

SCTE • ISBE[®]

S T A N D A R D S

Digital Video Subcommittee

AMERICAN NATIONAL STANDARD

ANSI/SCTE 223 2018

Adaptive Transport Stream

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) / International Society of Broadband Experts (ISBE) Standards and Operational Practices (hereafter called “documents”) are intended to serve the public interest by providing specifications, test methods and procedures that promote uniformity of product, interchangeability, best practices and ultimately the long-term reliability of broadband communications facilities. These documents shall not in any way preclude any member or non-member of SCTE•ISBE from manufacturing or selling products not conforming to such documents, nor shall the existence of such standards preclude their voluntary use by those other than SCTE•ISBE members.

SCTE•ISBE assumes no obligations or liability whatsoever to any party who may adopt the documents. Such adopting party assumes all risks associated with adoption of these documents, and accepts full responsibility for any damage and/or claims arising from the adoption of such documents.

Attention is called to the possibility that implementation of this document may require the use of subject matter covered by patent rights. By publication of this document, no position is taken with respect to the existence or validity of any patent rights in connection therewith. SCTE•ISBE shall not be responsible for identifying patents for which a license may be required or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention.

Patent holders who believe that they hold patents which are essential to the implementation of this document have been requested to provide information about those patents and any related licensing terms and conditions. Any such declarations made before or after publication of this document are available on the SCTE•ISBE web site at <http://www.scte.org>.

All Rights Reserved
© Society of Cable Telecommunications Engineers, Inc. 2018
140 Philips Road
Exton, PA 19341

Table of Contents

NOTICE	2
1 INTRODUCTION	7
1.1 Overview	7
1.2 Purpose of Document	8
1.3 Scope	8
1.4 Requirements	9
2 REFERENCES	10
2.1 Normative References.....	10
2.2 Informative References.....	10
2.3 Reference Acquisition.....	11
3 TERMS AND DEFINITIONS	12
4 ABBREVIATIONS AND ACRONYMS.....	15
5 ATS METADATA OVERVIEW.....	16
6 ATS SOURCE DESCRIPTION	17
6.1 Introduction	17
6.2 Identification.....	17
6.3 Non-HTTP URIs.....	17
6.3.1 <i>General</i>	17
6.3.2 <i>MPEG-2 TS over UDP</i>	17
6.3.3 <i>File</i>	17
6.4 Anchor addressing	18
6.4.1 <i>Introduction</i>	18
6.4.2 <i>MPEG-2 TS Addressing</i>	18
6.4.3 <i>EBP Addressing</i>	18
6.5 Carriage of ATS Source Description in MPEG-2 Transport Streams.....	18
7 ATS SPECIFICS.....	20
7.1 Format.....	20
7.2 General Makeup.....	20
7.3 Interleaving options	20
7.3.1 <i>Method 1 – All representations carry all audio streams</i>	21
7.3.2 <i>Method 2 – All representations carry a common subset of audio streams</i>	21
7.3.3 <i>Method 3 - Audio carried in separate Representations</i>	22
7.4 Late Binding Packaging Implications.....	22
7.5 Encoder Boundary Point (EBP).....	22
7.5.1 <i>EBP AF Descriptor Tags</i>	23
7.5.2 <i>EBP Structure Placement in MPEG-2 TS Elementary Streams</i>	23
7.5.3 <i>Usage of Timeline AF Descriptor Structure</i>	24
7.5.4 <i>Boundary AF Descriptor Structure</i>	26
7.5.5 <i>Boundary AF Descriptor Semantics</i>	26
7.5.6 <i>Labeling AF Descriptor Structure</i>	27
7.5.7 <i>Labeling AF Descriptor Semantics</i>	27
7.5.8 <i>Usage of Labeling Descriptor in MPEG-2 TS</i>	27
7.6 Partitions.....	28
7.7 PMT Descriptors.....	28
7.7.1 <i>PMT Descriptors related to EBP Data</i>	28
7.8 Use of Private Adaptation Descriptors	29

