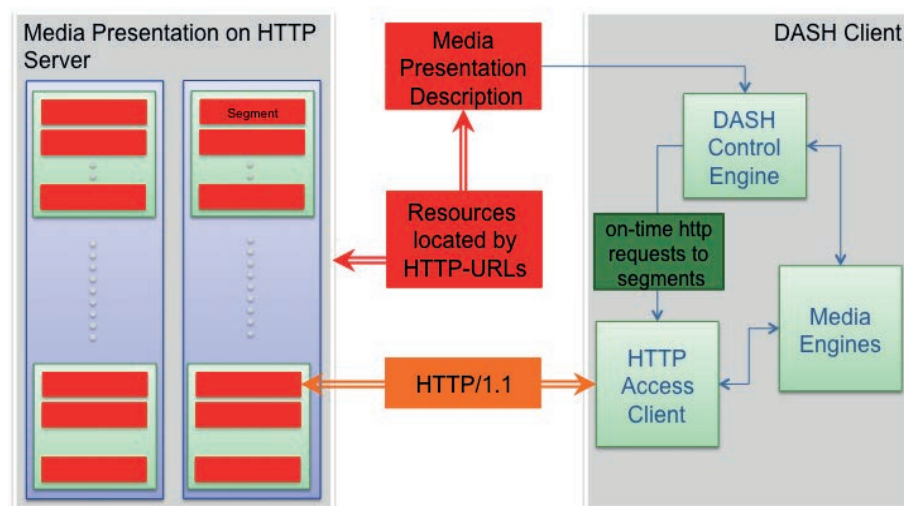


Figure 2 — Standardized aspects in DASH

NOTE Normative components are marked in red.

Other aspects, such as client implementations of control and media engines, are not defined as normative parts of the ISO/IEC 23009 specification.

4.3 Main design principles

4.3.1 Common timeline

ISO/IEC 23009 requires encoded versions of media content components (e.g. video, audio) to have a common timeline. The presentation time of access units within the media content is mapped to a global common presentation timeline, referred to as Media Presentation Timeline. This allows synchronization of different media components and enables seamless switching between different encoded versions of media content.

4.3.2 Data model

In ISO/IEC 23009, the organization of a multimedia presentation is based on a hierarchical data model shown in Figure 3.

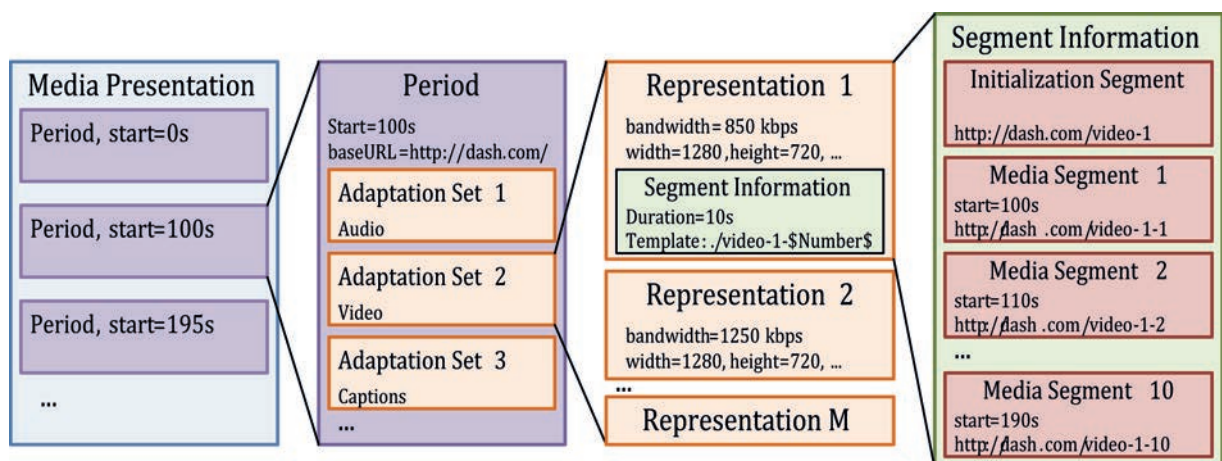


Figure 3 — DASH hierarchical data model

This model consists of the following elements.

