SCTE | STANDARDS

Digital Video Subcommittee

SCTE STANDARD

SCTE 138 2019 (R2024)

Stream Conditioning for Switching of Addressable Content in Digital Television Receivers

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) 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, interoperability, interchangeability, best practices, and the long term reliability of broadband communications facilities. These documents shall not in any way preclude any member or non-member of SCTE 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 members.

SCTE 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.

NOTE: The user's attention is called to the possibility that compliance with this document may require the use of an invention covered by patent rights. By publication of this document, no position is taken with respect to the validity of any such claim(s) or of any patent rights in connection therewith. If a patent holder has filed a statement of willingness to grant a license under these rights on reasonable and nondiscriminatory terms and conditions to applicants desiring to obtain such a license, then details may be obtained from the standards developer. SCTE 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 web site at https://scte.org.

All Rights Reserved
© 2024 Society of Cable Telecommunications Engineers, Inc.
140 Philips Road
Exton, PA 19341

Document Tags

☒ Specification	☐ Checklist	☐ Facility
☐ Test or Measurement	☐ Metric	☒ Access Network
☐ Architecture or Framework	☒ Cloud	☒ Customer Premises
☐ Procedure, Process or Method		

Document Release History

Release	Date
SCTE 138 2009	*2009*
SCTE 138 2013	*2013*
SCTE 138 2019	*2019*
SCTE 138 2019 (R2024)	*3/11/24*

Note: Standards that are released multiple times in the same year use: a, b, c, etc. to indicate normative balloted updates and/or r1, r2, r3, etc. to indicate editorial changes to a released document after the year.

Note: This document is a reaffirmation of SCTE 138 2019. No substantive changes have been made to this document. Information components may have been updated such as the title page, NOTICE text, headers, and footers.

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

1. Introduction

1.1. Executive Summary

Note: This document is a reaffirmation of SCTE 138 2019. No substantive changes have been made to this document. Information components may have been updated such as the title page, NOTICE text, headers, and footers.

This standard supports the delivery of household addressable advertising in linear programs. This document also describes the stream Conditioning required to enable Client-DPI Devices to implement Switching of compressed video streams with associated compressed audio streams. This standard enables Client-DPI Devices to support addressable advertising where the Addressable Content Set appears in an MPEG program outside the Primary Channel.

1.2. Scope

The Switch from the Primary Channel to the Addressable Content Set and back to the Primary Channel is defined by several levels.

For "Level 0", or "L0" systems, it defines Switching among compressed video and compressed audio streams in a non-seamless fashion (similar to a channel change).

For "Level 1", or "L1" systems, it defines Switching among compressed video and compressed audio streams in a seamless fashion.

All L0 or L1 Switches ***shall*** occur from a video stream to another video stream compressed using the same video codec. All L0 or L1 Switches ***shall*** occur from an audio stream to another audio stream compressed using the same audio codec.

1.2.1. Supported video and audio codecs

This standard supports the following video and audio codecs:

- Video: MPEG-2, AVC, and HEVC.
- Audio: AC-3, E-AC-3, AAC, AC-4, MPEG-H Audio, DTS-HD, and DTS-UHD.

Readers of this standard will find references to other relevant standards in later sections of this document. Section 2.1 SCTE References, section 2.2 Standards from Other Organizations and section 6 Standards Compliance describe specific requirements for each of these codecs.

Codecs refered to in requirements defined by this standard are specified in other sections of this document. For example, video Filler requirements and related standards for the AVC video codec can be found in section 8.4.2 AVC Video Filler.

Triggers and signaling are out of scope for this standard.

Note: Client-DPI Devices include Receivers conforming to OCAP as well as other cable-compatible Receivers.

1.3. Benefits

Client-DPI Devices will Switch to Addressable Content Sets either in the same multiplex as the primary Channel or in a new multiplex depending on the level. This standard ensures that conditioning occurs to properly formatted picture data and audio data in the transmitted Primary Channels and the Addressable Content Sets. This ensures that the Client-DPI Device will be able to acquire and begin decoding the new content without displaying any artifacts or losing any significant video or audio data. Without this conditioning, there would be visible artifacts during an addressable Client-DPI Device Switch.

1.4. Intended Audience

The intended audience for this standard is MVPDs, ad insertion equipment manufacturers, Client-DPI Device manufacturers, and addressable advertisement companies.

1.5. Areas for Further Investigation or to be Added in Future Versions

Areas for future consideration might include:

- Support for additional codecs beyond the ones included in this standard.
- Switches where there is a codec change in a Client-DPI Device implementation.
- Further consideration for C3/C7 rating periods.

2. Normative References

The following documents contain provisions, which, through reference in this text, constitute provisions of this document. At the time of Subcommittee approval, the editions indicated were valid. All documents are subject to revision; and while parties to any agreement based on this document are encouraged to investigate the possibility of applying the most recent editions of the documents listed below, they are reminded that newer editions of those documents might not be compatible with the referenced version.

2.1. SCTE References

1. [SCTE 30] ANSI/SCTE 30 2017: Digital Program Insertion Splicing API.
2. [SCTE 43] ANSI/SCTE 43 2015: Digital Video Systems Characteristics Standard for Cable Television.
3. [SCTE 54] ANSI/SCTE 54 2015: Digital Video Service Multiplex and Transport System for Cable Television
4. [SCTE 128-1] ANSI/SCTE 128-1 2018: AVC Video Constraints for Cable Television Part 1: Coding
5. [SCTE 128-2] ANSI/SCTE 128-2 2018: AVC Video Constraints for Cable Television Part 2: Transport
6. [SCTE 172] ANSI/SCTE 172 2017: Constraints On AVC and HEVC Structured Video Coding for Digital Program Insertion
7. [SCTE 193-1] SCTE 193-1 2014: MPEG-4 AAC Family Audio System - Part 1 Coding Constraints For Cable Television
8. [SCTE 193-2] SCTE 193-2: 2014: MPEG-4 AAC Family Audio System - Part 2 Constraints For Carriage Over MPEG-2 Transport
9. [SCTE 194-1] SCTE 194-1 2018: DTS-HD AUDIO SYSTEM – Part 1: Coding Constraints for Cable Television
10. [SCTE 194-2] SCTE 194-2: 2018 DTS-HD AUDIO SYSTEM – Part 2: Constraints for Carriage over MPEG-2 Transport

11. [SCTE 215-1] SCTE 215-1 2018: HEVC Video Constraints for Cable Television Part 1- Coding
12. [SCTE 215-2] SCTE 215-2 2018: HEVC Video Constraints for Cable Television Part 2-Transport
13. [SCTE 242-1] ANSI/SCTE 242-1 2017: Next Generation Audio Coding Constraints for Cable Systems: Part 1 – Introduction and Common Constraints
14. [SCTE 243-1] ANSI/SCTE 243-1 2017: Next Generation Audio Carriage Constraints For Cable Systems: Part 1 – Common Transport Signaling
15. [SCTE 242-2] ANSI/SCTE 242-2 2017: Next Generation Audio Coding Constraints for Cable Systems: Part 2 – AC-4 Audio Coding Constraints
16. [SCTE 243-2] ANSI/SCTE 243-2 2017: Next Generation Audio Carriage Constraints For Cable Systems: Part 2 – AC-4 Audio Carriage Constraints
17. [SCTE 242-3] ANSI/SCTE 242-3 2017: Next Generation Audio Coding Constraints for Cable Systems: Part 3 – MPEG-H Audio Coding Constraints
18. [SCTE 243-3] ANSI/SCTE 243-3 2017: Next Generation Audio Carriage Constraints For Cable Systems: Part 3 – MPEG-H Audio Carriage Constraints
19. [SCTE 242-4] ANSI/SCTE 242-4 2018: Next Generation Audio Coding Constraints for Cable Systems: Part 3 – DTS-UHD Audio Coding Constraints
20. [SCTE 243-4] ANSI/SCTE 243-4 2018: Next Generation Audio Carriage Constraints for Cable Systems: Part 3 –DTS-UHD Audio Carriage Constraints

2.2. Standards from Other Organizations

21. [13818-1] ISO/IEC 13818-1 (2018): International Standard, Information Technology – Generic coding of moving pictures and associated audio information: Systems.
22. [13818-2] ISO/IEC 13818-2 (2013): International Standard, Information Technology – Generic coding of moving pictures and associated audio information: Video
23. [14496-3] ISO/IEC 14496-3 (2009):Information technology – Coding of audio-visual objects – Part 3: Audio
24. [14496-10] ISO/IEC 14496-10 (ITU-T H.264): International Standard (2014), Advanced video coding for generic audiovisual services.
25. [23008-2] ISO/IEC 23008-2 (ITU H.265): International Standard (2017), Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding
26. [A/52] ATSC A/52 (2018): Digital Audio Compression Standard (AC-3, E-AC-3)
27. [A/53] ATSC A/53 Part 5 (2014): AC-3 Audio System Characteristics
28. [A/342-1]"ATSC Standard: A/342 Part 1, Audio Common Elements," Doc. A/342-1:2017
29. [A/342-2]"ATSC Standard: A/342 Part 2, AC-4 System," Doc. A/342-2:2017
30. [A/342-3] "ATSC Standard: A/342 Part 3, MPEG-H System," Doc. A/342-3:2017
31. [TS 102 114] ETSI TS 102 114 V1.5.1 (2018-05) DTS Coherent Acoustics; Core and Extensions with Additional Profiles
32. [TS 103 491] ETSI TS 103 491 V1.1.1 (2017-04) DTS - UHD Audio Format; Delivery of Channels, Objects and Ambisonic Sound Fields
33. [ITU-R BT.1359-1] Recommendation ITU-R BT.1359-1 Relative Timing Of Sound and Vision for Broadcasting.

2.3. Published Materials

- No normative references are applicable.

3. Informative References

The following documents might provide valuable information to the reader but are not required when complying with this document.

3.1. SCTE References

34. [SCTE 67] ANSI/SCTE 67 2017, Recommended Practice for Digital Program Insertion for Cable

3.2. Standards from Other Organizations

- No informative references are applicable.

3.3. Published Materials

- No informative references are applicable.

