

INTERNATIONAL
STANDARD

ISO/IEC
23001-13

First edition
2019-07

**Information technology — MPEG
systems technologies —**

**Part 13:
Media orchestration**



Reference number
ISO/IEC 23001-13:2019(E)

AVA Documents
Document Sharing Hub
© ISO/IEC 2019



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2019

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
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword	v
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms and definitions	2
4 Abbreviated terms, relationships between terms and mnemonics	5
4.1 Abbreviated terms	5
4.2 Relationships between defined terms	5
4.3 Mnemonics	6
5 Architecture	6
5.1 General	6
5.2 Architecture for temporal orchestration	9
5.2.1 General	9
5.2.2 Sink-side architecture for temporal orchestration	10
5.2.3 Source-side architecture for temporal orchestration	13
5.3 Architecture for spatial orchestration	15
5.3.1 General	15
5.3.2 M-processor side architecture for spatial orchestration	15
5.3.3 Sink-side architecture for spatial orchestration	17
6 Messaging and control	18
6.1 Wall clock	18
6.2 MORE communication channel	18
6.2.1 General	18
6.2.2 Signalling of the MORE communication channel	19
6.3 Websocket protocol	20
7 Metadata	20
7.1 Timed metadata	20
7.2 Position metadata	21
7.2.1 General	21
7.2.2 Global position	21
7.2.3 Altitude	21
7.2.4 Relative position	21
7.3 Orientation metadata	22
7.4 Regions of interest	23
7.4.1 General	23
7.4.2 Example implementation using DROP features	24
7.5 Quality metadata	25
7.5.1 General	25
7.5.2 Quality metadata	27
7.6 Stream monitor	27
7.6.1 General	27
7.6.2 MPEG-Audio-Sync-based stream monitor metadata	30
7.7 Capture mask	30
7.7.1 General	30
7.7.2 Capture-mask metadata	31
7.8 Camera metadata	32
7.9 Depth metadata	32
8 Transport	33
8.1 Carriage of timed metadata in ISO base media file format	33
8.1.1 General	33
8.1.2 Carriage of position metadata in ISO base media file format	34

8.1.3	Carriage of orientation metadata in ISO base media file format.....	35
8.1.4	Carriage of quality metadata in ISO base media file format.....	36
8.1.5	Carriage of audio features in ISO base media file format.....	37
8.1.6	Carriage of capture-mask metadata in ISO base media file format.....	38
8.1.7	Carriage of dvb-css correlation timestamps in ISO base media file format.....	38
8.2	Carriage of timed metadata in MPEG-2 systems.....	41
8.2.1	General.....	41
8.2.2	Timed metadata access unit.....	41
8.2.3	Timed metadata extension descriptor.....	42
8.2.4	Carriage of position metadata in MPEG-2 systems.....	43
8.2.5	Carriage of orientation metadata in MPEG-2 systems.....	46
8.2.6	Carriage of capture-mask metadata in MPEG-2 systems.....	47
8.2.7	Carriage of audio features in MPEG-2 systems.....	49
8.2.8	Carriage of quality metadata in MPEG-2 systems.....	50
8.2.9	Carriage of correlation timestamps in MPEG-2 systems.....	50
8.3	Carriage of MORE information in MMT.....	51
8.3.1	General.....	51
8.3.2	Carriage of MORE information in MMT.....	51
8.4	Carriage of MORE information in DASH.....	51
8.4.1	General.....	51
8.4.2	Carriage of quality metadata in MPEG DASH MPD.....	51
9	Orchestration data	51
9.1	General.....	51
9.2	Timeline synchronization.....	52
9.3	dvb-css correlation timestamps.....	52
	Annex A (normative) Source-side timeline synchronization	53
	Annex B (informative) Object model	56
	Annex C (informative) Control architecture	59
	Bibliography	64

Foreword

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

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

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) or the IEC list of patent declarations received (see <http://patents.iec.ch>).

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.

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

A list of all parts in the ISO/IEC 23001 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.

Introduction

Audiovisual equipment is pervasive; everyone with a smartphone, a tablet or a laptop has both recording and display equipment at their disposal, usually connected to a local or a public network. This equipment is increasingly sophisticated, with higher resolutions and better lenses (for video), often multiple microphones (for audio), coupled with other sensors (e.g. for geolocation) and increasingly sophisticated processing capabilities. These devices can not only decode in real time, but usually also perform decent encoding. Sensors and displays come in the form of personal smart phones, but also include smart watches, omnidirectional cameras and a variety of virtual reality and augmented reality head-mounted devices. All these devices can either be a source of multimedia content (audio, video or audiovisual) or consume such content. Often, devices play both roles.

The proliferation of these multimedia-capable devices combined with ever-increasing bandwidth, including mobile bandwidth, necessitates better and standardized mechanisms for coordinating such devices, the associated media streams and the available resources, like media processing and transmission. This process is called “orchestration”. Orchestration includes coordination in time (synchronization) and in space, as well as the coordination of computational resources to perform the coordination functions.

This document was developed to address the need for specifying this coordination between capture devices (“sources”) and play-back devices (“sinks”) in a heterogeneous environment. In this context, “heterogeneous” refers to the fact that devices can be of different nature (e.g. a mix of consumer and professional devices) and on different networks.

The International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that compliance with this document may involve the use of patents.

ISO and IEC take no position concerning the evidence, validity and scope of these patent rights. The holders of these patent rights have assured ISO and IEC that they are willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statements of the holders of these patent rights are registered with ISO and IEC. Information may be obtained from:

Electronics and Telecommunications Research Institute

218 Gajeong-ro, Yuseong-gu, Daejeon, 34129 Korea

Koninklijke KPN N.V.

Wilhelminakade 123, 3072 AP Rotterdam, The Netherlands

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

Information technology — MPEG systems technologies —

Part 13:

Media orchestration

1 Scope

This document specifies an architecture for media orchestration, as well as associated messaging and control, timed metadata, the carriage of that timed metadata, and orchestration data.

2 Normative references

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

ISO/IEC 13818-1:2019, *Information technology — Generic coding of moving pictures and associated audio information — Part 1: Systems*

ISO/IEC 14496-3:2009, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-3:2009/Amd 5:2015, *Information technology — Coding of audio-visual objects — Part 3: Audio — Amendment 5: Support for Dynamic Range Control, New Levels for ALS Profile, and Audio Synchronization*

ISO/IEC 14496-12:2015, *ISO Base Media File Format*

ISO/IEC 14496-12:2015/Amd 1:2017, *ISO Base Media File Format — Amendment 1: DRC extensions*

ISO/IEC 23001-10:—¹⁾, *Information technology — MPEG systems technologies — Part 10: Carriage of timed metadata metrics of media in ISO base media file format*

ISO/IEC 23005-6:2019, *Information technology — Media context and control — Part 6: Common types and tools*

ISO/IEC 23008-1, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 1: MPEG media transport (MMT)*

ETSI TS 103 286 02:2017, *Digital Video Broadcasting (DVB); Companion Screens and Streams; Part 2: Content Identification and Media Synchronization* [online]. Available at http://www.etsi.org/deliver/etsi_ts/103200_103299/10328602/

IETF RFC 3986, *Uniform Resource Identifier (URI): Generic Syntax* [online]. Edited by T. Berners-Lee. January 2005. Available at <https://www.ietf.org/rfc/rfc3986.txt>

IETF RFC 6455:2011, *The WebSocket Protocol* [online]. Edited by I. Fette. December 2011. Available at <https://www.ietf.org/rfc/rfc6455.txt>

IETF RFC 5905, *Network Time Protocol Version 4: Protocol and Algorithms Specification* [online]. Edited by D. Mills. June 2010. Available at <https://tools.ietf.org/html/rfc5905>

1) Under preparation. Stage at the time of publication: ISO/IEC DIS 23001-10:2019.