7.9	Video Boundary Points	29
7.10	Audio Boundary Points.....	30
7.11	Audio / Video Skew.....	31
7.12	ATS Boundaries in Relation to HLS/HDS	31
7.13	ATS Boundaries in Relation to HSS.....	32
7.14	Auxiliary data streams	33
7.14.1	Auxiliary data in PES Streams.....	33
8	ADAPTIVE STREAMING CONDITIONING	34
8.1	Chunk Conditioning and Synchronization.....	34
8.2	Video Conditioning and Synchronization.....	34
8.3	Video Chunk Sync: Start-up Considerations	34
8.4	Video Adaptive Sync	35
8.5	SCTE 35 and Splice Points.....	36
8.6	Ad Breaks	36
8.7	Data, Audio, Video Stream Alignment Considerations	37
8.8	Input Frame Loss	37
8.9	Audio Conditioning and Synchronization	38
8.10	AAC Family Audio Chunk Boundary requirements.....	40
8.11	Audio Alignment Across Representations.....	41
APPENDIX I	EBP STRUCTURE USE CASE EXAMPLES.....	42
I.1	Indicate a Base Fragment Boundary Point.....	42
I.2	Indicate a Common Segment Boundary and Fragment Boundary Point	43
I.3	Indicate a segmentation_upid_type (an identifier) and segmentation_type_id (stream point label) in Labeling Descriptor	43
I.4	Indicate NTP Time for Linear Services using TEMI AF_descriptor	44
APPENDIX II	PMT EBP VIRTUAL_SEGMENTATION_DESCRIPTOR USE CASE EXAMPLES	46
II.1	Signaling timing and partitions for HSS and HLS.....	46
II.2	Explicit and implicit partitions	46
APPENDIX III	DERIVATION OF ACQUISITION TIME (INFORMATIVE)	50
APPENDIX IV	DETERMINING NETWORK DRIFT EFFECTS ON ARRIVAL TIMES OF ATS SETS	52
APPENDIX V	GUIDELINES FOR BOUNDARY CREATION.....	53
V.1	Adjusting Chunk Durations	53
V.2	Frame Rate Decimation	54
V.3	Boundaries Desired at Unaligned AUs	56
V.4	Boundaries for Scene Changes in Multi Framerate Streams.....	58

Figures

Figure 1 - Unified Transcoder/Packager.....	7
Figure 2 - Separate Transcoder/Packager with ATS between	7
Figure 3 – ATS Streams with EBP data inserted at transcoder.....	8
Figure 4 – Fragmentation of an ATS generated stream	9
Figure 5 - Example of Audio and Video in a TS stream following HRD model and SCTE 128	20
Figure 6 - Interleave Method 1	21
Figure 7 - Interleave Method 1 with Audio 4 used by one Video representation	21
Figure 8 - Interleave Method 2	21
Figure 9 - Interleave Method 3, Hybrid of Method 2	22
Figure 10 - Another Interleave Method 3 Embodiment.....	22
Figure 11 - Boundary Points.....	23
Figure 12 - Candidate Reference Time Sample Points in Signal Chain in Nominal Transcoder.....	26
Figure 13 - Boundary Spec Indicating Fragment and Segment Boundaries	29
Figure 14 - Boundary Points at Ad Boundaries	30
Figure 15 - Implicit / Derived HLS Audio Segment Boundaries.....	30
Figure 16 - Implicit / Derived HDS Audio Fragment Boundaries.....	30
Figure 17 - Explicit Audio Fragment Boundaries.....	31
Figure 18 - Audio Skew from Video (e.g., HLS Segments).....	31
Figure 19 - Audio Skew from Video (e.g., HSS Fragments).....	31
Figure 20 - ATS Segment Byte Ranges.....	32
Figure 21 - ATS Video/Audio Fragments	32
Figure 22 - Video transcoding across one or more systems	34
Figure 23 - Conditioning Video Start	35
Figure 24 - Source Time Discontinuities	35
Figure 25 - New Segment at Ad Boundaries	36
Figure 26 - New Segment and Fragment at Ad Boundaries	36
Figure 27 - Multiple Break Boundaries	37
Figure 28 - Splice point and Video and Audio AUs.....	37
Figure 29 - ATS Transcoder with Audio distributed to Multiple Muxes.....	38
Figure 30 - ATS Transcoder with Audio encoded for each output.....	38
Figure 31 - Audio encoded in multiple ATS Transcoders.....	39
Figure 32 - Audio Access Units: Groups of Samples	40
Figure 33 - Nonaligned Audio Access Units	40
Figure 34 - ISO/IEC 13818-1 compliant transcoder model.....	50
Figure 35 - Candidate Reference Time Sample Points in Signal Chain in Nominal Transcoder.....	51
Figure 36 – Effects on Network Drift on Arrival Times.....	52
Figure 37 - Adjusting Durations Around/After Splice Points.....	53
Figure 38 - Example 2, Adjusting Durations Around/After Splice Points.....	54
Figure 39 - Adjusting Durations with DASH-IF IOP 50% Tolerance to Align to Original Timeline	54
Figure 40 - One Second of Frame Rate Decimation (25 to 12.5)	55
Figure 41 - Two Seconds of Frame Rate Decimation (25 to 12.5).....	55
Figure 42 - Non integer frame rate reduction	55