4. Compliance Notation

shall	This word or the adjective "***required***" means that the item is an absolute requirement of this document.
shall not	This phrase means that the item is an absolute prohibition of this document.
forbidden	This word means the value specified shall never be used.
should	This word or the adjective "*recommended*" means that there may exist valid reasons in particular circumstances to ignore this item, but the full implications should be understood and the case carefully weighted before choosing a different course.
should not	This phrase means that there may exist valid reasons in particular circumstances when the listed behavior is acceptable or even useful, but the full implications should be understood and the case carefully weighed before implementing any behavior described with this label.
may	This word or the adjective "*optional*" means that this item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because it enhances the product, for example; another vendor may omit the same item.
deprecated	Use is permissible for legacy purposes only. Deprecated features may be removed from future versions of this document. Implementations should avoid use of deprecated features.

5. Abbreviations and Definitions

5.1. Abbreviations

AFD	active format descriptor
ATSC	Advanced Television Systems Committee
AVC	advanced video coding
CAT	Conditional Access Table. [13818-1]
DPB	decoded picture buffer
DPI	Digital Program Insertion.
FPP	forward predicted picture [SCTE 128-1]

GOP	group of pictures [13818-2]
HDR	High Dynamic Range
HEVC	High Efficiency Video Coding
I	intraframe
IDR	instantaneous decoding refresh
IP	internet protocol
MPEG	Moving Picture Experts Group
MPTS	Multiprogram Transport Stream. [13818-1]
MVPD	Multichannel Video Programming Distributor
NAL	network abstraction layer as defined in [14496-10].
OCAP	Open Cable Application Platform
P	predictive frame
PAT	program association table [13818-1]
PCR	program clock reference 13818-1
PES	packetized elementary stream 13818-1
PID	packet identifier 13818-1
PMT	program map table 13818-1
PSI	program specific information 13818-1
PTS	presentation time stamp 13818-1
QAM	quadrature amplitude modulation
SCTE	Society of Cable Telecommunications Engineers
SDR	Standard Dynamic Range
SRAP	SCTE (AVC) Random Access Point as defined in[SCTE 128-2]
T-STD	Transport Stream System Target Decoder 13818-1
TRAIL_N	Non-reference trailing picture in HEVC [SCTE 215-1]
SHRAP	SCTE HEVC Random Access Point as defined in [SCTE 215-1]

5.2. Definitions

Addressable Content Set	A collection of simultaneously transmitted elementary streams of equal presentation duration typically containing appropriate combinations of media streams (video, audio and data) that may be selected as the result of a Switch.
Advanced Video Coding	Refers specifically to video compression standardized in [14496-10]
C3	A metric for the ratings for average commercial minutes in live programming plus total playback by digital video recorder out to three days after live playback.
C7	A metric for the ratings for average commercial minutes in live programming plus total playback by digital video recorder out to seven days after live playback.
Client-DPI Device	A Receiver or other device that is able to select among elementary streams from within an Addressable Content Set.
Conditioning	This term refers to the set of rules that specify constraints on transport, video and audio that enable both a seamless and a non-seamless Switch within Client-DPI Devices.
Decision Engine	The algorithm in a Client-DPI Device that selects which particular content should be selected from an Addressable Content Set and presented to the viewer.
Filler	Used in Level 0 Streams, this is the region within which the Client-DPI Device begins and completes a Switch. Filler is inserted into all

 © 2024 SCTE

	video and audio elementary streams that comprise an Addressable Content Set.
Gap	Used in Level 1 Streams, this is the region within which the Client-DPI Device begins and completes a Switch. The Gap is inserted into all video and audio elementary streams that comprise an Addressable Content Set.
Instantaneous Decoding Refresh	A picture type defined in [14496-10]
IDR NAL unit	An IDR slice [14496-10]. (A NAL unit with nal_unit_type = 5.)
Insertion Channel	The MPEG Programs comprising the Addressable Content Set that replaces the Primary Channel in whole or in part for the duration of a Switch Event.
Insertion Sequence	A set of back-to-back Switch Events.
Level 0 / L0	Streams or Switches resulting in non-seamless transitions.
Level 1 / L1	Streams or Switches resulting in seamless transitions.
Level 0 Stream	A Primary Channel or Insertion Channel that has been conditioned in order to enable a Level 0 Switch. [SCTE 67]
Level 1 Stream	A Primary Channel or Insertion Channel that has been conditioned in order to enable a Level 1 Switch. [SCTE 67]
Level 0 Switch	This term refers to a Switch from one video elementary stream to another video elementary stream and one audio elementary stream to another audio elementary stream without any perceptible audio or video artifacts, but which *may* include the use of black pictures and audio silence to mask any artifacts during the latency of the Switch.
Level 1 Switch	This term refers to a Switch from one video elementary stream to another video elementary stream of the same stream_type and/or one audio elementary stream to another audio elementary stream without any perceptible delay, or audio or video artifacts.
Primary Channel	The MPEG Program that is replaced in whole or in part by members of the Addressable Content Set during a Switch Event.
Receiver	A cable television receiving device which is typically a set-top box or other digital video device .
Seamless	A transition from one content set to another (either video or audio, or both) without any artifacts or discontinuities. This transition is usually invisible and inaudible to the viewer.
Signaled Switch Point	The point within the MPEG-2 transport stream where the Switching Engine *may* initiate a Switch.
Switch	The action performed by a Client-DPI Device, which chooses certain elementary streams out of an Addressable Content Set. Note: switch and splice are typically used interchangeably.
Switch Event	The time between two adjacent switches during which Client-DPI Devices present an Insertion Channel.
Switching Engine	The functionality that executes a Switch in a Client-DPI Device.
Trigger Signal	An element within the MPEG-2 transport stream multiplex which indicates the location of a Signaled Switch Point.

6. Standards Compliance

All transport streams (with video and audio components) delivered to Client-DPI Devices ***shall*** be compliant with [SCTE 54]. In addition:

- MPEG-2 video coding shall be compliant with [SCTE 43].
- AVC video coding shall be compliant with [SCTE 128-1] and carriage shall be compliant with [SCTE 128-2].
- HEVC video coding shall be compliant with [SCTE 215-1] and carriage shall be compliant with [SCTE 215-2].
- AC-3 audio coding shall be compliant with ATSC [A/52] as constrained by [A/53] Part 5.
- E-AC-3 audio coding shall be compliant with ATSC [A/52] as constrained by SCTE 54.
- AC-4 audio coding shall be compliant with [SCTE 242-2] and carriage shall be compliant with [SCTE 243-2]
- DTS-HD audio coding shall be compliant with [SCTE 194-1] and carriage shall be compliant with [SCTE 194-2].
- DTS-UHD coding shall be compliant with [SCTE 242-4] and carriage shall be compliant with [SCTE 243-4].
- MPEG-H audio coding shall be compliant with [SCTE 242-3] and carriage shall be compliant with [SCTE 243-3].

7. The PMT

SCTE 54 constrains each PMT PID to contain a single TS_program_map_section. Any PMT structures used by Addressable Content *should* conform to this constraint. [SCTE 54]

In systems where addressable content is present in the same transport stream as the Primary Channel, all of the PMTs ***shall*** be present in the transport stream at all times. Changes to any PMT in the vicinity of a Filler or Gap *should* be avoided.

Notes:

- The preceding requirement will result in one or more PMTs being present in the transport stream that describe PIDs for which data will be present only when addressable content is present.
- It may be necessary to transmit PAT, CAT and PMT more frequently than required by SCTE 54 during the Filler in Level 0 Streams. The aggregate of PAT, PMT and CAT transmitted in this interval should not exceed 80,000 bits/second.

8. Level 0 Requirements

Level 0 Conditioning consists of a Filler, as defined in section 8.4, and video coding requirements defined in sections 8.8, 8.9, and 8.10. The Conditioning described in this section ***shall*** be present in the transport stream at the input to a Client-DPI Device when a Level 0 Switch is enabled. In this context, "enabled" means that a Client-DPI Device *may*, as determined by its Decision Engine, perform a Level 0 Switch. The following requirements for Level 0 apply to both standard definition and high definition formats in coded video streams.

8.1. Use of Transport Streams and Services

In Level 0 systems, one or more transport streams *may* be used to convey the elementary streams comprising an Addressable Content Set. All elementary streams selected as the result of a Level 0 Switch ***shall*** reside in the same transport stream and in the same service.

8.2. Video and Audio Elementary Stream Presentation Duration Differences

Due to the fact that addressable content is preceded by Filler, a time difference (positive or negative) is introduced between corresponding audio and video presentation time stamps (PTS). This difference is introduced even when the original addressable content has been created with perfect synchronization. However, it can be shown (see Appendix B: Bound on A/V Presentation Time Mismatch Due to Splicing) that under the following assumptions, the absolute value of the PTS difference introduced is bounded. Hence, the PTS difference between audio and video introduced by any technique under the following assumptions ***shall*** be less than or equal to one-half of the audio frame duration.

The presentation duration variance value calculations can be found in Appendix B: Bound on A/V Presentation Time Mismatch Due to Splicing. Those calculations ***shall*** be based on the following:

- A/V Filler (black video, muted audio) is precomputed with a target duration (e.g., 0.5 sec)
- Ad is spliced in after Filler, and PTS values are restamped to be continuous with Filler
- Video and audio frame periods are fixed and remain constant throughout Filler and ad
- PES packets contain one video or audio access unit (AU)
- The duration of the video Filler is calculated first. The duration of audio Filler is calculated to match the video Filler.

8.3. Conditioning of Primary Channel Streams

All of the video and audio elementary streams comprising the Primary Channel ***shall*** be Level 0 conditioned, as defined in section 8.0 of this standard, when a Level 0 Switch is enabled.

8.4. Filler

Filler ***shall*** be present in all of the audio and video elementary streams which comprise an Addressable Content Set when a Level 0 Switch is enabled. Filler in video elementary streams ***shall*** consist of picture data that will be decoded as black pictures. Filler in audio elementary streams ***shall*** consist of audio data that will be decoded as silence.

Filler ***shall*** be comprised of compressed video and compressed audio matching the video and audio codecs used to compress the Primary Channel and the addressable ads.