3 Terms and definitions

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

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

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

3.1 content

media data

data (3.3) that can be rendered, including audio, video, text, graphics, images, haptic and tactile information

Note 1 to entry: This data can be timed or non-timed.

3.2 controller

elementary function that controls one or more other elements, outputs *control data* (3.5) and can receive control data

3.3 data

bytes of information that are carried, processed or stored

3.4 timed data

data (3.3) that has an intrinsic *timeline* (3.25)

3.5 control data

data (3.3) exchanged for messaging and control

3.6 metadata

data (3.3) about other data that cannot be rendered independently and can affect rendering, processing or orchestration of the associated *media data* (3.1)

3.7 timed metadata

metadata (3.6) that has an intrinsic *timeline* (3.25)

3.8 orchestration data

data (3.3) that orchestrates any number of *timed data* (3.4) streams

Note 1 to entry: Orchestration data itself can also be timed data.

3.9 device

physical entity that can embody a number of elementary functions

3.10 m-processor

media processor

elementary function that can process *media data* (3.1) and/or *metadata* (3.6), that can receive media data and/or metadata, that can optionally receive *orchestration data* (3.8), that can output media data, metadata or both, and that can optionally exchange *control data* (3.5)

3.11**media orchestration**

orchestration of media and *metadata* (3.6) capture, processing and presentation involving multiple *devices* (3.9)

3.12**media orchestration session**

coherent application context that includes media capture and/or media processing and/or media rendering

3.13**media stream**

timed *media data* (3.1)

3.14**mos more**

mean opinion score of the subjective quality of a piece of timed *media data* (3.1)

3.15**sink**

elementary function that can receive *media data* (3.1), and optionally *orchestration data* (3.8) and/or *metadata* (3.6), can present the media data, can optionally output metadata and can optionally exchange *control data* (3.5)

3.16**source**

elementary function that can output *media data* (3.1) and/or *metadata* (3.6) and can optionally exchange *control data* (3.5)

3.17**timestamp**

number that is used to identify an audio sample, video frame or *metadata* (3.6) sample

EXAMPLE MPEG-2 TS PTS (presentation timestamp) or ISOBMFF CT (composition time).

Note 1 to entry: The meaning of the term “timestamp” depends on the context. Within the MPEG context, the term is usually used as a single number to identify a point on the timeline of an MPEG-2 transport stream or within an ISOBMFF file. In the context of DVB-CSS (ETSI TS 103 286 02), the term “timestamp” is used to provide an association between a wall clock time and a point on a certain media timeline. This document uses the “dvb-css” prefix to flag where the DVB-CSS (ETSI TS 103 286 02) notion of timestamp is used.

3.18**dvb-css timestamp**

pair of two values each of which represents a time value on a timeline such that the two-time values correspond to the same moment in time

[SOURCE: ETSI TS 103 286 02:2017, 3.1]

3.19**dvb-css control timestamp**

dvb-css timestamp (3.18) representing the time (in relation to the wall clock) at which a *synchronization client* (3.24) is recommended to present a particular point in time (in relation to the synchronization timeline (3.25)) of its *timed data* (3.4)

[SOURCE: ETSI TS 103 286 02:2017, 3.1, modified — «timed content» has been replaced with «timed data» and the note has been deleted]

3.20

dvb-css correlation timestamp

time values on two or more *timelines* (3.25) that correlate with each other

[SOURCE: ETSI TS 103 286 02:2017, 3.1, modified — «ore more” has been added and the note has been deleted]

3.21

dvb-css earliest presentation timestamp

dvb-css timestamp (3.18) representing the earliest time (in relation to the wall clock) at which a *synchronization client* (3.24) believes it can present a particular point in time of its *timed data* (3.4)

[SOURCE: ETSI TS 103 286 02:2017, 3.1, modified — “(in relation to the Synchronization Timeline) of its Timed Content” has been replaced with “of its timed data” and the note has been deleted]

3.22

user

human that uses one or more *device(s)* (3.9) and/or service(s)

Note 1 to entry: A user can be a producer and consumer of media, and both at the same time.

3.23

media synchronization application server

elementary function that coordinates the process of obtaining a shared agreement on the progress of *timelines* (3.25) among all *synchronization clients* (3.24)

Note 1 to entry: This is for the purpose of enabling the synchronization clients to present timed data simultaneously with respect to each other.

[SOURCE: ETSI TS 103 286 02:2017, 3.1]

3.24

synchronization client

elementary function that aligns its presentation of *timed data* (3.4) with other synchronization clients by communicating with a *media synchronization application server* (3.23)

[SOURCE: ETSI TS 103 286 02:2017, 3.1, modified — “wishes to align” has been replaced with “aligns” and “timed content” has been replaced with “timed data”]

3.25

timeline

reference frame for describing time, represented as a linear scale against which time can be measured for a particular system

Note 1 to entry: This could manifest in various ways, such as: as a local oscillator, the progress of a broadcast or the time position within an item of media content.

[SOURCE: ETSI TS 103 286 02:2017, 3.1]

3.26

timeline correlation

correlation between two *timelines* (3.25)

[SOURCE: ETSI TS 103 286 02:2017, 3.1]

3.27

wall clock

linear monotonic clock that is not assumed to represent real date and time

Note 1 to entry: The wall clock is intended for sharing between two or more entities for the purposes of having a common synchronized time reference frame.

Note 2 to entry: DVB-CSS (ETSI TS 103 286 02) wall clock is NOT a “UTC-type date&time”. It has a random offset and it has no leap seconds or leap years. It is a clock that ticks like a metronome, and it has only a local meaning in the context of media synchronization between devices. This is a different notion for a timestamp than what is commonly used across MPEG. This document uses the definition from ETSI TS 103 286 02 to maintain consistency with that specification.

[SOURCE: ETSI TS 103 286 02:2017, 3.1, modified — Note 2 to entry has been added]

3.28

wall clock synchronization

achieving a shared agreement on the progress of time of a *wall clock* ([3.27](#))

Note 1 to entry: This is a process that takes place between devices.

[SOURCE: ETSI TS 103 286 02:2017, 3.1]

4 Abbreviated terms, relationships between terms and mnemonics

4.1 Abbreviated terms

The following abbreviated terms are used in this document:

CSA	Companion screen application
CT	Composition time
DROP	Distinctive region or pattern
GPS	Global positioning system
MORE	Media orchestration
MORE-TS	MORE timeline synchronization (interface)
MORE-WC	MORE wall clock (interface)
MSAS	Media synchronization application server
URL	Uniform resource locator
PTS	Presentation timestamp
SC	Synchronization client
SNR	Signal-to-noise ratio
WCC	Wall clock client
WCS	Wall clock server

4.2 Relationships between defined terms

[Figure 1](#) sketches relationships between some of the defined terms, including synonyms.