Figure 43 - Naturally Aligned Chunk Boundary	56
Figure 44 - Desire for new Chunk	56
Figure 45 - First AU >= Splice Point.....	56
Figure 46 - Chunk simply starting at next frame rate decimated frame AU >= splice point	57
Figure 47 - Using AUs to Complete Chunk	57
Figure 48 - Skipping AUs leading up to the next Chunk.....	57
Figure 49 - Skipping AUs leading up to the next Chunk and Extending previous AU's Duration	58
Figure 50 - Selecting first AU aligned across four bitrates.....	58
Figure 51 - Selecting first AU aligned across two bitrates	58
Figure 52 - Scene change in close proximity to anticipated chunk boundary.....	59
Figure 53 - I (blue) and IDR (red) frames due to Scene Change and Chunk Boundaries.....	59
Figure 54 - Adjusted Chunk Boundary at Scene Change	59

TABLES

Table 1- Parameters for MPEG-2 TS Anchors	18
Table 2- Parameters for EBP Anchors.....	18
Table 3 – AF Descriptor Tags	23
Table 4- TEMI Timeline Descriptor.....	25
Table 5 – Segment SN Properties	32
Table 6- Fragment FN Properties	33
Table 7 - EBP Structure for Fragment Boundary Point.....	42
Table 8 - EBP Structure for Segment Boundary Point	43
Table 9- Carriage of Identifier and Program Start using the Labeling Af_descriptor	44
Table 10- Carriage of NTP for Linear Services using the TEMI af_descriptor.....	45
Table 11 - Example of PMT EBP virtual_segmentation_descriptor() for HLS and HSS video PID.....	46
Table 12 - Example of PMT EBP virtual_segmentation_descriptor() for HLS and HSS audio PID.....	47
Table 13 - Example of PMT EBP virtual_segment_descriptor() for HLS, HSS, HDS (implicit to partition=2) video PID.....	47
Table 14 - Example of PMT EBP virtual_segmentation_descriptor() for HLS, HSS, HDS (implicit partition=2) audio PID.....	48
Table 15 - Example of PMT EBP virtual_segmentation_descriptor() for HLS, HSS, and HDS (explicit partition=3) video	48
Table 16 - Example of PMT EBP virtual_segmentation_descriptor() for HLS, HSS, and HDS (implicit partition=3) audio PID.....	49

1 INTRODUCTION

1.1 Overview

There are a variety of Adaptive Streaming wire formats. Some are based on an MPEG-2 Transport Stream container such as HLS (HTTP Live Streaming: Apple) and others on a fragmented MP4 container such as HSS (HTTP Smooth Streaming: Microsoft) and HDS (HTTP Dynamic Streaming: Adobe); whereas DASH (Dynamic Adaptive Streaming over HTTP: MPEG) supports both containers. While different, they utilize common video and audio compression formats; namely: ISO/IEC 14496-10 (AVC) and ISO/IEC 14496-3 (AAC). Additional audio formats, such as Dolby Digital Plus and DTS-HD, *may* also be supported by these or a subset of these Adaptive Bit Rate (ABR) formats.

In a unified ABR encoding and packaging system, video and audio data are encoded and conditioned for adaptive streaming purposes and the resultant elementary compressed access units are fed to one or more ABR packagers or encapsulators to be formatted into ABR-specific wire formats.

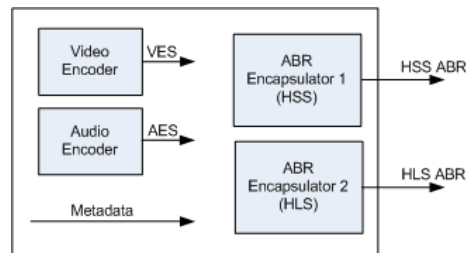


Figure 1 - Unified Transcoder/Packager

The Adaptive Transport Stream (ATS) format described in this specification allows for streaming/storage of adaptive streaming content originating as Transport Streams in a generic manner without restricting this to a particular adaptive streaming delivery technology (HSS/HLS/HDS). As Figure 2 illustrates, this allows for a separation of the transcoding process from the encapsulation process that produces ABR-specific formats.