- **Media Presentation Description (MPD):** Describes the sequence of Periods that make up a DASH Media Presentation.
- **Period:** Interval of the Media Presentation, where a contiguous sequence of all Periods constitutes the Media Presentation.
- **Adaptation Set:** Represents a set of interchangeable encoded versions of one or several media content components. For example, there may be an Adaptation Set for video, one for primary audio, one for secondary audio, and one for captions. Adaptation Sets may also be multiplexed, in which case, interchangeable versions of the multiplex may be described as a single Adaptation Set. For example, an Adaptation Set may contain both video and main audio for a Period.
- **Representation:** Describes a deliverable encoded version of one or several media content components. Any single Representation within an Adaptation Set should be sufficient to render the contained media content components. Clients may switch from Representation to Representation within an Adaptation Set in order to adapt to network conditions or other factors.
- **Segment:** Content within a Representation may be further divided in time into Segments of fixed or variable length. Each segment is referenced in the MPD by means of a URL. Thus, a Segment defines the largest data unit that can be accessed by means of a single HTTP request.

4.3.3 Segments

Segments contain encoded chunks of media components. They may also include information on how to map the media segments into the media presentation timeline for switching and synchronous presentation with other Representations.

4.3.3.1 Segment availability timeline

The Segment Availability Timeline is used to signal clients the availability time of segments at the specified HTTP URLs. These times are provided in wall-clock times. Before accessing the Segments at the specified HTTP URL, clients compare the wall-clock time to Segment availability times.

For on-demand content, the availability times of all Segments are identical. All Segments of the Media Presentation are available on the server once any Segment is available. Thus, the MPD is a static document.

For live content, the availability times of Segments depend on the position of the Segment in the Media Presentation Timeline. Segments become available with time as the content is produced. Thus, the MPD is updated periodically to reflect changes in the presentation over time. For example, Segment URLs for new segments may be added to the MPD; old segments that are no longer available may be removed from the MPD. Updating the MPD may not be necessary if Segment URLs are described using a template.

4.3.3.2 Segment duration

The duration of a segment represents the duration of the media contained in the Segment when presented at normal speed. Typically, all Segments in a Representation have the same or roughly similar duration. However, Segment duration may differ from Representation to Representation. A DASH presentation can be constructed with relative short segments (for example a few seconds) or longer Segments including a single Segment for the whole Representation.

Segments cannot be extended over time; a Segment is a complete and discrete unit that shall be made available in its entirety.

4.3.3.3 Sub-segments

Segments may be further subdivided into Sub-segments.

If a Segment is divided into Sub-segments, these are described by a Segment Index, which provides the presentation time range in the Representation and corresponding byte range in the Segment occupied by each Sub-segment. Clients may download this index in advance and then issue requests for individual Sub-segments using HTTP partial GET requests.

The Segment Index may be included in the Media Segment, typically at the beginning of the file. Segment Index information may also be provided in separate file containing Index Segments.

4.3.4 Segment types

4.3.4.1 General

ISO/IEC 23009-1 defines the following four types of segments:

- Initialization Segments;
- Media Segments;
- Index Segments;
- Bitstream Switching Segments.

4.3.4.2 Initialization segments

Initialization Segments contain initialization information for accessing the Representation and it does not contain any media data with an assigned presentation time. Conceptually, the Initialization Segment is processed by the client to initialize the media engines for enabling play-out of Media Segments of the containing Representation.

4.3.4.3 Media segments

A Media Segment contains and encapsulates media streams that are either described within this Media Segment or described by the Initialization Segment of this Representation or both. Media Segments must contain a whole number of complete Access Units and should contain at least one Stream Access Point (SAP) for each contained media stream. Other requirements applicable to Media Segments are described in ISO/IEC 23009-1:2014, 6.2.3.

4.3.4.4 Index segments

Index Segments contain information that is related to Media Segments, including timing and access information for Media Segments or Subsegments. An Index Segment may provide information for one or more Media Segments. The Index Segment may be media format specific and more details are defined for each media format that supports Index Segments.

4.3.4.5 Bitstream switching segments

A Bitstream Switching Segment contains data enabling switching to the Representation it is assigned to. It is media format specific and more details are defined for each media format that permits Bitstream Switching Segments. At most, one bitstream switching segment can be defined for each Representation.

4.3.5 Segment addressing schemes

ISO/IEC 23009 allows several alternative schemes for addressing Segments from an MPD. Specifically, a segment list for a Representation or Sub-Representation can be specified by either

- *SegmentBase* element, provided when Representation contains a single media Segment,
- *SegmentList* elements, providing a set of exact URL(s) for Media Segments, and

— *SegmentTemplate* element, providing a template form of URL(s) for Media Segments.

Details of these schemes are described in ISO/IEC 23009-1:2014, 5.3.9.

4.3.6 Stream access points

A Stream Access Point (SAP) is a position in a Representation enabling playback of a media stream to be started using only the information contained in Representation data starting from that position onwards (preceded by initializing data in the Initialization Segment, if any).