All Filler, Primary Channel and the addressable ad content ***shall not*** change PMTs during an Insertion Sequence.

Picture format changes during an Insertion Sequence ***shall*** be restricted to changes in horizontal resolution and ***shall*** be compliant with either [SCTE 43], [SCTE 128-1], or [SCTE 215-1].

The transition from Filler to subsequent non-Filler content and non-Filler content to subsequent Filler in both video and audio streams *should not* introduce any visible or audible artifacts.

Notes:

- There is no requirement that the codec used on any Insertion Channel be the same as any other Insertion Channel or Primary Channel.
- Filler may be produced at any point in the signal processing chain. For example, it could be included in stored advertising content or inserted by a splicing device.

- Client-DPI Devices may utilize any appropriate methods to minimize the time required to perform a Level 0 Switch, including communicating PID values in a manner that eliminates the need to parse the PSI (See section 7.0).

8.4.1. MPEG-2 Video Filler

The following requirements apply when performing an L0 Switch when MPEG-2 video content is present:

In MPEG-2 video streams, Filler ***shall*** begin with an I picture, and ***shall*** consist of alternating I and P pictures in transmission order throughout the duration of the Filler. Filler *may* end on an I or P picture. All I pictures ***shall*** be preceded by a sequence header and appropriate sequence extension. The transition between Filler and non-Filler content ***shall*** be compliant with MPEG-2. Field parity ***shall*** be maintained during these transitions.

In each service carrying Filler there ***shall*** be a PCR transmitted on the designated PCR_PID within 5 milliseconds prior to the transmission of the PES header that begins each I picture.

8.4.2. AVC Video Filler

The following requirements apply when performing an L0 Switch when AVC video content is present:

In AVC video streams, Filler ***shall*** begin with an IDR picture, and ***shall*** consist of alternating IDR and FPP pictures in transmission order throughout the duration of the Filler. Filler *may* end on an IDR or FPP picture. All IDR pictures ***shall*** be preceded by an SPS and appropriate VUI. The coded video sequence comprised of Filler and the adjacent content ***shall*** be compliant with [SCTE 128-1]. Field parity ***shall*** be maintained during these transitions.

In each service carrying Filler there ***shall*** be a PCR transmitted on the designated PCR_PID within 5 milliseconds prior to the transmission of the PES header that begins each IDR picture.

8.4.3. HEVC Video Filler

The following requirements apply when performing an L0 switch in the presence of HEVC video:

In HEVC video streams, Filler ***shall*** begin with an IDR picture, and ***shall*** consist of alternating IDR pictures and TRAIL_N pictures in transmission order throughout the duration of the Filler. Filler may end on an IDR or TRAIL_N picture. Each IDR picture ***shall*** be contained within a SHRAP access unit as defined in [SCTE 215-1]. The coded video sequence comprised of Filler and the adjacent content ***shall*** be compliant with [SCTE 215-1]. Field parity ***shall*** be maintained during these transitions.

In each service carrying Filler there ***shall*** be a PCR transmitted on the designated PCR_PID within 5 milliseconds prior to the transmission of the PES header that begins each IDR picture. The transport of Filler ***shall*** meet the requirements of [SCTE 215-2].

8.4.4. AC-3 or E-AC-3 Audio Filler

The following requirements apply when performing an L0 Switch when AC-3 or E-AC-3 audio content is present:

Filler in AC-3 or E-AC-3 audio elementary streams ***shall*** be comprised of complete AC-3 or E-AC-3 access units. Each AC-3 or E-AC-3 access unit ***shall*** be in its own PES packet. The transition between Filler and non-Filler content ***shall*** be compliant with AC-3 standards [A/52] and [A/53] Part 5.

8.4.5. AAC Audio Filler

The following requirements apply when performing an L0 Switch when AAC audio content is present:

Filler in AAC audio elementary streams ***shall*** be comprised of complete AAC access units. Each AAC access unit ***shall*** be in its own PES packet.

The first AAC access unit shall be a RAP AU as defined in [SCTE 193-1].

The transition between Filler and non-Filler content ***shall*** be compliant with AAC standards [SCTE 193-1] and [SCTE 193-2].

8.4.6. AC-4 Audio Filler

The following requirements apply when performing an L0 Switch when AC-4 audio content is present:

Filler in AC-4 audio elementary streams ***shall*** be comprised of complete AC-4 access units. Each AC-4 access unit ***shall*** be in its own PES packet. The transition between Filler and non-Filler content ***shall*** be compliant with AC-4 standards [A/342-1] and [A/342-2].

8.4.7. MPEG-H Audio Filler

The following requirements apply when performing an L0 Switch when MPEG-H Audio content is present:

Filler in MPEG-H Audio elementary streams ***shall*** be comprised of complete MPEG-H Audio access units. Each MPEG-H Audio access unit ***shall*** be in its own PES packet.

The MPEG-H Audio elementary stream shall start with a RAP as defined in [SCTE 243-3].

The transition between Filler and non-Filler content ***shall*** be compliant with MPEG-H Audio standards [A/342-1]and [A/342-3].

8.4.8. DTS-HD Audio Filler

The following requirements apply when performing an L0 Switch when DTS-HD audio content is present:

Filler in DTS-HD audio elementary streams ***shall*** be comprised of complete DTS-HD access units. Each DTS-HD access unit ***shall*** be in its own PES packet. The transition between Filler and non-Filler content ***shall*** be compliant with DTS-HD standards SCTE 194-1 and SCTE 194-2.

8.4.9. DTS-UHD Audio Filler

The following requirements apply when performing an L0 Switch when DTS-UHD audio content is present:

Filler in DTS-UHD audio elementary streams ***shall*** be comprised of complete DTS-UHD access units. Each DTS-UHD access unit ***shall*** be in its own PES packet and ***shall*** be a synch frame. The transition between Filler and non-Filler content ***shall*** be compliant with DTS-UHD standards [SCTE 242-4] and [SCTE 243-4].

The first access unit of Filler in DTS-UHD audio streams ***shall*** meet the requirements for a Random Access Point as defined in [SCTE 243-4], with the exception that if the DTS-UHD BroadcastChunk [SCTE 242-4] is present, it ***shall*** appear in the main stream prior to the first synch frame.

When multiple DTS-UHD streams are used to define playback, the Random Access Points of each of the component streams ***shall*** be aligned.

8.5. Filler Alignment

For a given Switch, all Filler ***shall*** be aligned. This requirement ***shall*** apply to all streams involved in a Switch regardless of whether they are contained within the same transport stream or in different transport streams.

For alignment purposes, within a service an alignment point is defined as the later of the start of Filler in either the video or audio elementary stream. Across all services involved in a given Switch, all alignment points ***shall*** occur within 35 milliseconds.

Following the alignment point, all video Filler ***shall*** have the same transmission duration. Following the alignment point, all audio Filler ***shall*** have the same transmission duration.

Note: The alignment point establishes the earliest point at which a Client-DPI Device can initiate a Switch. The start of Filler is aligned to guarantee that a Client-DPI Device which completes a Switch very quickly does not join the destination service before the transmission of Filler on that service has started. The transmission duration of Filler is chosen to guarantee that a Client-DPI Device which completes a Switch slowly does not join the destination service after the transmission of Filler on that service has ended. The minimum Filler transmission duration *should* be chosen to accommodate the tuning time of the Client-DPI Device. For example if the Client-DPI Device can complete a tune in 500 milliseconds or less the minimum Filler duration would be 533 milliseconds. In certain implementations, Filler transmission duration *may* be greater than 533 milliseconds based upon other system requirements, for example to accommodate time base discontinuities or to allow for changes in an encryption key.

8.6. Filler Presentation Duration

The presentation duration of Filler in video elementary streams is defined as the time from the PTS of the first presented video access unit in Filler to the PTS of the last presented video access unit in the Filler plus the presentation duration of the last presented video access unit in the Filler. The presentation duration of all Filler in video elementary streams ***shall*** be the same across all video elementary streams involved in a given Switch.

The presentation duration of Filler in audio elementary streams is defined as the time from the PTS of the first presented audio access unit in Filler to the PTS of the last presented audio access unit in the Filler plus the presentation duration of the last presented audio access unit in the Filler. The presentation duration of all Filler in audio elementary streams ***shall*** be the same across all audio elementary streams involved in a given Switch.

8.7. Addressable Content Set Duration

The presentation duration of the video within an Addressable Content Set is defined as the time from the PTS of the first presented video access unit in the Addressable Content Set to the PTS of the last presented video access unit in the Addressable Content Set plus the presentation duration of the last presented video access unit in the Addressable Content Set. The presentation duration of all video streams ***shall*** be the same across all video streams in an Addressable Content Set.

The presentation duration of the audio within an Addressable Content Set is defined as the time from the PTS of the first presented audio access unit in the Addressable Content Set to the PTS of the last presented audio access unit in the Addressable Content Set plus the presentation duration of the last presented audio access unit in the Addressable Content Set. The presentation duration of all audio streams ***shall*** be the same across all audio streams in an Addressable Content Set.

In the case where Filler is included in the encoded Addressable Content Set, these constraints apply only to the non-Filler portion, while the Filler portion is constrained as required in section 8.5.

Note: The presentation duration of video and audio in an Addressable Content Set might not be identical due to the fact that the access unit duration of video might not be the same as the access unit duration of audio.

8.8. MPEG-2 Video Coding Requirements

Following the Filler, the first transport packet carrying video elementary stream payload ***shall*** start with a PES header, and the payload of that PES packet ***shall*** start with a sequence header, and a GOP header followed by an I picture. In the first GOP header, the closed_gop bit ***shall*** be set to '1', indicating that this first GOP is closed. In order to accommodate changes to horizontal_size_value in the video stream between Filler and Addressable Content, a sequence end code ***shall*** be used at the end of the last access unit, in decode order, in the Filler.

8.9. AVC Video Coding Requirements

Following the Filler, content ***shall*** begin with an IDR access unit. The initial IDR access unit *should* have no_output_of_prior_pics_flag set to '1' in all slice headers within that access unit, unless there is another way to guarantee conformance with both [SCTE 128-1] or [SCTE 172].