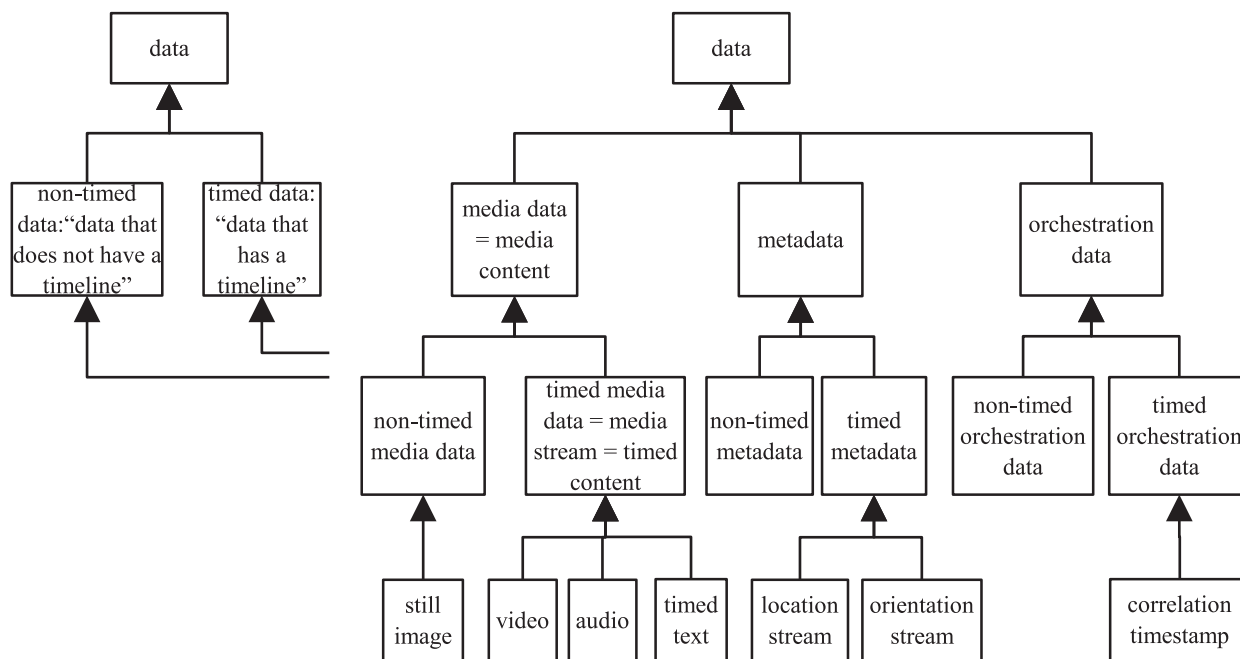


Figure 1 — Relationships between defined terms

4.3 Mnemonics

rpchof	remainder polynomial coefficients, highest order first
bslbf	bit string, left bit first
fsbf	float, sign bit first
uimsbf	unsigned integer, most significant bit first

5 Architecture

5.1 General

The term “media orchestration” denotes the orchestration of media and metadata capture, presentation and processing involving multiple devices.

Orchestration of media capture refers to metadata and control signalling to determine which device captures what, when and how. **What** to capture is about what media to capture and which part of that media should be captured. **When** to capture is about capture synchronization with other devices, as well as start and stop of capture. **How** to capture is about location, orientation, capture capabilities, frame rate, resolution, microphone gain, white balance settings, etc. as well as codecs used, metadata delivered and possible processing to be applied.

Orchestration of media presentation refers to metadata and control signalling to determine which device presents what, when and how. **What** to present is about what media to retrieve and which parts of that media should be presented. **When** to present is about presentation synchronization with other devices. **How** to present is about where exactly to play out something, e.g. positioning of a media part in a screen, positioning of an audio object in a room and possible processing to be applied.

Orchestration of processing refers to metadata and control signalling for applying processing to sets of captured media and/or metadata. This includes single-media processing (e.g. media synchronization in case of transcoding), as well as processing of multiple media and/or metadata together, e.g. performing

video stitching, changing arrangements of media in space and time, or automated editing and selection processes.

A **media orchestration session** is a coherent application context that includes media capture and/or media rendering and/or media processing. It is defined by its scope: a concert lasting one evening in one concert hall, a week-long festival in a park, the vacations of a family in Norway over a month, the synchronized presentation of media across multiple screens, the reconstruction of a crime captured over many different cameras...

Media orchestration is independent from media type or encoding or packaging or distribution.

DVB CSS introduces the concept of timed data, see TS 103 286 02, 5.2.1. DVB uses the term “timed content” for this concept. Timed data has an intrinsic timeline. It may have a start and/or end (e.g. content on-demand) or it may be continuous (broadcast). It may be atomic (video-only) or it may be composite (A/V, 3D-audio, multiplex). Classic examples of timed data are video, audio and timed text.

In the context of media orchestration, streams of location and orientation as well as other timed sensor outputs are also timed data; see [Figure 2](#). This can also be called timed metadata, as defined in DVB CSS.

NOTE 1 The notion of time has various applications in media orchestration. It can pertain to the *delivery* or to the media itself, or the *presentation*. It can also pertain to the *capture* of media.

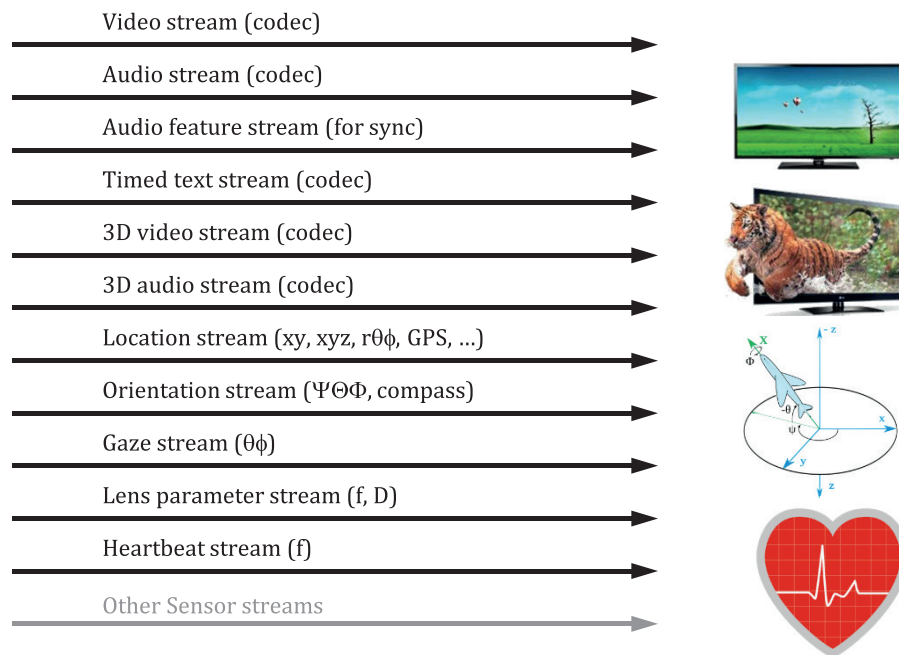


Figure 2 — Examples of timed data in the context of media orchestration

This document distinguishes a number of functional entities for media orchestration, all derived from the same abstract entity. It first describes the entities and then gives an example of how they can be configured for an actual use case.