Stream access points are usually defined for each Media Segment, and used to support variety of streaming client operations, including the following:

- stream switching, e.g. for adaptation to changes in network bandwidth or other events;
- random access (seek and rewind operations);
- trick modes.

ISO/IEC 14496-12:2015, Annex I defines six possible types of SAPs. The mappings between SAP types and commonly used video prediction structures, such as Open GOP and Closed GOP structures are explained in [5.1.1.4](#).

4.3.7 Remote elements

Remote elements are elements that are not fully contained in the MPD document but are referenced in the MPD with an HTTP-URL using a simplified profile of XLink.

A remote element has two attributes, `@xlink:href` and `@xlink:actuate`. `@xlink:href` contains the URL for the remote element entity, while `@xlink:actuate` specifies the resolution model. The value "onLoad" requires immediate resolution at MPD parse time (i.e., beginning of time or after an MPD update), while "onRequest" allows deferred resolution at a time when an XML parser accesses the remote element. While there is no explicit timing model for earliest time when deferred resolution can occur, ISO/IEC 23009-1 strongly suggests it should be close to the expected playout time of the corresponding period.

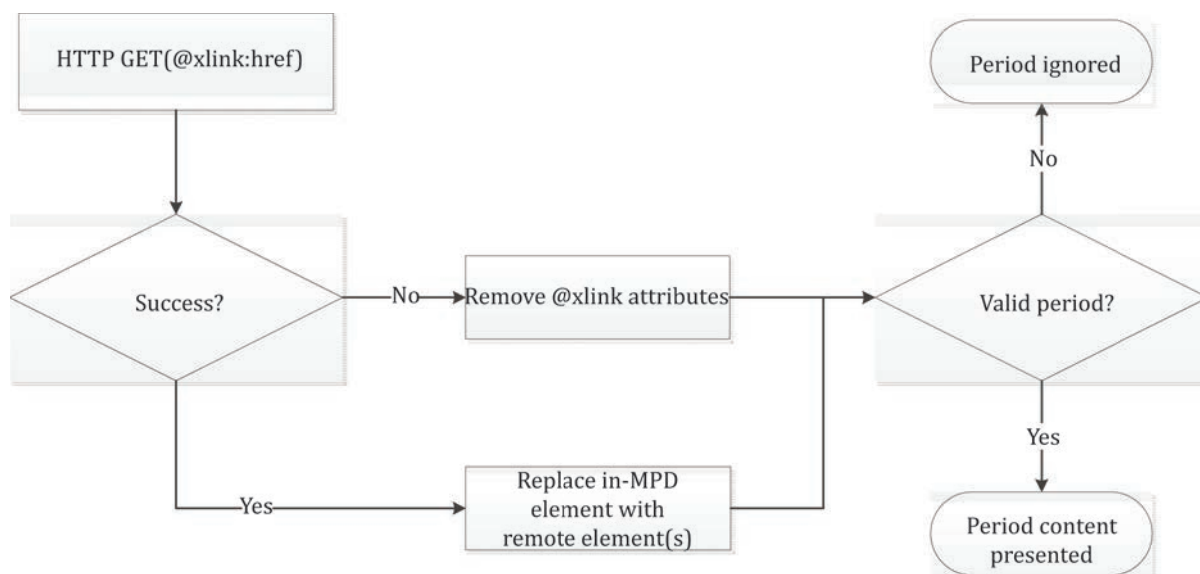


Figure 4 — XLink resolution

Resolution (a.k.a. dereferencing) consists of two steps. Firstly, a DASH client issues an HTTP GET request to the URL contained in the `@xlink:href` attribute of the *in-MPD element*, and the XLink

resolver responds with a *remote element entity* in the response content. In case of an error response or syntactically invalid remote element entity, the latter is ignored, and the `@xlink:href` and `@xlink:actuate` attributes are removed from the *in-MPD element*.

If syntactically valid remote element entity was received, it will replace the in-MPD element. “Slate” functionality, i.e. having default content that plays if dereferencing fails, can be achieved if remote periods are also valid when XLink attributes are removed.

Once a remote element entity is resolved into a fully specified element, XLink attributes that originated from the in-MPD element are removed, hence it cannot be resolved again. However, the remote element entity may contain an `@xlink:href` attribute, which contains a new XLink URL allowing repeated resolution.

If the value of the `@xlink:href` attribute is `urn:mpeg:dash:resolve-to-zero:2013`, HTTP GET request is not issued and the in-MPD element is removed from the MPD. This special case is used when a remote element can be accessed (and resolved) only once during the time at which a given version of MPD is valid.