8.10. HEVC Video Coding Requirements

Following the Filler, content ***shall*** begin with an IDR picture. The initial IDR picture should have no_output_of_prior_pics_flag set to '1' in all slice headers within that access unit, unless there is another way to guarantee conformance with both [SCTE 215-1] or [SCTE 172]. The initial IDR picture should be contained in an SHRAP access unit which meets the requirements as defined in [SCTE 215-1].

In Level 0 systems, HEVC video streams refers to SDR video encoded using HEVC compression. Transitions between SDR and HDR video streams and between HDR video streams compressed using HEVC is not defined.

8.11. DTS-UHD Audio Coding Requirements

Following the Filler, content ***shall*** begin with a Random Access Point as defined in [SCTE 243-4], with the exception that if the DTS-UHD BroadcastChunk [SCTE 242-4] is present, it ***shall*** appear in the main stream prior to the first synch frame. When multiple DTS-UHD streams are used to define playback, the Random Access Points of each of the component streams ***shall*** be aligned.

8.12. Level 0 Special Cases

There are several special cases of Level 0 systems that involve Switching from a Primary Channel to an addressable ad and back to the Primary Channel. This section defines those special cases.

8.12.1. DVR-Based

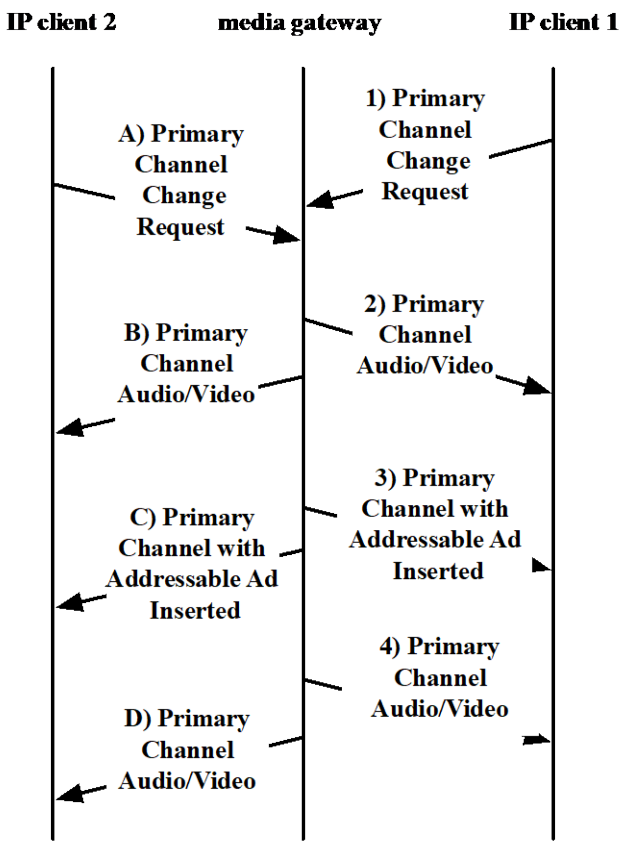
In DVR-based systems, the Primary Channel can either be live, or prerecorded. In these systems, there are no Insertion Channels because the addressable ads are stored on the DVR's hard drive. The Switch from Primary Channel to an addressable ad(s) ***shall*** be performed by selecting an addressable ad(s) by the Client-DPI Device's Decision Engine and Switching from the Primary Channel to an addressable ad(s) stored on the DVR's hard drive. The Client DPI Device ***shall*** Switch back to the Primary Channel after the addressable ad(s) have ended.

In order for Client-DPI Devices to perform a DVR-based Switch, the Client DPI Device ***shall*** find the appropriate addressable ad(s) on its hard drive; stop decoding the audio and video from the Primary Channel; start streaming, decoding and displaying the audio and video from the addressable ad(s); and ***shall*** Switch back to the Primary Channel after the addressable ads have ended. Since this type of Level 0 Switch could take some time on the Client-DPI Device, Filler *may* have to be added to the addressable ads stored on the DVR's hard drive. The amount of Filler to be chosen is cable system dependent and *should* be selected by each MVPD to mask audio and video artifacts during the Switch time for the set of Client-DPI Devices in the MVPD's cable system.

8.12.2. Media Gateway/IP Client

Media gateway/IP client systems are similar to DVR-based systems where the Primary Channel can either be live, or prerecorded. The difference in media gateway/IP client systems is that the storage for Primary Channels and addressable ads reside in the media gateway device and are accessed by the IP client device over a home network. The Switch from Primary Channel to an addressable ad(s) ***shall*** be performed by selecting an addressable ad(s) by the Decision Engine and Switching from the Primary Channel to an addressable ad(s) stored on the media gateway's hard drive. The media gateway ***shall*** Switch back to the Primary Channel after the addressable ad(s) have ended. In media gateway/IP client systems, the Decision Engine resides in the media gateway.

To perform a media gateway IP client Switch, the media gateway ***shall*** find the appropriate addressable ad(s) on the media gateway's hard drive; stop decoding the audio and video from the Primary Channel; start streaming, decoding and displaying the audio and video from the addressable ad(s); and ***shall*** Switch back to the Primary Channel after the addressable ad(s) playback has completed. Since this type of Level 0 Switch ***shall*** take place in the media gateway device and could take some time, Filler *may* have to be added to the addressable ads stored on the media gateway's hard drive. The amount of Filler to be chosen is cable system dependent and *should* be selected by each MVPD to mask audio and video artifacts during the Switch time.



1) IP client 1 sends a request to the media gateway to Switch to a new Primary Channel.

2) Media gateway performs the acquisition of the new Primary Channel and provides the new audio/video to the IP client 1.

3) Addressable ad insertion is performed by the media gateway with no interaction with IP client 1. The decision engine runs on the media gateway for all connected clients. The addressable ad audio and video is provided to IP client 1 after the Switch is performed from the Primary Channel to the addressable ad stored in the media gateway.

4) The media gateway Switches back to the Primary Channel after the addressable ad(s) have ended and provides the new audio/video to the IP client 1.

A) IP client 2 sends a request to the media gateway to Switch to a new Primary Channel.

B) Media gateway performs the acquisition of the new Primary Channel and provides the new audio/video to the IP client 2.

C) Addressable ad insertion is performed by the media gateway with no interaction with IP client 2. The decision engine runs on the media gateway for all connected clients. The addressable ad audio and video is provided to IP client 2 after the Switch is performed from the Primary Channel to the addressable ad stored in the media gateway.

D) The media gateway Switches back to the Primary Channel after the addressable ad(s) have ended and provides the new audio/video to the IP client 2.

Figure 1 – Media Gateway IP Client Communications

8.12.3. Connected TV/IP Devices

In a connected TV or IP device where the Switch from a Primary Channel to an addressable ad carried by a different transport stream and back to the Primary Channel is performed in the Client-DPI Device, there are several systems issues to consider when choosing a Filler. In these types of systems, the Primary Channel could be provided via a transport stream or streamed over an IP network. The same is true for the addressable ads. There may be some latency associated with Switching from a Primary Channel carried by a transport stream or streamed over an IP network to an addressable ad carried by a different transport stream or streamed over an IP network. Filler must be chosen in order to mask any Switch latency that could be present in either connected TV, or IP devices.

9. Level 1 Requirements

Level 1 Conditioning consists of a Gap, as defined in section 9.3, and specific MPEG coding requirements defined in sections 9.4 and 9.5. This Conditioning ***shall*** be present in the transport stream at the input to a Client-DPI Device when a Level 1 Switch is enabled. In this context, “enabled” means that a Client-DPI Device *may*, as determined by its Decision Engine, perform a Level 1 Switch. The following requirements for Level 1 apply to both standard definition and high definition coded video streams.

9.1. Use of Transport Streams

In Level 1 systems, a Switch ***shall*** be limited to elementary streams within the same MPTS.

9.2. Use of Services

PMTs within the MPTS ***shall*** be constructed such that all elementary streams that are selected as the result of a Level 1 Switch, including a video-only or audio-only Switch, ***shall*** reside in the same service.

9.3. The Gap

When a Level 1 Switch is to be enabled, a Gap ***shall*** be inserted into all coded video elementary streams that *may* be Switched away from or Switched to. There ***shall*** exist a minimum 10 millisecond period (as measured in transmission time) that is wholly contained within all of the video Gaps. A separate Gap ***shall*** be inserted into all coded audio elementary streams that *may* be Switched away from or Switched to. There ***shall*** exist a minimum 10 millisecond period (as measured in transmission time) that is wholly contained within all of the audio Gaps.

During the Gap, there ***shall*** be no transmission of transport packets containing PES headers or PES data in the video or audio elementary streams that *may* be Switched away from or Switched to.

Note: For a given Switch opportunity, the Gap in video streams will not necessarily be aligned with the Gap in audio streams. The offset between them will be equal to the difference in decoder buffer delay between video and audio, which is a function of a number of encoding parameters affecting both the video and audio streams.

9.4. MPEG Coding Requirements

9.4.1. The System Clock and Program Clock Reference

A Switch from one stream to another (of the same type) during the Gap ***shall*** maintain the continuity of the system clock in a Client-DPI Device. All of the services comprising an Addressable Content Set ***shall*** either reference a common PCR_PID or individual services *may* each contain a different PCR_PID all of which ***shall*** carry the same program clock.

9.4.2. MPEG Syntax

A Switch from one stream to another (having the same stream_type) during the Gap in a Client-DPI Device ***shall not*** cause a discontinuity in or non-compliance of MPEG syntax at any layer, except for the continuity_counter in the transport packet header.

9.4.3. Constraints on MPEG-2 Video Streams Adjacent to a Gap

In video streams, the last byte of the payload of the transport packet transmitted prior to the Gap ***shall*** be the last byte of a video access unit and the last byte of a PES packet.

The last picture in presentation order prior to a Gap ***shall*** be either a P or an I picture.

Prior to a Gap, the value of DTS for the last decoded picture ***shall*** be the same in all video streams that *may* be Switched from. The value of PTS for the last presented picture ***shall*** be the same in all video streams that *may* be Switched from.

To accommodate horizontal resolution changes after the Gap, the last access unit prior to the Gap ***shall*** end with a sequence end code.