[Table 1](#) shows the meaning of the arrows used in [Figures 3](#) to [6](#).

Table 1 — Types of data

	media data (timed or non-timed, unidirectional)
	metadata (timed or non-timed, unidirectional)
	orchestration data (timed or non-timed, unidirectional)

Table 1 (continued)

	control data for messaging and control (non-timed, possibly bidirectional)
---	--

Figures 3 to 6 illustrate the concepts of source, sink, m-processor and controller, respectively, as defined in Clause 3.

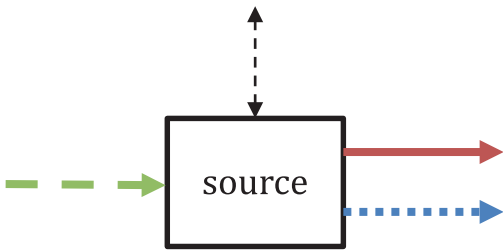


Figure 3 — Source

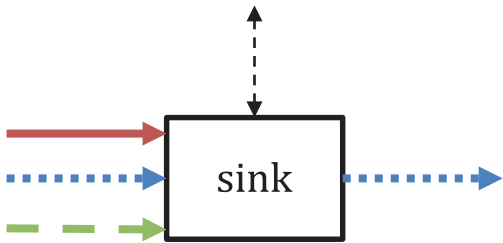


Figure 4 — Sink

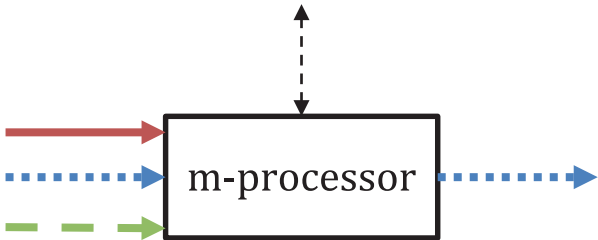


Figure 5 — m-processor

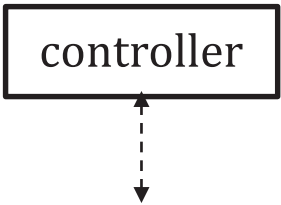


Figure 6 — Controller

A device may have one or more sources. Each source may generate timed media data or timed metadata. When an m-processor is located within the same device, the device can also generate orchestration data.

The timed media data and metadata generated from sources can be delivered to other functional elements, i.e. m-processor and sink. The timed metadata may also be multiplexed with associated media data.

NOTE 2 The following applies:

- a user device can (and very often will) have both source and sink functions;
- an application server can combine controller and m-processing functions;
- m-processing function can include transcoding, adding or changing timelines, multiplexing, demultiplexing, selection (editor), stitching, tiling (e.g. using MPEG DASH spatial relationship description), translation, extracting timed metadata (e.g. CDVS metadata) and more.

An example instantiation of this model is as follows:

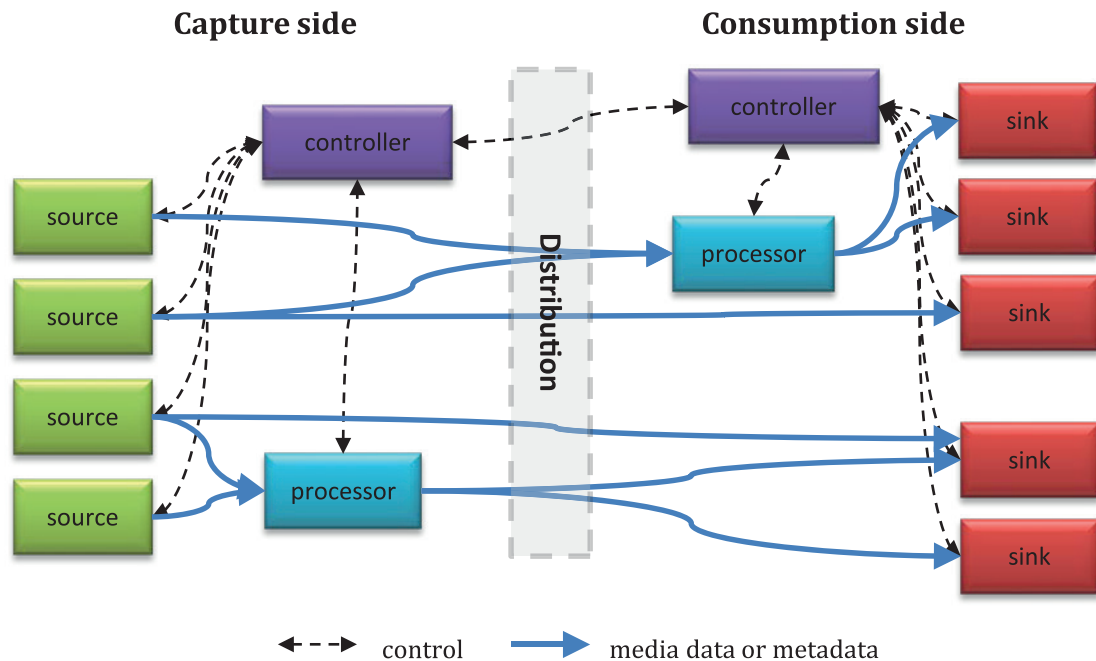


Figure 7 — Examples of timed data in the context of media orchestration

This model shows a complex scenario where many sources contribute to an experience with many orchestrated sinks.

[Figure 7](#) is broadcast-oriented, with a clear distinction between the production side and the consumption side. The architecture also supports mixed cases, for example shared experiences, where sinks are used for broadcast media consumption in orchestration with sources and sinks used for communication between users.

[Annexes B](#) and [C](#) provides illustrations of control information and its exchanges.

5.2 Architecture for temporal orchestration

5.2.1 General

Temporal orchestration, or media synchronization, is the process of time alignment of timed media data and metadata. When media data and metadata are captured at different source devices, their timelines will generally be different. Audio samples, video frames and metadata samples are dvb-css timestamped against the clock of the source device. Such dvb-css timestamps identify audio samples, video frames and metadata samples by their presentation timestamp (PTS, MPEG-2 transport stream), composition time (CT, MPEG-4 ISOBMFF) or external timeline (e.g. TEMI, MPEG-2 timeline for external data). Clocks (“wall clocks”) of source devices are typically free running and have clock skew, clock drift and clock drift variation. Moreover, transcoders and multiplexers may change the time base of timed media data and metadata. Finally, there are typically significant differential delays between distribution chains

to different sink devices. As a consequence, timed media data and metadata can arrive severely out of sync at sink devices, and temporal orchestration is needed to get them synchronized again.