4.3.8 Events

DASH events are messages having type, timing, and optional payload. They can appear either in MPD (as period-level event stream) or inband, as ISO-BMFF boxes of type ``emsg``. The ``emsg`` boxes appear at the very beginning of the segment, so that DASH client will need a minimal amount of parsing to detect them.

DASH defines several events that are processed directly by a DASH client. MPD Validity Expiration, MPD Patch and MPD Update events signal to the client that the MPD needs to be updated and provide the publish time of the MPD that should be used. MPD Patch provides an additional XML patch that can be applied to the client’s in-memory representation of MPD, as an alternative to requesting a new MPD. MPD Update event provides a complete new MPD.

NOTE Unless segments carrying MPD Patch or MPD Update use secure transport, such as HTTPS, a man-in-the-middle attack altering these can be mounted by a malicious entity.

User-defined events are also possible. The DASH client does not deal with them directly, they are passed to an application or discarded if there is no application willing to process these events. A possible client implementation would register callbacks to specific event types, and these would be called on arrival of these events.

In the ad insertion context, user-defined events can be used to signal information, such as cue messages.

Care should be taken to make sure that addition of events would not violate the declared bitrate constraints.

Events should be used for purposes that are related to media and are meaningful, e.g. when consumed in non-realtime applications, such as VoD, nPVR, etc.

4.3.9 General-purpose descriptors

DASH defines descriptors with clear scope. Some descriptors, such as `FramePacking`, convey specific properties of coded media that need to be supported in order to play a representation, while others (e.g. `Rating`) convey informational metadata. `SupplementalProperty` and `EssentialProperty` are an exception to the rule, these two general-purpose descriptors that are provided for user-defined extensions without a pre-defined scope.

`SupplementalProperty` is strictly informational, a client recognizing the specific scheme may benefit from the information provided in the descriptor, while a client that does not recognize it will function correctly.

`EssentialProperty` is used to provide information that is essential to the ability to present one or more representations. If a client does not recognize the scheme, it may be unable to present the media correctly. The `@id` property of DASH descriptors provides grouping semantics: if two descriptors

EssentialProperty descriptors with different schemes all having same @id value, recognizing just one of these is sufficient for correct playback.

General purpose descriptors can be present at several different levels in the MPD, from MPD down to Representation.

4.4 Background on DASH profile concept

Profiles of DASH are defined so as to enable interoperability and the signalling of the use of features. A DASH client complying with a specific profile does not necessarily have to implement features that are not required by it. By complying with a profile, a content author declares the boundaries on a set of DASH features that will be needed for successful presentation.

ISO/IEC 23009-1 defines six profiles, shown in [Figure 5](#). Profiles are organized in two categories based on file format used for segments. Three profiles use ISO Base media file format, two profiles use MPEG-2 transport stream (TS), and full profile supports both segment formats. These profiles do not impose restrictions on codecs encapsulated by these file formats.

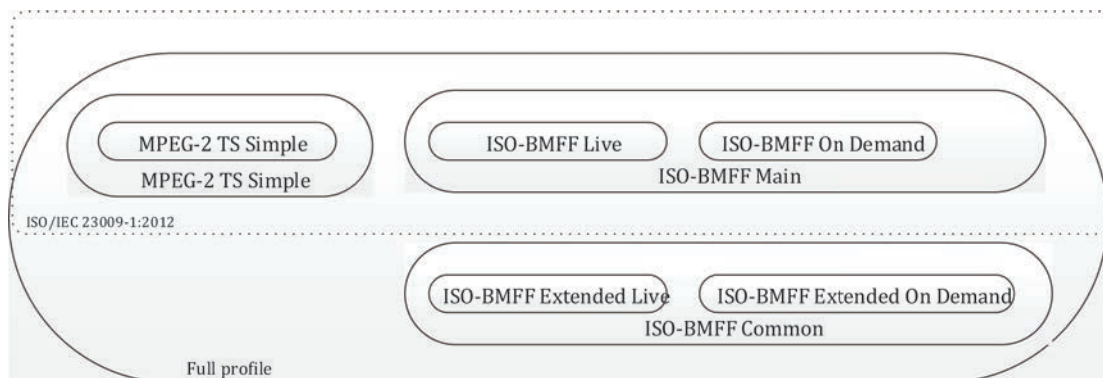


Figure 5 — DASH Profiles

In addition to six profiles defined in ISO/IEC 23009-1:2012, additional profiles may also be defined by both ISO/IEC and external entities. Such profiles may also impose constraints on codecs, resolutions, bit-rates, segment durations, and other system parameters.

Features added by the second edition of DASH, ISO/IEC 23009-1:2014, are excluded from these profiles. These are covered by the ISO-BMFF Common, Extended ISO-BMFF On Demand, and Extended ISO-BMFF Live profiles, added in ISO/IEC 23009-1:2014/Amd 1:2015.