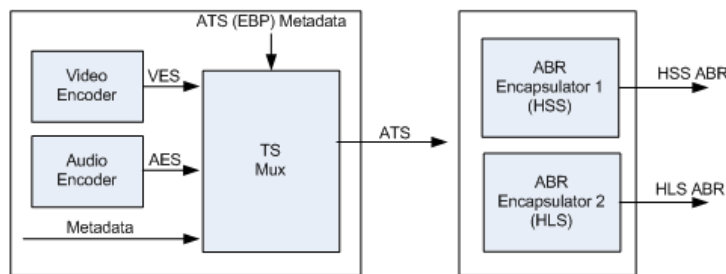


Figure 2 - Separate Transcoder/Packager with ATS between

This specification defines a fully **compliant** continuous single program MPEG-2 Transport Stream which follows the HRD model and provides markers in the stream through the use of PES frame encapsulated `af_descriptors` (adaptation field descriptors) [11]. These markers identify conditioned points in the stream that are virtual segments that can be partitioned into segments used for ABR applications [11]. Downstream encapsulation is expected to re-encapsulate an ATS, which *may* involve partial or complete de-encapsulation (demux) prior to encapsulating into a target ABR format. Since this downstream encapsulation does no re-encoding of the media data, the video and audio access units in the ATS need to be pre-conditioned for adaptive streaming purposes [in accordance with section 8 of this specification]. Additionally this specification defines `af_descriptor` metadata, Boundary Descriptor /Labeling Descriptor/Timeline Descriptor and collectively called Encoder Boundary Points (EBP) data [as defined in section 7.5], that are injected and carried in the transport stream layer to provide adaptive boundary information to downstream processing tasks. The EBP data provides a hinting mechanism for taking continuous streams conditioned for adaptive streaming and creating discrete chunks of decodable content with boundaries in one

component stream in the multiplex (Fragment) or across the multiplex (Segment). This specification also references additional PMT descriptors as specified in 13818-1 [11] that may be carried in the ATS for informative purposes to describe the various conditioning and boundary points used in the stream(s). Lastly, structural information on a related set of ATS streams *may* be carried through an ATS Source Description.

The EBP data contained within an ATS stream is carried as data descriptors of the public adaptation field (af_descriptors) of an MPEG-2 TS packet for video or audio and can be applied to each video packetized elementary stream (PES) and audio PES packet, resolved down in many cases to a single Video access unit (AU) or a group of audio AUs [11]. It contains a set of af_descriptors to indicate Boundary, Labeling, and Timeline information.

1.2 Purpose of Document

The purpose of this document is to define an ATS stream, the boundary points within it, both explicit and implicit, how boundary points map to various ABR formats such as HSS Fragments and HLS Segments (both in the video and audio domain), and the time stamps, durations and byte ranges of these chunks.

To that end, a significant portion of this document describes the basic requirements for adaptive video and audio conditioning. These sections detail conditioning considerations such as varying frame rates, advertising splice points and input loss handling.

1.3 Scope

This standard describes the requirements and constraints on a single program transport stream (SPTS) that allow it to be used as an Adaptive Transport Stream, including stream conditioning and signaling of segment boundary points. Typically, multiple ATSS will be generated from a single input and sent to a packager, recorder or other device. The EBP structure can be inserted at the time of encoding or added during the transcoding process. This specification does not describe how an ATS is stored or how it *may* be converted to target delivery formats.

This document describes the wrapping, chunking, and conditioning of packetized elementary streams carried over MPEG-2 TS. These elementary streams are codec independent and could carry AVC, HEVC, or even MPEG-2 video. Reference is made to SCTE 128 [3][33] to be compliant with MPEG-2 Systems layer constraints on the use of adaptation field public data.

These created ATS streams are then sent to a packager (also called a fragmentor or encapsulator) directly or stored to be sent to a packager upon request at a later time. Upon receiving such streams, a packager then processes these streams with EBP data and produces chunks according to the one or more adaptive streaming encapsulating technologies.

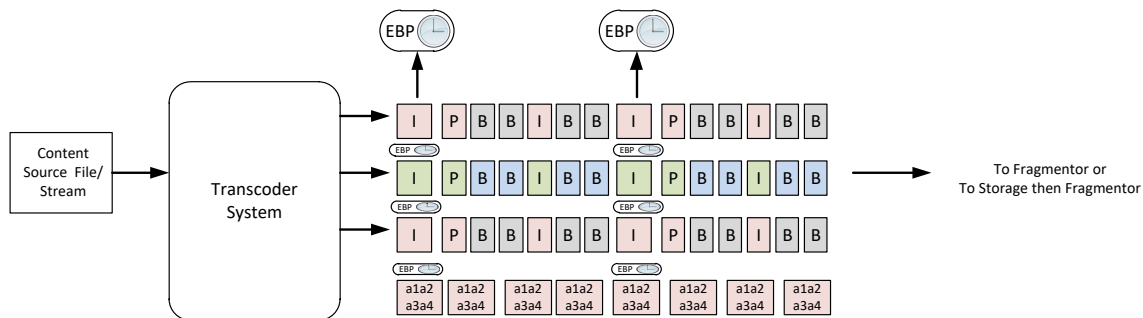


Figure 3 – ATS Streams with EBP data inserted at transcoder