Following a video Gap, the first transport packet carrying a payload ***shall*** start with a PES header. The payload of that PES packet ***shall*** start with a sequence header, sequence extension, and GOP header.

The sequence header fields vertical_size_value, aspect_ratio_information, frame_rate, progressive_sequence, and constrained_parameters_flag ***shall*** contain values identical to those fields within the last sequence header transmitted prior to the Gap. The field horizontal_size_value is permitted to change within the constraints specified by [SCTE 43]. The sequence extension ***shall*** be identical to the last sequence extension transmitted prior to the Gap. Field parity ***shall*** be maintained across the Gap.

In the first GOP header, the closed_gop bit ***shall*** be set to '1', indicating that this first GOP is closed. The first coded picture in the first GOP ***shall*** be an I picture.

Following a Gap, the first video transport packet carrying payload ***shall*** start with a PES header and the start of a video access unit. The values of PTS of the first video access unit in all of the MPEG-2 video streams that *may* be selected as the result of a Switch ***shall*** be the same. The value of that PTS ***shall*** be equal to the PTS of the last video access unit prior to the Gap plus the presentation duration of that video access unit.

The first picture in decode order following a Gap ***shall*** have a DTS such that the decoding of this picture follows the decoding of the last picture decoded prior to the Gap at the proper time as determined by the access unit duration.

Following a Gap, the values of DTS of the first coded picture in all video streams that *may* be selected as the result of a Switch ***shall*** be the same. The value of PTS for the first presented picture in all video streams that *may* be selected as the result of a Switch ***shall*** be the same.

Note: Changes in bar data and/or active format description (AFD) data across the Gap *may* affect the presentation of selected content at the Client-DPI Device.

9.4.4. Constraints on AVC Video Streams Adjacent to a Gap

In video streams, the last byte of the payload of the transport packet transmitted prior to the Gap ***shall*** be the last byte of a video access unit and the last byte of a PES packet.

Prior to a Gap, the value of DTS for the last decoded picture ***shall*** be the same in all video streams that *may* be Switched from.

When the last access unit prior to the Gap is decoded, all of the pictures in the DPB which are not yet output (displayed/presented) ***shall*** be, starting immediately, contiguously displayable (no discontinuity in their PTS values).

The value of PTS for the last presented picture in all video streams that *may* be Switched from ***shall*** be the same.

Following a video Gap, the first transport packet carrying a payload ***shall*** start with a PES header. The payload of that PES packet ***shall*** start with an SRAP containing an IDR access unit constrained by [SCTE 172] and [SCTE 128-1]. The SPS and VUI parameters of this IDR access unit ***shall*** be the same as the SPS and VUI parameters of coded video sequence transmitted prior to the Gap except for the field 'PicWidthInMBs' as constrained by SCTE 128-1. The no_output_of_prior_pics_flag in this IDR access unit ***shall*** be set to '0'.

Following a Gap, the first video transport packet carrying payload ***shall*** start with a PES header and the start of a video access unit. The values of PTS of the first video access unit in all of the AVC video streams that *may* be selected as the result of a Switch ***shall*** be the same. The value of that PTS ***shall*** be equal to the PTS of the last video access unit prior to the Gap plus the presentation duration of that video access unit.

The first picture in decode order following a Gap ***shall*** have a DTS such that the decoding of this picture follows the decoding of the last picture decoded prior to the Gap at a proper time as determined by the access unit duration.

Following a Gap, the values of DTS of the first coded picture in all video streams that *may* be selected as the result of a Switch ***shall*** be the same. The value of PTS for the first presented picture in all video streams that *may* be selected as the result of a Switch ***shall*** be the same.

Note: Changes in bar data and/or active format description (AFD) data across the Gap *may* affect the presentation of selected content at the Client-DPI Device.

9.4.5. Constraints on HEVC Video Streams Adjacent to a Gap

In video streams, the last byte of the payload of the transport packet transmitted prior to the Gap ***shall*** be the last byte of a video access unit and the last byte of a PES packet.

Prior to a Gap, the value of DTS for the last decoded picture ***shall*** be the same in all video streams that may be switched from.

When the last access unit prior to the Gap is decoded, all of the pictures in the DPB which are not yet output (displayed/presented) ***shall*** be, starting immediately, contiguously displayable (no discontinuity in their PTS values).

The value of PTS for the last presented picture in all video streams that may be switched from ***shall*** be the same.

Following a video Gap, the first transport packet carrying a payload ***shall*** start with a PES header. The payload of that PES packet ***shall*** start with an SHRAP access unit constrained by [SCTE 215-1]. The VPS, SPS, and VUI parameters of this SHRAP ***shall*** be the same as the VPS, SPS, and VUI parameters of coded video sequence transmitted prior to the Gap except for the field 'pic_width_in_luma_samples' as constrained by [SCTE 215-1]. The no_output_of_prior_pics_flag in this SHRAP access unit ***shall*** be set to '0'.

Following a Gap, the first video transport packet carrying payload ***shall*** start with a PES header and the start of a video access unit. The values of PTS of the first video access unit in all of the HEVC video streams that may be selected as the result of a Switch ***shall*** be the same. The value of that PTS ***shall*** be equal to the PTS of the last video access unit prior to the Gap plus the presentation duration of that video access unit.

The first picture in decode order following a Gap ***shall*** have a DTS such that the decoding of this picture follows the decoding of the last picture decoded prior to the Gap at a proper time as determined by the access unit duration.

Following a Gap, the values of DTS of the first coded picture in all video streams that may be selected as the result of a Switch ***shall*** be the same. The value of PTS for the first presented picture in all video streams that may be selected as the result of a Switch ***shall*** be the same.

In Level 1 systems, HEVC video streams refers to SDR video encoded using HEVC compression. Transitions between SDR and HDR video streams and between HDR video streams compressed using HEVC is not defined.

Note: Changes in bar data and/or active format description (AFD) data across the Gap may affect the presentation of selected content at the receiver.

9.4.6. Constraints on AC-3 or E-AC-3 Audio Streams Adjacent to a Gap

In audio streams, the last byte of the payload of the transport packet transmitted prior to the Gap ***shall*** be the last byte of an audio access unit and the last byte of a PES packet.

Preceding a Gap, the values of PTS of the last presented audio access unit in all of the audio streams that *may* be Switched from ***shall*** be the same.

Following a Gap, the first audio transport packet carrying payload ***shall*** start with a PES header and the start of an audio access unit. The values of PTS of the first audio access unit in all of the audio streams that *may* be selected as the result of a Switch ***shall*** be the same. The value of that PTS ***shall*** be equal to the PTS of the last audio access unit prior to the Gap plus the presentation duration of that access unit.

9.4.7. Constraints on AAC Audio Streams Adjacent to a Gap

In audio streams, the last byte of the payload of the transport packet transmitted prior to the Gap ***shall*** be the last byte of an audio access unit and the last byte of a PES packet.

Preceding a Gap, the values of PTS of the last presented audio access unit in all of the audio streams that *may* be Switched from ***shall*** be the same.

Following a Gap, the first audio transport packet carrying payload ***shall*** start with a PES header and an audio Random Access Point (RAP) as defined in [SCTE 193-1].

The values of PTS of the first audio access unit in all of the audio streams that *may* be selected as the result of a Switch ***shall*** be the same. The value of that PTS ***shall*** be equal to the PTS of the last audio access unit prior to the Gap plus the presentation duration of that access unit.

9.4.8. Constraints on AC-4 Audio Streams Adjacent to a Gap

In audio streams, the last byte of the payload of the transport packet transmitted prior to the Gap ***shall*** be the last byte of an audio access unit and the last byte of a PES packet.

Preceding a Gap, the values of PTS of the last presented audio access unit in all of the audio streams that *may* be Switched from ***shall*** be the same.

Following a Gap, the first audio transport packet carrying payload ***shall*** start with a PES header and the start of an audio access unit. The values of PTS of the first audio access unit in all of the audio streams that *may* be selected as the result of a Switch ***shall*** be the same. The value of that PTS ***shall*** be equal to the PTS of the last audio access unit prior to the Gap plus the presentation duration of that access unit.

9.4.9. Constraints on MPEG-H Audio Streams Adjacent to a Gap

In audio streams, the last byte of the payload of the transport packet transmitted prior to the Gap ***shall*** be the last byte of an audio access unit and the last byte of a PES packet.

Preceding a Gap, the values of PTS of the last presented audio access unit in all of the audio streams that *may* be Switched from ***shall*** be the same.

Following a Gap, the first audio transport packet carrying payload ***shall*** start with a PES header and an audio Random Access Point (RAP) as defined in [SCTE 243-3].

The values of PTS of the first audio access unit in all of the audio streams that *may* be selected as the result of a Switch ***shall*** be the same. The value of that PTS ***shall*** be equal to the PTS of the last audio access unit prior to the Gap plus the presentation duration of that access unit.

9.4.10. Constraints on DTS-HD Audio Streams Adjacent to a Gap

In audio streams, the last byte of the payload of the transport packet transmitted prior to the Gap ***shall*** be the last byte of an audio access unit and the last byte of a PES packet.

Preceding a Gap, the values of PTS of the last presented audio access unit in all of the audio streams that *may* be Switched from ***shall*** be the same.

Following a Gap, the first audio transport packet carrying payload ***shall*** start with a PES header and the start of an audio access unit. The values of PTS of the first audio access unit in all of the audio streams that *may* be selected as the result of a Switch ***shall*** be the same. The value of that PTS ***shall*** be equal to the PTS of the last audio access unit prior to the Gap plus the presentation duration of that access unit.

9.4.11. Constraints on DTS-UHD Audio Streams Adjacent to a Gap

In audio streams, the last byte of the payload of the transport packet transmitted prior to the Gap ***shall*** be the last byte of an audio access unit and the last byte of a PES packet.

Preceding a Gap, the values of PTS of the last presented audio access unit in all of the audio streams that *may* be Switched from ***shall*** be the same.

Following a Gap, the first audio transport packet carrying payload ***shall*** start with a PES header and the start of an audio access unit. The values of PTS of the first audio access unit in all of the audio streams

© 2024 SCTE

that *may* be selected as the result of a Switch ***shall*** be the same. The value of that PTS ***shall*** be equal to the PTS of the last audio access unit prior to the Gap plus the presentation duration of that access unit.