4.5 Dynamic aspects

DASH Media Presentations with **MPD@type** set to “dynamic” enable that media is made available over time and its availability may also be removed over time. This has two major effects, namely:

- The content creator can announce a DASH Media Presentation for which not all content is yet available, but only gets available over time.
- Clients are forced into some timed schedule for the playout, such that they follow the schedule as desired by the content author.

Dynamic services may be used for different types of services.

- Dynamic Distribution of Available Content:** Services for which content is made available as dynamic content, but the content is entirely available prior to distribution. In this case, the details of the Segments are known and can be announced in a single MPD without MPD updates.

- b) **MPD-controlled Live Service:** Services for which the content is typically generated on the fly and the MPD needs to be updated occasionally to reflect changes in the service offerings. For such a service, the DASH client operates solely on information in the MPD.
- c) **MPD and Segment-controlled Live:** Services for which the content is typically generated on the fly and the MPD may need to be updated on short notice to reflect changes in the service offerings. For such a service, the DASH client operates on information in the MPD and is expected to parse segments to extract relevant information for proper operation. This addresses the use cases 4 and 5, but also takes into account the advanced use cases.

Dynamic and Live services are typically controlled by different client transactions and server-side signalling.

For initial access to the service and joining the service, an MPD is required. MPDs may be accessed at join time or may have been provided earlier, for example, along with an Electronic Service Guide. The initial MPD or join MPD is accessed and processed by the client and the client having a globally accurate clock can analyze the MPD and extract suitable information in order to initiate the service. This includes, but is not limited to the following:

- identifying the currently active Periods in the service and the Period that expresses the live edge (for more details see below);
- selecting the suitable media components by selecting one or multiple Adaptation Sets. Within each Adaptation Set selecting an appropriate Representation and identifying the live edge segment in each Representations. The client then issues requests for the segments.

The MPD may be updated on the server based on certain rules and clients consuming the service are expected to update MPDs based on certain triggers. The triggers may be provided by the MPD itself or by information included in segments. Depending on the service offering, different client operations are required as reflected in [Figure 6](#).

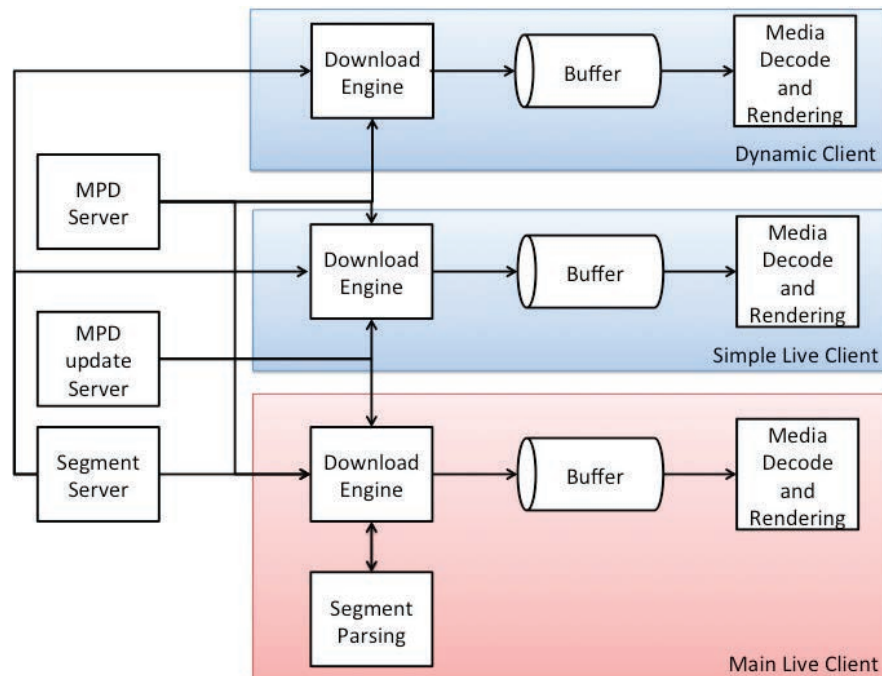


Figure 6 — Different client models

The basic operations of the different clients are as follows.

- a) **Dynamic DASH Client:** Such a client is creating a list of available segments based on a single MPD. It then joins by downloading segments at the live edge or may use the available time shift buffer.

- b) **Simple Live Client:** Such a client includes all features of the dynamic client. In addition, it updates the MPD based on information in the MPD in order to extend the segment list at the live edge. MPDs are refetched and revalidated when the currently available MPD expires, i.e. it can no longer be used for segment URL generation.
- c) **Main Live Client:** Such a client includes all features of the dynamic client and an MPD-based DASH client. In addition, it updates the MPD based on information in the segments if the service offering provides this feature. MPDs are refetched and revalidated when the currently available MPD expires based on expiry information in the Segments.