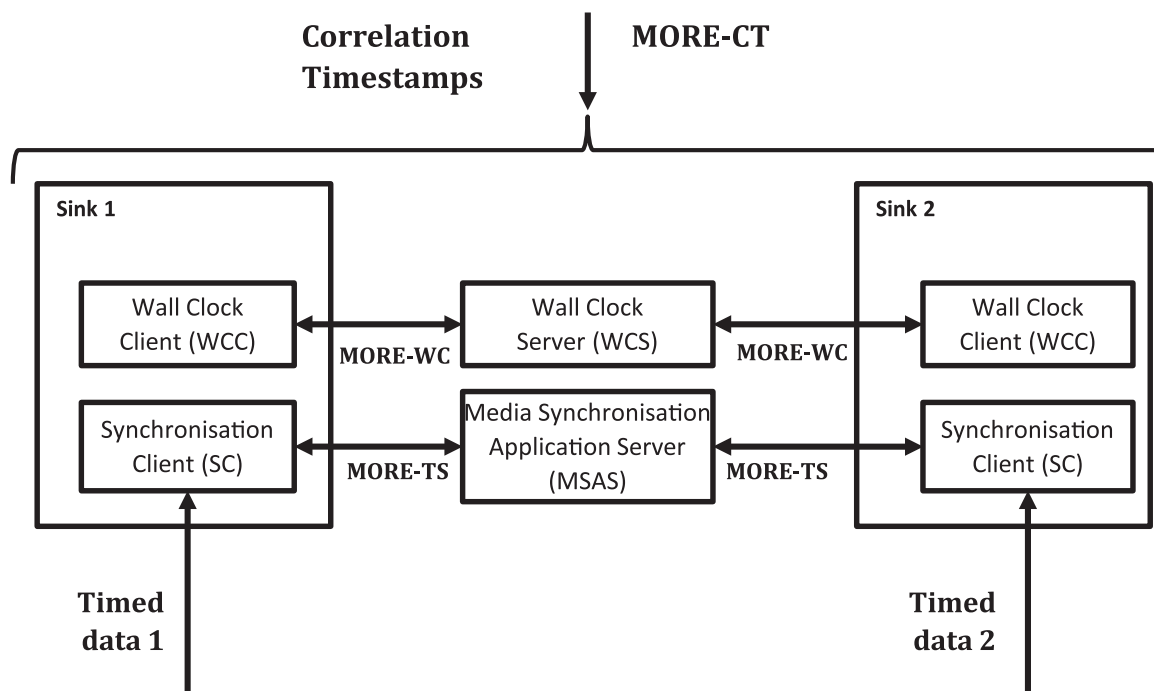
NOTE Clock skew is an offset. Clock drift is where two clocks have a nominal relationship to each other that is not observed in practice (e.g. an audio clock that claims to be 48 KHz but does not actually sample at that rate). Clock drift variation is when the actual sampling rate is not only off, but also not constant.

The introduction clause of ETSI TS 103 286 02:2017 provides a useful introduction into sink-side media synchronization, including use cases, the need for content identification functionality and the need for playback synchronization functionality, as well as issues with timelines and a conceptual model for time-controlled playback.

[5.2.2](#) and [5.2.3](#) provide architectures for temporal orchestration at the sink and source sides, respectively.

5.2.2 Sink-side architecture for temporal orchestration

[Figure 8](#) provides a sink-side architecture for temporal orchestration, based on ETSI TS 103 286 02. The purpose of this architecture is to synchronize two or more pieces of incoming timed data (timed media, timed metadata or timed orchestration data) within/between one or more sinks.



SOURCE: ETSI TS 103 286 02, reproduced with the permission of DVB

Figure 8 — Sink-side architecture for temporal orchestration

The sink-side architecture has the following elements.

- There are sinks, each of which receives pieces of timed data. Each of those pieces of timed data can be timed media data, timed metadata or timed orchestration data.
- Each sink has a synchronization client (SC) per received piece of timed data (stream), which takes care of the synchronization of the timed data, delaying the presentation/use of the timed data when appropriate.
- An SC coordinates the timeline synchronization of its timed data with a media synchronization application server (MSAS) via the MORE-TS interface.

- Each sink has a wall clock client (WCC), which provides the wall clock time that is used for the dvb-css timestamping in the MORE-TS interface.
- A WCC synchronizes its wall clock against a wall clock server (WCS) via the MORE-WC interface.
- dvb-css correlation timestamps are a form of orchestration data that may be received and processed by SCs and/or MSAS via the MORE-CT interface; see also 9.3. dvb-css correlation timestamps are used to perform timeline translations, in the generic case where the timelines of different received timed data are different.

NOTE 1 Pieces of timed data (e.g. media streams) usually have different timelines, as each source typically has its own local oscillator. A common example is the case where timed data 2 is a transcoded (and remultiplexed) version of timed data 1.

Figure 9 shows a simplified case of the above architecture. This simplified case corresponds with the ETSI TS 103 286 02 inter-device media synchronization standard. The following applies to this example:

- A television device integrates WCC, WCS, SC and MSAS functions.
- Only timed media data is considered, not timed metadata or timed orchestration data.
- The timed media data at the companion screen application (CSA) is enslaved to the timed media data received by the TV device, ensuring that the playout of the timed media data at the TV device is never delayed.
- dvb-css correlation timestamps are received and processed by the CSA (design choice by ETSI TS 103 286 02).
- Timed media data are presented synchronized between TV device and CSA.

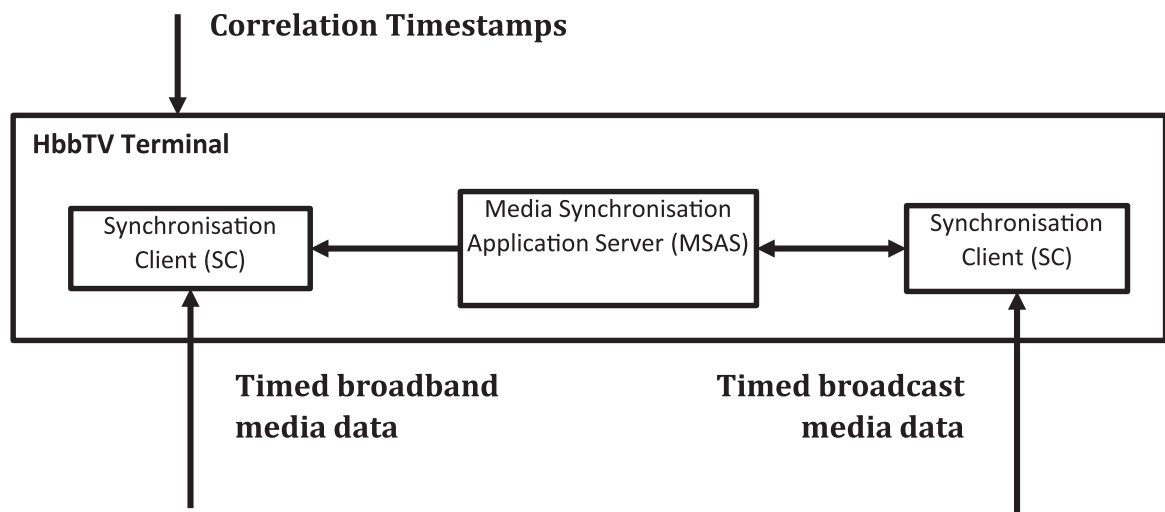


Figure 9 — Simplified case of the sink-side architecture, corresponding to HbbTV 2.0 ETSI TS 102 796

Figure 10 illustrates temporal orchestration at the sink side.