Following a Gap, content shall begin with a Random Access Point as defined in [SCTE 243-4], with the exception that if the DTS-UHD BroadcastChunk [SCTE 242-4] is present, it shall appear in the main stream prior to the first synch frame.

When multiple DTS-UHD streams are used to define playback, the Random Access Points of each of the component streams ***shall*** be aligned.

9.5. Buffer Management

At each Gap, all possible Switches ***shall not*** result in any buffer or time base discontinuities, and ***shall*** maintain full compliance with the T-STD model.

Appendix A: System Overview (Informative)

This standard enables a form of addressable content delivery in digital television systems by conditioning video and audio elementary streams in an MPEG-2 transport stream. The following is an example application of telecast addressable advertising that employs this Conditioning.

In this telecast addressable advertising implementation, multiple advertisement streams are simultaneously delivered to a Client-DPI Device. The Client-DPI Device selects one of the advertisements based upon received addressing information and locally stored selection criteria, and Switches at the appropriate time. The Switch may be accomplished in a non-seamless manner, herein referred to as Level 0 (L0) Switching, or in a seamless manner, herein referred to as Level 1 (L1) Switching, as described below.

An addressable advertising system may employ either or both Level 0 and/or Level 1 Switching.

A.1 Client-DPI Device Functionality

In order for a Receiver to be Client-DPI capable, specific functionalities herein referred to as the Decision Engine and Switching Engine are present on that Receiver.

The Decision Engine implements a methodology for determining which stream(s) should be selected for presentation at the time of each Switch.

The Switching Engine implements Level 0 and/or Level 1 Switching, and provides an interface that enables a Decision Engine to cause the required Switching to take place.

A.2 The MPEG-2 Transport Stream

The transmission of addressable advertising within a digital cable television system is in compliance with MPEG-2 and SCTE standards. This section reviews some of the basic concepts, the knowledge of which is fundamental to understanding how addressable advertising systems might operate.

Typically, a number of separate television channels are combined in a single transport stream, as described in [13818-1]. The MPEG-2 transport stream is a packetized multiplex. Packets are fixed in length, and are comprised of a header portion and a payload portion. Video and audio coded data from each elementary stream within the multiplex is carried in the payload portion. The header portion carries, among other items a unique PID value that indicates to which elementary stream the contents of the payload portion belongs.

In addition to the data comprising encoded video and audio information for each television channel, the transport stream carries System Information streams which describe the video and audio data and enable a Client-DPI Device to select the proper streams to be decoded based on channel selection. There may also be data associated with applications, sometimes specific to the television channels contained within the multiplex, and there may be data transmitted to support decryption of encrypted video and audio elementary streams.

In accordance with the MPEG-2 Systems standard [6], the transmission of video, audio, and data is timed to insure that data from each stream is available at the Client-DPI Device in time for it to be decoded without overflowing that stream's buffer in the Client-DPI Device. Null packets may be used in the stream as necessary to satisfy QAM modulator requirements. The result, as seen at the input to the Client-DPI Device, is a continuous stream of packets.

 © 2024 SCTE

A.3 Level 0 Systems

A.3.1 Level 0 Stream Conditioning

Level 0 systems permit addressable advertisement streams designated for insertion into a particular Primary Channel to be transmitted in an MPEG-2 transport multiplex different from the one that is carrying the Primary Channel. The Client-DPI Device, then, acquires the alternate multiplex before it can select and decode the required addressable advertisement. Typically, this will involve tuning to the RF channel carrying the multiplex that carries the addressable advertisements.

Figure 2 below depicts portions of two different MPEG-2 transport stream multiplexes showing the presence of a single Level 0 Addressable Content Set, comprised of four addressable advertisements. Also shown are the time boundaries within which the transport packets containing video, audio and data associated with the Addressable Content Set may be transmitted. Packets carrying streams not related to the Addressable Content Set, including null packets, would also be present in a real system. For clarity, this unrelated data is not shown in the Figure.

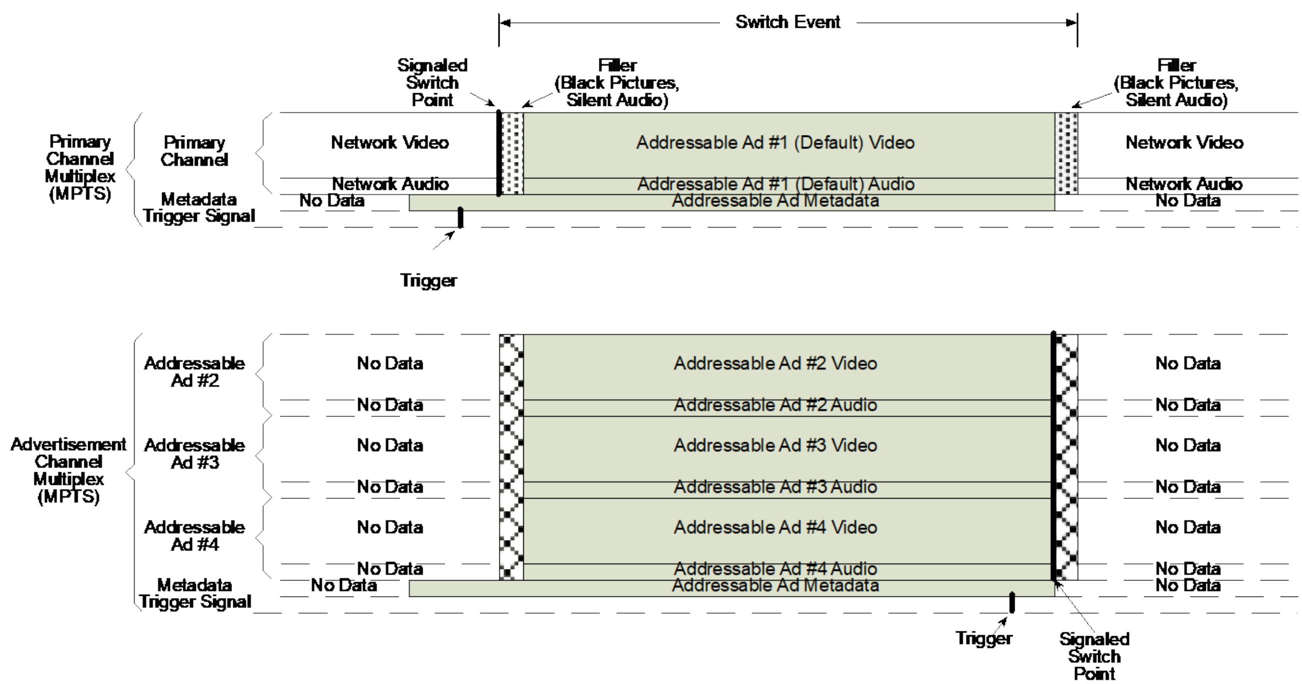


Figure 2 - Example Level 0 Transport Multiplex

In this example, one of the addressable advertisements, designated as the default advertisement, is present in the Primary Channel and exists on the same PIDs as the Network's video and audio PIDs. The default addressable advertisement will be displayed by Client-DPI Devices that do not implement Level 0 Switching. Also, the default addressable advertisement will be selected by Level 0 capable Client-DPI Devices that, as determined by the Decision Engine, do not respond to selection criteria for this Addressable Content Set. The remaining addressable advertisements are present in the advertisement channel multiplex.

An opportunity to Switch from the Primary Channel video and audio streams is shown as a Signaled Switch Point at the start of the Addressable Content Set. Another Signaled Switch Point marks the opportunity to Switch back to the Primary Channel video and audio streams at the end of the Addressable Content Set. A transmission Filler follows each of these Signaled Switch Points. During these Fillers, the

Client-DPI Device will tune to and begin decoding the new multiplex. Properly formatted picture data for black pictures, and audio data that represents silent audio, are transmitted during this time. This insures that the Client-DPI Device will be able to acquire and begin decoding the new content without displaying any artifacts or losing any significant video or audio data as defined in OC-SP-HOST2.1 [14]. The duration of Filler is selected to allow for the time required for the slowest Client-DPI Device to tune to and begin decoding the new multiplex.

The Signaled Switch Points and Filler occurring in both the Primary Channel and advertisement channel multiplexes are aligned in time. Similarly, the transmission of addressable advertisement #1 in the Primary Channel multiplex is aligned in time with the transmission of addressable advertisements #2, #3, and #4 in the advertisement channel multiplex. This standard specifies the stream properties at the receiving Client device. The preparation of Filler for the Primary Channel may be performed either in the head-end or at the network feed origination site. When Filler is present in the Primary Channel it will always be entirely within Switch Event boundaries.

A single Trigger Signal is associated with each Signaled Switch Point. The Trigger Signal indicates the location of the Signaled Switch Point in time. Even though Figure 2 shows the Trigger Signal as data that is not contained within the video or audio streams, other methods of providing the Trigger Signal not shown here may be used.

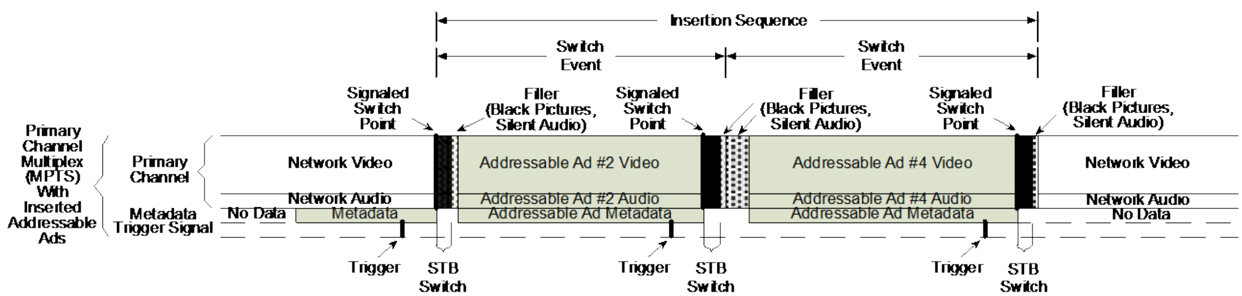


Figure 3 - Back to Back Level 0 Addressable Insertions

Figure 3 illustrates a back-to-back Level 0 addressable insertion. This diagram shows the result of 2 addressable ads that have been selected by a Client-DPI Device.