5 Guidelines for content generation

5.1 General guidelines

5.1.1 Video content generation

5.1.1.1 General

This section presents a set of recommended practices for preparing video content for streaming using MPEG video codecs. It uses terms and concepts defined in ISO/IEC 13818-2, ITU-R Rec.H.264 | ISO/IEC 14496-10, and ITU-T Rec. H.265 | ISO/IEC 23008-2.

5.1.1.2 Enabling bandwidth adaptation

In order to support bandwidth adaptation, video content should be encoded at plurality of rates, covering the range of operational rates of the network. Such encodings should be declared as Representations with different `@bandwidth` attributes and presented in MPD as an Adaptation Set.

Video encodings within each Representation may be produced by using either Variable Bit-Rate (VBR) or Constant Bit-Rate (CBR) modes. In both cases, the `@bandwidth` attributes should be set to the maximum of the bitrate averaged over `@minBufferTime` in each Representation. When VBR encoding is used, it is also recommended to provide Index Segments, allowing client to access exact length information for each encoded segment.

The rates of different encodings should not present large gaps, as switching between them could become noticeable. Ideally, the rates in adjacent operating points should not be more than 2x apart. For example, for mobile clients connected over 3G networks and bandwidth in the range from 100 Kbps to 3 Mbps, a reasonable set of operating points may include: 100 kbps, 200 kbps, 400 kbps, 600 kbps, 1,2 Mbps, 1,8 Mbps, and 2,5 Mbps.

In order to produce encodings with different target bitrates, video encoders may be instructed to use different spatial resolutions, framerates, or other parameters needed to achieve acceptable visual quality. Such parameters should be signaled by `@width`, `@height`, `@frameRate` or other relevant attributes of Adaptation Sets and Representations. Additional properties of encoded video may be also signaled by means of `EssentialProperty` or `SupplementalProperty` descriptors with URNs and schemas defined in ISO/IEC 23001-8 or other relevant specifications.

In performing encoding, it may be desirable to minimize fluctuations of quality in the encoded content. VBR encoding is naturally more suitable for this purpose. Additionally, if system design permits the use of segments with variable durations, such durations may also be adjusted to accommodate changing complexities in video content. For example, segment boundaries may be placed at frames that are naturally encoded as I- or IDR-frames due to scene changes in video content. However, it is always a good practice to minimize deviation of segment durations to the extent possible.

5.1.1.3 Initialization segments

Initialization segments should be used to communicate sequence- and picture-level parameters needed to initialize the decoder and be able to decode media segments in the Representation.

For example, when the ITU-R Rec.H.264 | ISO/IEC 14496-10 codec is used, the initialization segment may be used to communicate SPS, and PPS elements of the bitstream.

5.1.1.4 GOP structure and Stream Access Points

In preparation of video segments, it is important to ensure that each Media Segment has at least one Stream Access Point (SAP). There are several different types of SAPs and corresponding encoding structures that can be employed.

5.1.1.4.1 SAP type 1

SAP type 1 is characterized by full alignment of time-domain parameters: $T_{EPT} = T_{DEC} = T_{SAP} = T_{PFT}$. It can be implemented by using video structure known as “Closed GoP” (Group of Pictures). An example of such a structure is shown in [Figure 7](#).

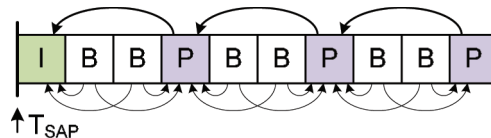


Figure 7 — Type 1 SAP implemented using Closed GOP structure

In this figure, boxes labeled as I, B, and P – correspond to I-(or IDR-), B-, and P-type pictures correspondingly. The I- (or IDR-) picture is independently decodable, while P-pictures require prior frames to be decoded first, and B-pictures require both prior and following I- or P- pictures to be decoded first.

In closed GoP structure, the I- (or IDR-) picture is transmitted first, allowing all subsequent pictures to be sequentially decoded. The first Access Unit in decoding order is also the first access unit in presentation order.

5.1.1.4.2 SAP type 2

SAP type 2 allows the presentation time of the first access unit to be delayed: $T_{EPT} = T_{DEC} = T_{SAP} < T_{PFT}$. This can be implemented by using modified Closed GoP structure, as illustrated in [Figure 8](#).