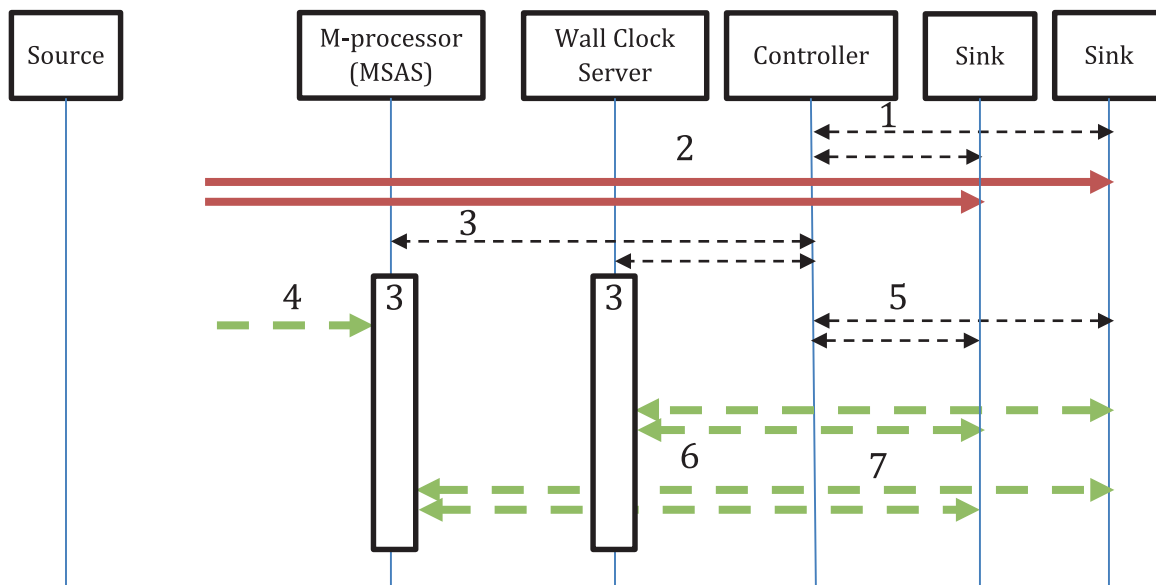


Figure 10 — Orchestration of timeline synchronization at the sink side

The sequence diagram of [Figure 10](#) has the following steps.

1. Discovery/registration processes, during which one or more sinks are associated with a controller for a specific media orchestration session.
2. The sinks start receiving timed media data and/or timed metadata. The sources of these timed media data and/or timed metadata have independent and free-running (not genlocked) clocks, hence necessitating sink-side timeline synchronization.
3. The controller initiates instances of a wall clock server and a media synchronization application server.

NOTE 2 The control signalling for this is out of scope of this document.

4. The media synchronization application server starts receiving dvb-css correlation timestamps.

NOTE 3 This example assumes that it is the MSAS that calculates or processes the dvb-css correlation timestamps. In other examples, it can be the sinks that process the dvb-css correlation timestamps, e.g. TV-to-companion synchronization in DVB-CSS ETSI TS 103 286 02.

5. The controller provides the sinks with the end-point addresses of the wall clock server and the media synchronization application server; see also below.
6. The sinks synchronize their wall clocks with the wall clock server using the wall clock protocol specified in [6.1](#).
7. The sinks start exchanging dvb-css presentation timestamps and dvb-css control timestamps with the media synchronization application server using the sink-side timeline synchronization protocol specified in [9.1](#). The media synchronization application server uses the dvb-css correlation timestamps to translate dvb-css presentation timestamps between the time bases (timelines) of the different timed media data and/or timed metadata and calculate dvb-css control timestamps, which sinks may use to delay and achieve synchronous playback.

NOTE 4 The source is drawn only for illustration purposes. It is not involved in this orchestration example.

The wall clock server end-point address is represented as a URL of the form "udp://<ip-address>:<port-number>", where <ip-address> and <port-number> are the IP address and port number, at which the wall clock server is providing its service endpoint; see also ETSI TS 103 286 02:2017, 8.4.

The media synchronization application server end-point address is a websocket URI as specified in IETF RFC 6455:2011, Clause 3, at which the media synchronization application server is providing its service endpoint; see also ETSI TS 103 286 02:2017, 9.3.

5.2.3 Source-side architecture for temporal orchestration

[Figure 11](#) provides a source-side architecture for temporal orchestration. It is a mirror image of the sink-side architecture of [Figure 8](#). The purpose of this architecture is to enable the future synchronization of two or more pieces of timed data (timed media, timed metadata or timed orchestration data) that are produced by one or more sources. The word "future" implies that the actual synchronization takes place at playout with the sink(s). To enable this, sources report about the production timing of their produced timed data.

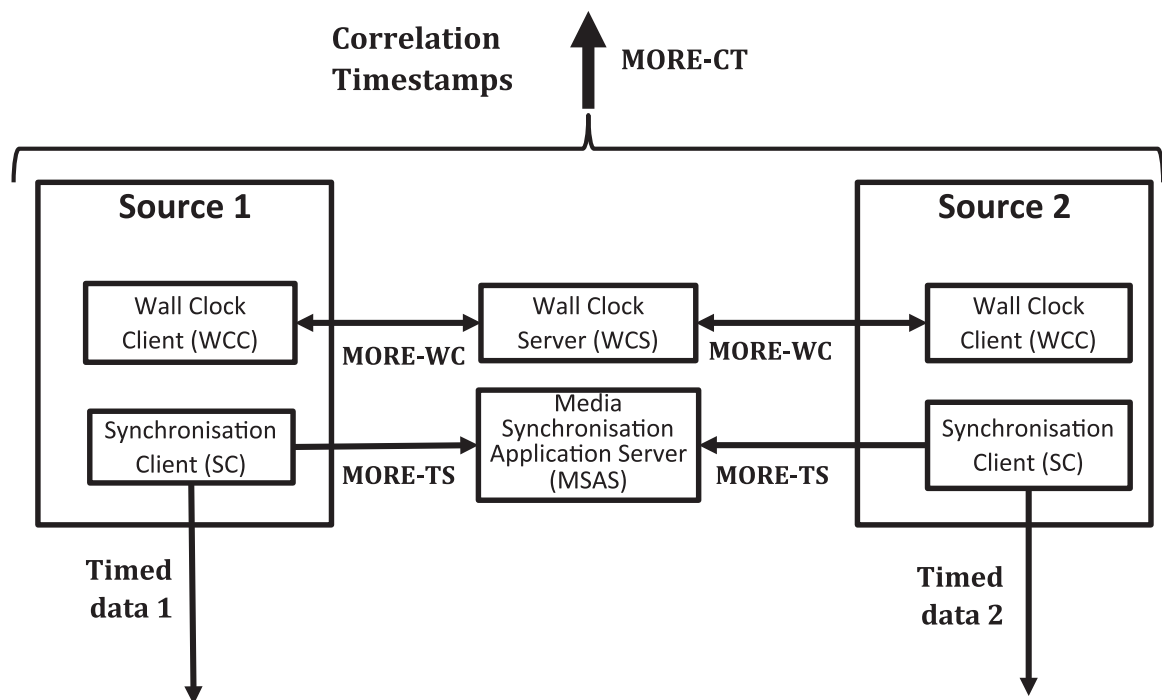


Figure 11 — Source-side architecture for temporal orchestration

The source-side architecture has elements similar to the sink-side architecture, with the following differences.