In this example, Insertion Channels 1, 2 and 3 contained the following addressable advertisements:

Insertion Channel 1	Addressable Ad #1	Addressable Ad #4
Insertion Channel 2	Addressable Ad #2	Addressable Ad #5
Insertion Channel 3	Addressable Ad #3	Addressable Ad #6

This figure shows the bit stream being presented to the Client-DPI Device's decoder after Switching. As a result of Switches, Addressable Ads 2 and 4 were selected from two separate Insertion Channels: 2 and then 1. There is no data during the STB Switch before the Client-DPI Device acquires valid data in the next Insertion Channel or Primary Channel.

A.3.2 Level 0 Head-End

Figure 4 and Figure 5 below show hypothetical head-end architectures suitable for inserting Level 0 Addressable Content Sets.

In Figure 4, the splicer inserts addressable advertisements using both the Primary Channel multiplex and the advertisement channel multiplex, as well as advertisement selection criteria and Trigger Signals.

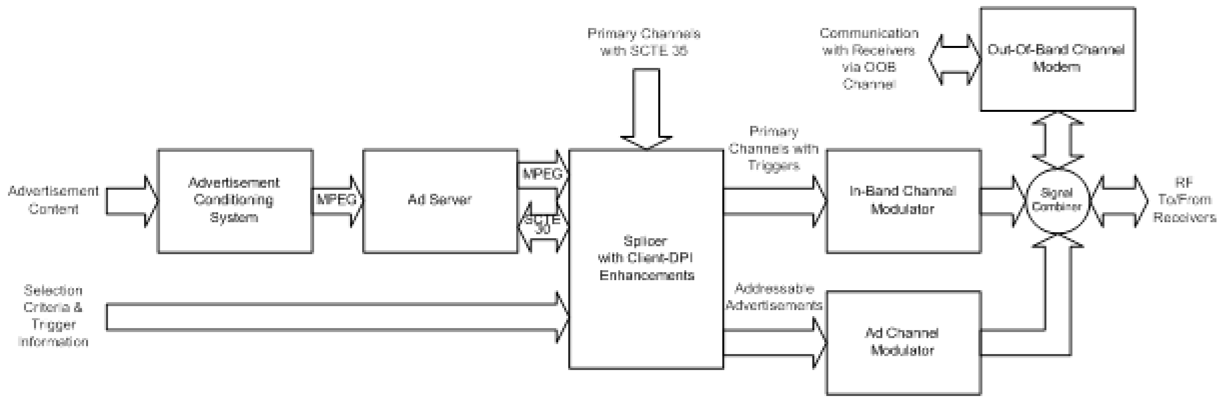


Figure 4 - Example Level 0 Head-End Configuration #1

In the example shown in Figure 5, the Client Edge Device is a specialized device for implementing certain types of client-based systems. It can handle all signaling, can stream out addressable advertisements on a dedicated advertisement channel, interface with cable plant to negotiate and acquire bandwidth, and perform basic conditioning on addressable advertisements and network feeds to enable Level 0 Switching.

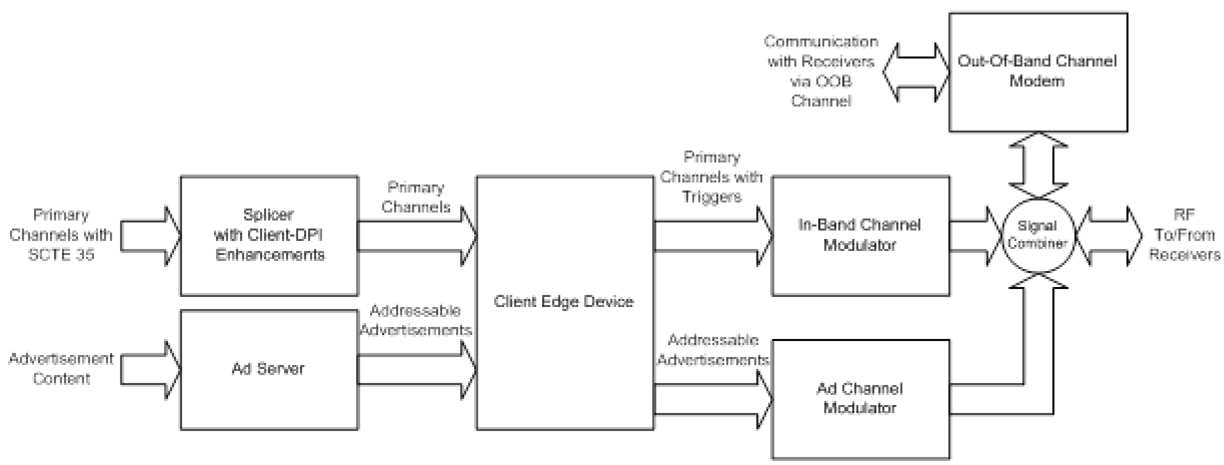


Figure 5 - Example Level 0 Head-End Configuration #2

A.3.3 Level 0 Client-DPI Device

Figure 6 below depicts a hypothetical Client-DPI Device architecture suitable for use in a Level 0 system. The Decision Engine and Switching Engine are shown, along with relevant signal flows. The Trigger Signal is assumed to be in the form of a data message carried on a separate stream from the streams carrying the video and audio data.

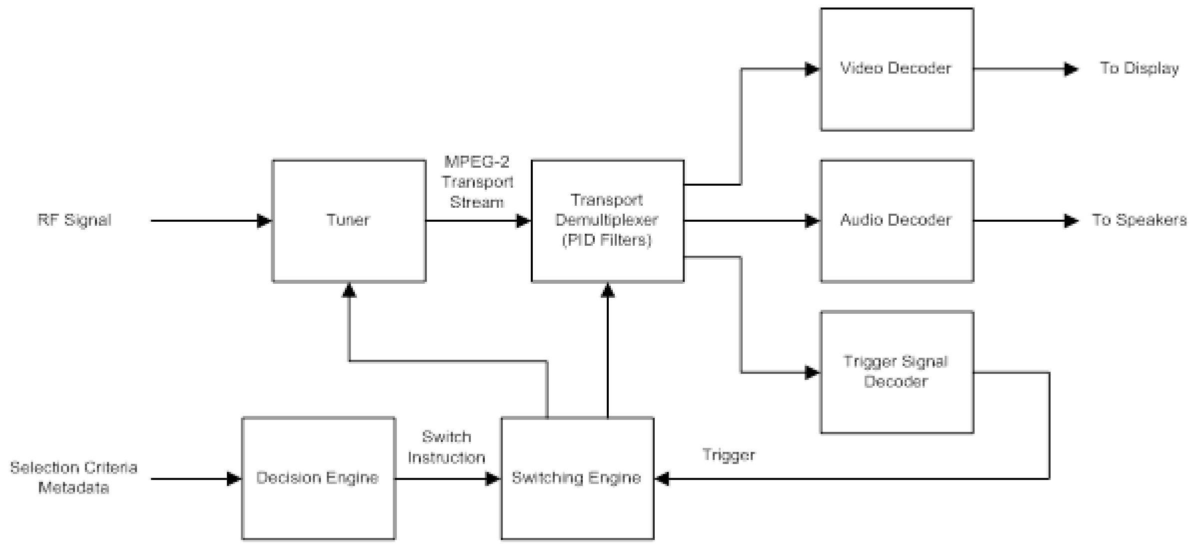


Figure 6 - Example Level 0 Set-Top Decoder Architecture

In this Figure, the selection of the required transport multiplex is performed by the tuner, and selection of the required addressable advertisement within the multiplex is performed by the demultiplexer. There are three registers within the demultiplexer; one that holds the PID value for transport stream packets containing coded picture data that is to be forwarded to the video decoder, another that holds the PID value for transport stream packets containing coded audio data that is to be forwarded to the audio decoder, and a third that holds the PID value for transport stream packets containing Trigger Signal data that is to be forwarded to the Trigger Signal decoder. The Switching Engine directs the tuner to the proper RF channel, and then selects the streams comprising the addressable advertisement by loading the proper PID values into these registers. The Trigger Signal arrives sufficiently early to permit the Client-DPI Device to process it, and the presence of Filler following the Switch Point ensures that there is time for the Client-DPI Device to perform this operation without losing important video and audio data. Implementations may utilize any appropriate methods to ensure that these timing requirements are met, including communicating PID values and stream_types in a manner that eliminates the need to parse the PSI. (See section 7.0)

To provide a transparent viewing experience, there may be other software in the Client-DPI Device that maintains a state associated with the currently tuned Primary Channel (for example, to display the channel number on the front panel). In such a case, to avoid confusing the viewer, it is necessary that the Switching Engine be able to control the tuner and the demultiplexer without altering this state. Worst case Level 0 Switching occurs when Switching across 2 separate frequencies although Level 0 Switches can occur within a single multiplex.

A.4 Level 1 Systems

A.4.1 Level 1 Stream Conditioning

In this example of a Level 1 addressable advertising system, all of the addressable advertisement streams comprising an Addressable Content Set for a particular Primary Channel are present in the same MPEG-2 transport stream multiplex as the Primary Channel. The Primary Channel stream and addressable advertisement streams are conditioned to enable the Client-DPI Device to Switch to the selected advertisement seamlessly.