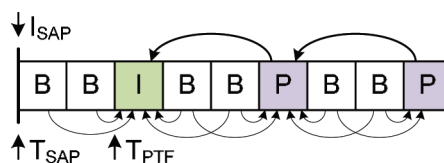


Figure 8 — Type 2 SAP implemented using modified Closed GOP structure

In this example, the first two pictures are backward predicted (syntactically they can be coded as forward-only B-pictures), and they both need 3rd picture to be decoded first. The 3rd picture is an I- (or IDR-) picture and it is transmitted in the first Access Unit.

5.1.1.4.3 SAP type 3

SAP type 3 imposes the following order on time-domain parameters: $T_{EPT} < T_{DEC} = T_{SAP} \leq T_{PFT}$. This can be implemented by using Open GoP structure, as illustrated in [Figure 9](#).

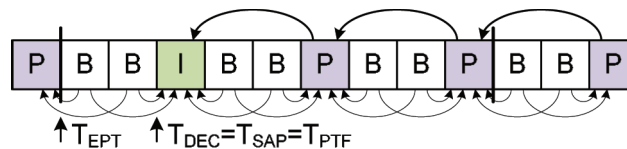


Figure 9 — Type 3 SAP implemented using Open GoP structure

In this example, Access Units corresponding to first 2 B-pictures cannot be decoded by using data within a segment, as they also rely on P-picture from a previous segment.

Open GoP structure has the advantage of being most efficient from coding efficiency standpoint. It can be used to implement random access in a single Representation. In order to enable switching between different Representations with Open GoP-based access points, such Representations should be prepared using identical codecs, resolutions, and prediction structures (codecs should have identical states of Decoded Picture Buffers).

5.1.2 Audio content generation

5.1.2.1 General

This section presents a set of recommended practices for preparing audio content for streaming when using MPEG audio codecs. It uses terms and definitions provided in ISO/IEC 13818-3, ISO/IEC 14496-3, ISO/IEC 23003-1, and ISO/IEC 23003-3.

5.1.2.2 Enabling bandwidth adaptation

In order to support bandwidth adaptation, audio content should be encoded at a plurality of rates, covering the range of operational rates of the network. Such encodings should be presented as Representations with different `@bandwidth` attributes, and presented in the MPD as an Adaptation Set.

Audio encoding characteristics such as codecs, sampling rates, and channel configurations should be communicated by means of `@codecs`, `@audioSamplingRate`, and `AudioChannelConfiguration` attributes and Elements of Representations and Adaptation Sets. Additional characteristics may be also signaled by means of `EssentialProperty` or `SupplementalProperty` descriptors with URNs and schemas defined in ISO/IEC 23001-8 or other relevant specifications. Unless declared otherwise, Media Segments carrying audio data should be understood as ones having SAP type 1, and where the first Access Unit serves as SAP.

5.1.2.3 Restrictions

To avoid discontinuities in the decoder output, the following parameters should be the same across all Representations in an Adaptation Set for a Segment:

- audio object type of the audio codec;
- channel configuration;
- sampling frequency.

Period Boundaries, however, signal a change (discontinuity) in content, e.g. ad insertion. At these boundaries, any codec reconfiguration is possible, including the above parameters.

5.1.2.4 Delay alignment

All bit streams should be delay adjusted. Encoder implementations may add additional delay depending on how they are configured (e.g. bit rate dependent Low Pass filter with varying tap count). The delay

for all configurations should be pre-compensated, so that all Segments in an Adaptation Set contain the same audio information and all streams have the same framing.

5.1.2.5 AAC-LC bitrate switching

When switching between different AAC-LC Streams, the following restrictions should be taken into account, to guarantee seamless switching between different AAC-LC bit streams.

5.1.2.5.1 Window type and Window sequence

To avoid artifacts due to time aliasing components not being cancelled, window type and window shape of the AAC-LC streams should be synchronized across all bit streams at the Stream Access Point (SAP). All audio streams should use the same overlap (either long or short) and window shape (either KBD or Sine) at the Segment boundary (right window half of frame preceding SAP and left window half of SAP frame).

5.1.2.6 HE-AAC bitrate switching

As HE-AAC consists of AAC-LC and SBR audio objects, all restrictions for AAC-LC should also apply for HE-AAC. Additional adaptations for the AAC-LC core are needed to avoid small spectral holes when switching between streams with different AAC/SBR crossover frequencies. Also for the SBR part, restrictions to some SBR tools are necessary to avoid artifacts for switching between different streams.

5.1.2.6.1 Additional restrictions for AAC-LC core

To avoid gaps in the frequency range between AAC/SBR crossover frequencies of the low bit rate and high bit rate stream, the AAC-LC core bandwidth of the last frame of a segment should match the highest AAC/SBR crossover frequency of all streams in an Adaptation Set. To properly encode the additional bandwidth, extra bits are necessary. The bit reservoir control should be adapted to compensate for this extra bit demand.