- The direction of timed data (timed media data, timed metadata and timed orchestration data) is outgoing.
- The SCs report to the MSAS, but they do not need to delay the produced timed data.
- The dvb-css correlation timestamps (MORE-CT interface, see also [9.3](#)) are outgoing, to be used for the future synchronization of timed data at the sink side.

NOTE 1 A media-orchestrated system can have multiple independent wall clock servers (WCS) and media synchronization application servers (MSAS). For example, an MSAS used at the source/production side can be different from an MSAS used at the sink/consumption side.

Simplified cases are possible, similar to the ones described for the sink-side architecture; for example, WCS and MSAS may be integrated with a source.

A single source device may integrate multiple sources, producing multiple pieces of timed data (streams).

Figure 12 provides an example of the simplified case that integrates a source and a sink function. This example is an m-processor (e.g. a transcoder or a CDN node) that remultiplexes and possibly delays the timed media data, resulting in that the timeline for the incoming timed media data is different than for outgoing timed media data. The m-processor produces dvb-css correlation timestamps in order to enable future synchronization between the two pieces of timed media data.

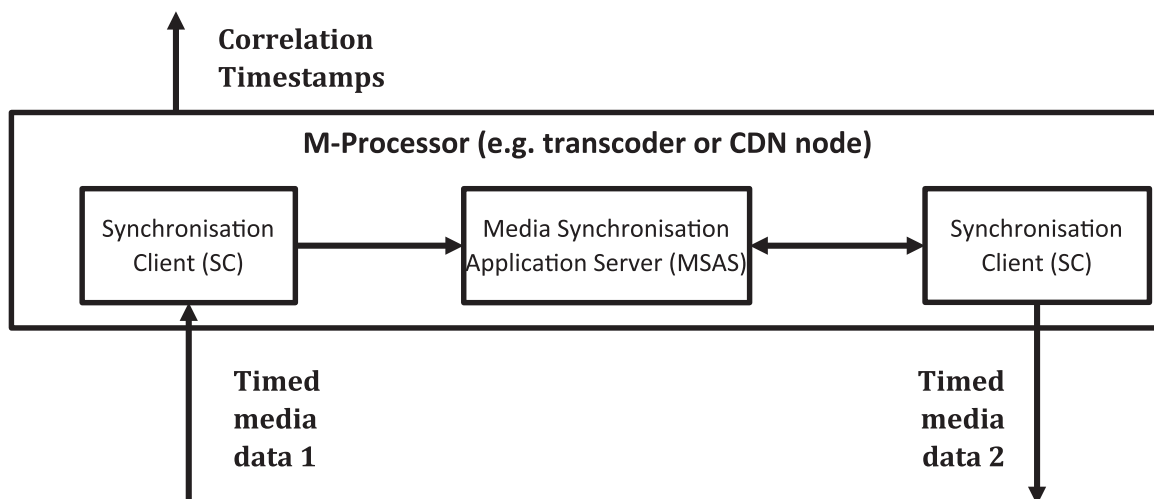


Figure 12 — Simplified source-side architecture for temporal orchestration

Figure 13 illustrates temporal orchestration at the source side.

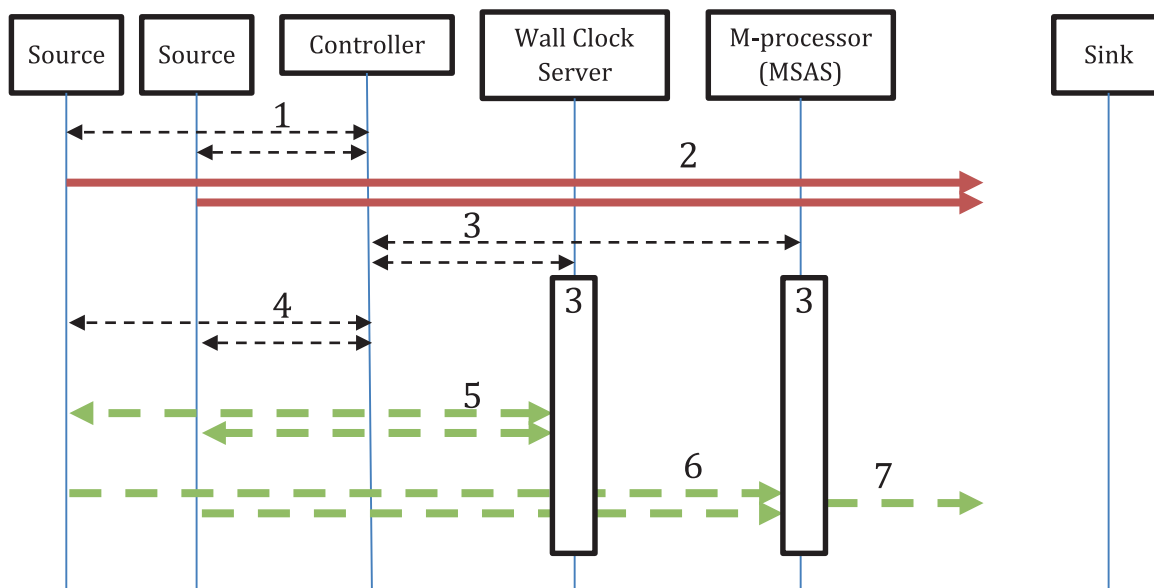


Figure 13 — Orchestration of timeline synchronization at the source side

The sequence diagram of Figure 13 has the following steps.

1. Discovery/registration processes, during which one or more sources are associated with a controller for a specific media orchestration session.
2. The sources start sending timed media data and/or timed metadata. The sources have independent and free-running (not genlocked) clocks, hence necessitating source-side timeline synchronization.
3. The controller initiates instances of a wall clock server and a media synchronization application server.

NOTE 2 The control signalling for this is out of scope of this document.

4. The controller provides the sources with the end-point addresses of the wall clock server and the media synchronization application server, see also below.
5. The sources synchronize their wall clocks with the wall clock server using the wall clock protocol specified in [6.1](#).
6. The sources start sending dvb-css capture timestamps to the media synchronization application server using the source-side timeline synchronization protocol specified in [9.2](#) and [Annex A](#).
7. The media synchronization application server starts calculating and sending dvb-css correlation timestamps based on the received dvb-css capture timestamps.