5.1.2.6.2 SBR header and Time-differential coding

In contrast to the AAC configuration that is completely described by the audio specific configuration (ASC), the SBR decoder needs additional configuration parameters. These parameters are transmitted inside the SBR Header which may not be contained in every access unit (AU). The MPEG-4 standard recommends a transmission interval of 500 ms or whenever an instantaneous change of header parameters is required (see ISO/IEC IS 14496-3:2009, 4.5.2.8.2.1).

To allow seamless switching of HE-AAC bit streams, it is necessary to transmit an SBR header in the first Access Unit of a Segment (SAP frame). As the MPEG-4 Audio Conformance (ISO/IEC IS 14496-26) forbids the use of tools that rely on preceding frames for frames containing an SBR Header, with this restriction it is also assured that the SAP frame can be completely decoded and processed.

5.1.2.6.3 SBR frame class

SBR envelopes can reach over frame borders, i.e. "VARVAR" and "FIXVAR" frames may overlap the SBR frame border. A SAP Frame should always start with a FIX border ("FIXVAR" or "FIXFIX") to make sure all necessary information to fully decode the audio contained in that frame is available. Consequently, the last frame in a segment (the frame before the SAP) should end with a "FIX" border (i.e. a "FIXFIX" or "VARFIX" frame).

5.1.2.7 HE-AACv2 bitrate switching

As HE-AACv2 relies on an HE-AAC core and adds the Parametric Stereo (PS) audio object, all requirements listed for AAC-LC and HE-AAC streams should also apply for HE-AACv2.

5.1.2.7.1 PS header and Time-differential coding

As with the SBR payload, also within the PS payload, Configuration Parameters may be transmitted not with every frame, but on a less regular basis. Also, time differential coding of certain parameters can be used to increase compression efficiency.

To allow for a seamless switching of HE-AACv2 bit streams, it is necessary to transmit a PS header with each SAP frame. For frames containing a PS Header, the MPEG-4 Audio Conformance forbids the use of tools that rely on past information, i.e. time differential coding of parameters.

With the above restriction, it is assured that the SAP frame can be completely decoded and processed. As HE-AACv2 Conformance requires a PS Header with every SBR header, this requirement is also implicitly inherited from the HE-AAC requirements.

5.1.2.7.2 PS tools and parameters

All PS Tools are designed to allow for continuous remapping of different configurations (e.g. frequency resolution of parameter bands). For the baseline version of the PS Tool, no extra care has to be taken at Stream Access Points.

5.1.2.8 AAC-LC/HE-AAC plus MPEG Surround bitrate switching

As MPEG Surround is based on an AAC-LC or HE-AAC core coder, all restrictions for AAC-LC and HE-AAC should also apply for MPEG Surround. Further restrictions are as follows, for details see ISO/IEC 23003-1.

The MPEG Surround data shall be conveyed in the AAC extension payload using implicit signaling. Each SAP frame shall contain the syntactic element *SpatialSpecificConfig* that contains the MPEG Surround configuration data. Additionally, the coding of SAP frames should be independent of previous frames. Therefore, the bitstream payload element *bsIndependencyFlag* should be set to 1.

The MPEG Surround tool *residual coding* employs a representation of differential signals using the AAC-LC syntax. As described in [5.1.2.5.1](#), the window type of the residual signal should be synchronized at the SAP.

5.1.3 Content preparation for live streaming

The DASH media presentation description and encoded segments should be prepared in conformance to timing model described in [6.4](#).

5.1.4 Guidelines for generation of segment file names

5.1.4.1 General

In preparation of DASH content for distribution, it is important to ensure that encoded segments are given unique names, allowing their storage in same location and referencing from an MPD file.

This section provides recommended process for generation of file names, based on mapping of content parameters to file names. This process includes mapping of static content parameters, as well as dynamic parameters, such as \$number\$ and \$time\$ enabled by **SegmentTemplate** elements.

5.1.4.2 Segment URL generation

Mapping of content parameters to URLs is accomplished by using a set of key/value pairs provided in [Table 1](#). Such key/value pairs shall only be used right of the rightmost character “/”. The first character after “/” does not represent a key. The first key shall only be present right of the first ‘_’ which is right of the rightmost character “/”. The ordering of the key/value pairs in the string is arbitrary. The key/value pairs shall only use delimiters “__”. No delimiter is added between the key and the value. Note that the syntax and semantic is compatible with ISO/IEC 23009-1 and only provides an additional restriction to the exact syntax of the HTTP URL.