NOTE 3 The sink is drawn only for illustration purposes. It is not involved in this orchestration example.

5.3 Architecture for spatial orchestration

5.3.1 General

Spatial orchestration is the process of recognizing the spatial relationships between various media data (or elements like source, sink) using various types of metadata, and describing these relationships using orchestration data. There are various sources and sinks in a MORE session. Those sources capture something and send captured media data to m-processor. The m-processor needs to know which media data is related to what media data and how they are related. These relation data can be achieved with various metadata, such as GPS, position, orientation, camera parameters, image features, depth map, etc. Those relation data can be spatial orchestration data and spatial orchestration data can be helpful to carry out some processes of m-processor, such as stitching, tilting, etc. and furthermore, video summarization, user selection, generating VR contents, etc. On the sink side, when media data are consumed by sinks, sinks can communicate with m-processor with various types of metadata and receive spatial orchestration data for more interactive and immersive services.

5.3.2 M-processor side architecture for spatial orchestration

[Figure 14](#) provides an m-processor side architecture for spatial orchestration. The purpose of this architecture is letting m-processor to gather media data with associated metadata from sources, recognise relationship between media data or sources, and generate orchestration data. Those metadata can be orientation (gaze) data, capture mask, absolute and/or relative locations of sources, etc. Also, m-processor can generate metadata which can be helpful to recognise relations between media data, such as image features, depth map, etc., by itself. All of those metadata can be used for spatial orchestration and generating spatial orchestration data.

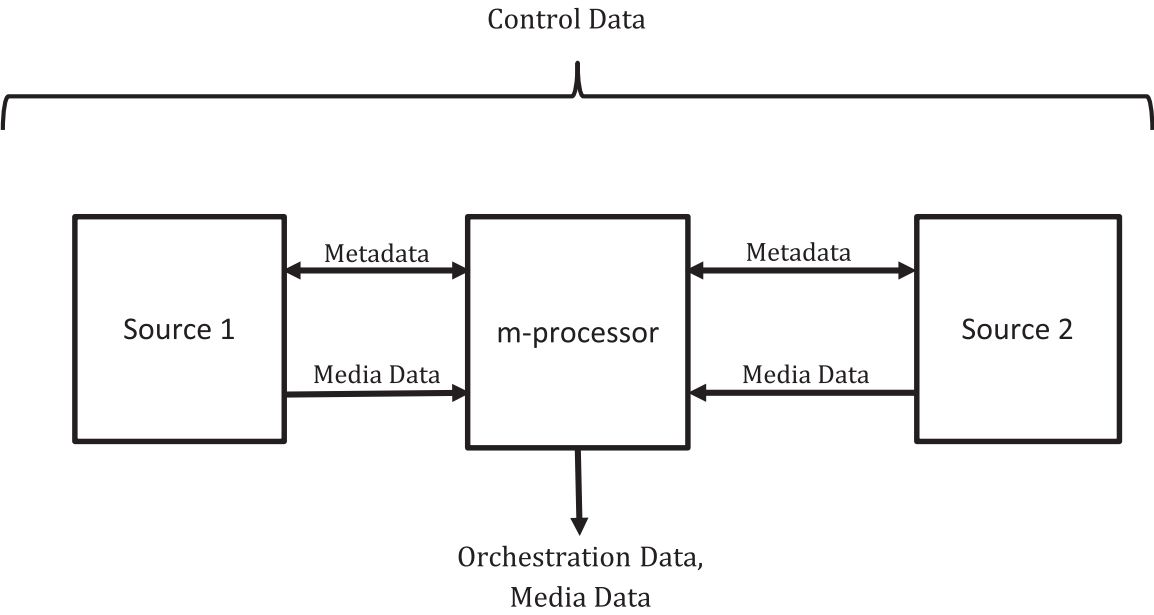


Figure 14 — A simple structure of end-to-end system in terms of spatial orchestration

Figure 15 illustrates spatial orchestration at the m-processor side.

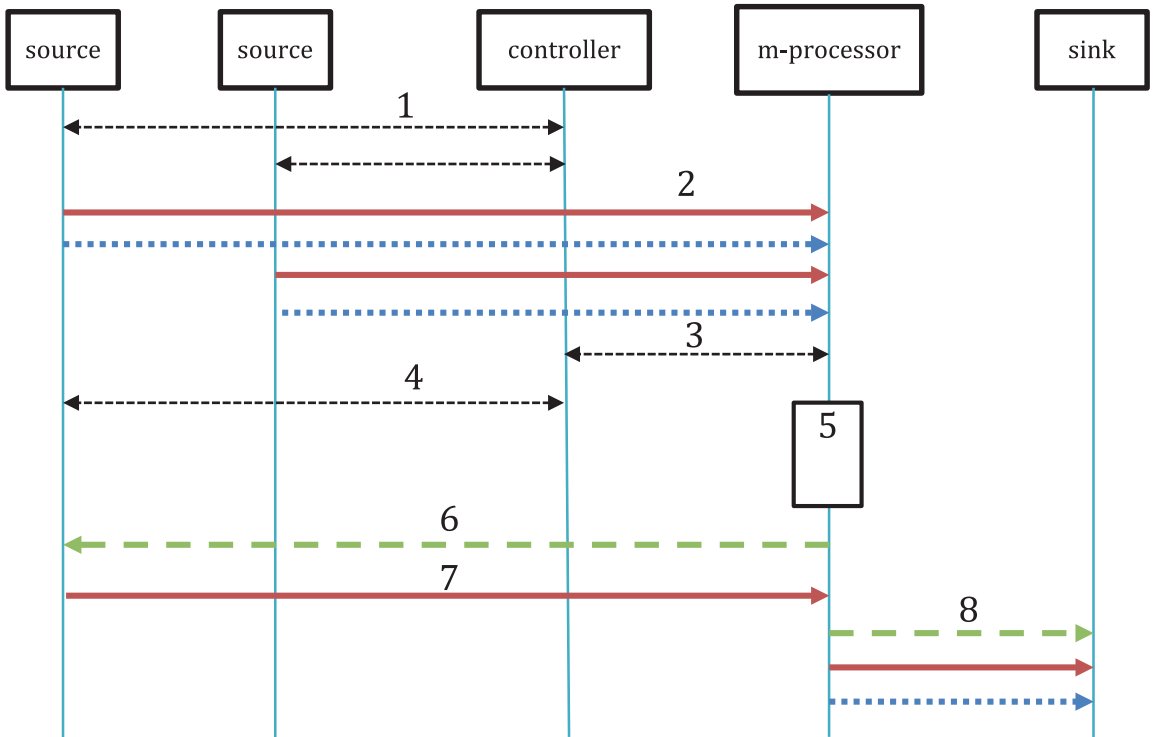


Figure 15 — Spatial orchestration at the m-processor side

The sequence diagram of Figure 10 has the following steps.

1. Discovery/registration processes, during which one or more sources are associated with a controller for a specific media orchestration session.
2. The sources start sending media data and/or metadata. The sources can have their own sensors, which can generate metadata associated with media data they generate.