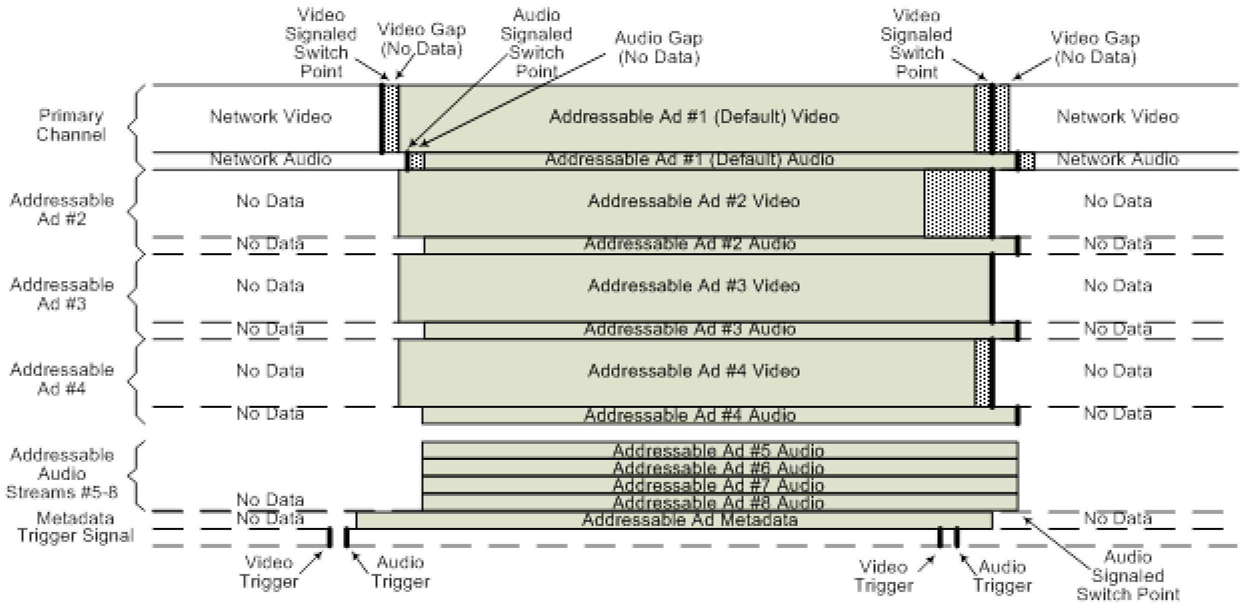


Figure 7 - Example Level 1 Transport Multiplex

Figure 7 depicts a portion of an MPEG-2 MPTS showing the insertion of a single Level 1 Addressable Content Set, comprised of four addressable advertisements. Also shown are the time boundaries within which transport packets containing video, audio and data associated with the Addressable Content Set may be transmitted. Packets carrying streams not related to the Addressable Content Set, including null packets, would also be present in a real system. For clarity, this unrelated data is not shown in the Figure.

One of the addressable advertisements is designated as the default and is present in the Primary Channel and exists on the same PIDs as the network's video and audio PIDs. The default addressable advertisement will be displayed by Client-DPI Devices that do not implement Level 1 Switching. Also, the default addressable advertisement will be selected by Level 1 capable Client-DPI Devices that, as determined by the Decision Engine, do not respond to the selection criteria for this Addressable Content Set.

The Signaled Switch Point represents an opportunity to Switch streams. Because the transmission of video and audio intended to be presented simultaneously may occur at slightly different times, two Signaled Switch Points are indicated, one for video streams and one for audio streams. An opportunity to Switch from the Primary Channel video and audio streams is shown as a pair of Signaled Switch Points at the start of the Addressable Content Set. Another pair of Signaled Switch Points marks the opportunity to Switch back to the Primary Channel video and audio streams at the end of the Addressable Content Set. In Level 1 systems, the Switching of audio streams without Switching video streams is also possible as illustrated by addressable audio streams 5-8 in Figure 7. A transmission Gap follows each of these Signaled Switch Points. During these Gaps, no transport packets carrying PES headers or data for any of the component streams comprising the Addressable Content Set are transmitted. This Gap insures that the Client-DPI Device will not Switch while it is receiving data associated with the Addressable Content Set. The duration of the Gap and its location following the Signaled Switch Point are selected to allow for a small amount of uncertainty in the timing of the stream selection process within the Client-DPI Device. The transport stream is a continuous data stream; therefore, packets carrying data from streams not related to the Addressable Content Set, or null transport packets, would be transmitted during the Gap to maintain the continuity of the stream.

A Trigger Signal is associated with each Signaled Switch Point. The Trigger Signal indicates the location of the Signaled Switch Point in time. Even though this Figure shows the Trigger Signal as data that is not contained within the video or audio streams, other methods of providing the Trigger Signal not shown here may be used.

At the start of the Addressable Content Set, the start of transmission of all of the addressable advertisement video streams is aligned in time. Likewise, the start of all of the advertisement audio streams is aligned in time, and that time may differ from the start of the video streams. This alignment is enforced to make it possible for the Client-DPI Device to Switch from the Primary Channel video and audio streams to the selected addressable advertisement video and audio streams without loss of data.

It should be noted that each of the video streams within the Addressable Content Set might require a different amount of time to be transmitted. The transmission of the Primary Channel video and audio streams is able to resume only after the transmission of all of the addressable advertisement video and audio streams is complete. Addressable advertisement #3 contains the video stream having the longest transmission time, and therefore determines the location of the Signaled Switch Point, and when the Gap can begin. Packets carrying data from streams not related to the Addressable Content Set, or null transport packets, would be transmitted during this time.

A.4.2 Level 1 Head-End

Figure 8 shows an example of a head-end architecture suitable for inserting Level 1 Addressable Content Sets.

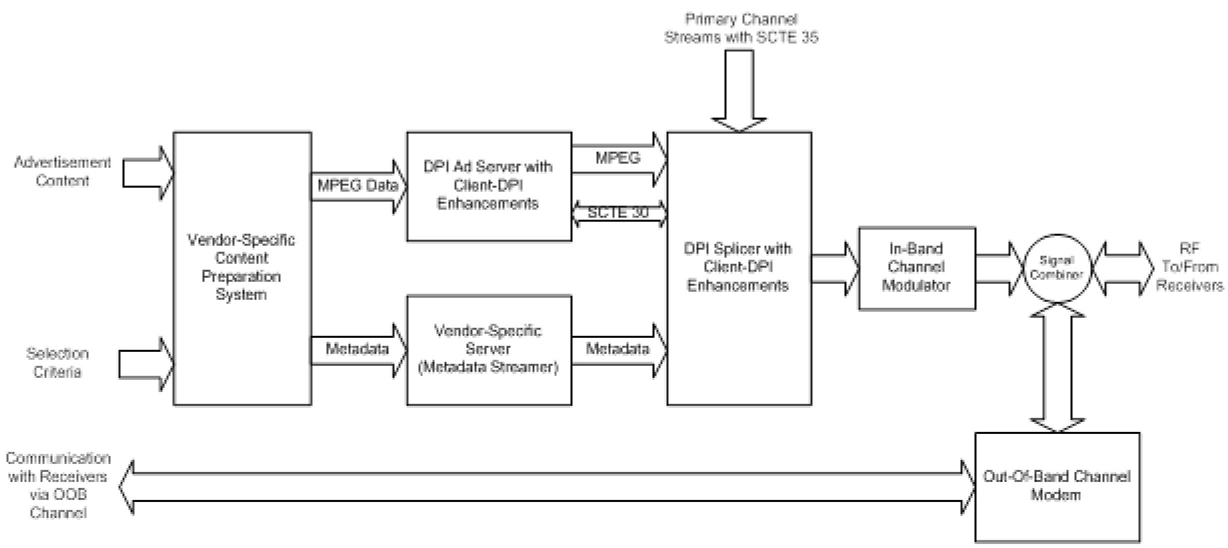


Figure 8 - Example Level 1 Head-End Configuration

In Figure 8, the insertion of Addressable Content Sets and Trigger Signals is handled by the splicer. Some of the conditioning of the video and audio streams within the Addressable Content Set may be done by the content preparation system, and some may be done by the splicer.

Insertion of selection criteria is performed by the metadata streamer and the splicer. The metadata streamer transmits packetized selection criteria metadata to the splicer when that data is to be inserted into the multiplex. The splicer inserts that data into the multiplex on an opportunistic basis, using a PID set aside for that purpose. There is sufficient bandwidth reserved in the multiplex to allow the opportunistic insertion to take place in a timely manner.

Command and control signaling, typically delivered using an out-of-band channel, may be used to configure and deliver addressing criteria to Decision Engines.

Note that the conditioning functions to implement a Level 1 Switch are very similar to the functionality in existing DPI head-end splicing devices. The goal of this conditioning is to insure, as much as possible, that no video or audio artifacts are produced at the Client-DPI Device. Some of the features of this conditioning are:

- Conformance with appropriate standards (ATSC, MPEG, and SCTE) with respect to transport, video, audio, buffer management and system clock is maintained across the splice.
- Vertical resolution and presentation duration of inserted content match those of the Primary Channel. Note: It is intended that the horizontal resolution of the Primary Channel will not be altered except for the Ads placed within the allowed Switch Event within the Primary Channel.
- Field parity is maintained across the splice.
- There are no timebase discontinuities.
- The number of pictures in the inserted content is adjusted, if necessary, to be consistent with the time slot into which it is being inserted. This might involve dropping pictures if the inserted content is too long, or adding black pictures if the inserted content is too short.
- The duration of audio in the inserted content is adjusted, if necessary, to be consistent with the time slot into which it is being inserted.
- The number of bits used to encode pictures and the timing of the transmission of video and audio data are adjusted so that the buffers in the Client-DPI Device do not underflow or overflow at any time.
- Video data transmitted following a splice can be decoded without referring to the video data transmitted prior to the splice.
- No portion of an audio access unit transmitted before a splice overlaps in presentation time with audio access units transmitted after that splice.
- Additional conditioning for addressable advertisements within Level 1 Addressable Content Sets is required to insure that the Client-DPI Device is able to select one of the addressable advertisements from the Addressable Content Set without introducing visible or audible artifacts. This includes:
- Assuring that the conditioning described above is applied to all of the addressable advertisement streams.
- Assuring that the same PCR, PTS and DTS adjustments are applied to all of the addressable advertisement streams.
- Inserting Trigger Signals.
- Inserting the Gap.

A.4.3 Level 1 Client-DPI Device

Figure 9 depicts a hypothetical Client-DPI Device architecture suitable for use in a Level 1 system. The Decision Engine and Switching Engine are shown, along with relevant signal flows. The Trigger Signal is assumed to be in the form of a data message on a stream separate from those carrying the video and audio elementary streams.

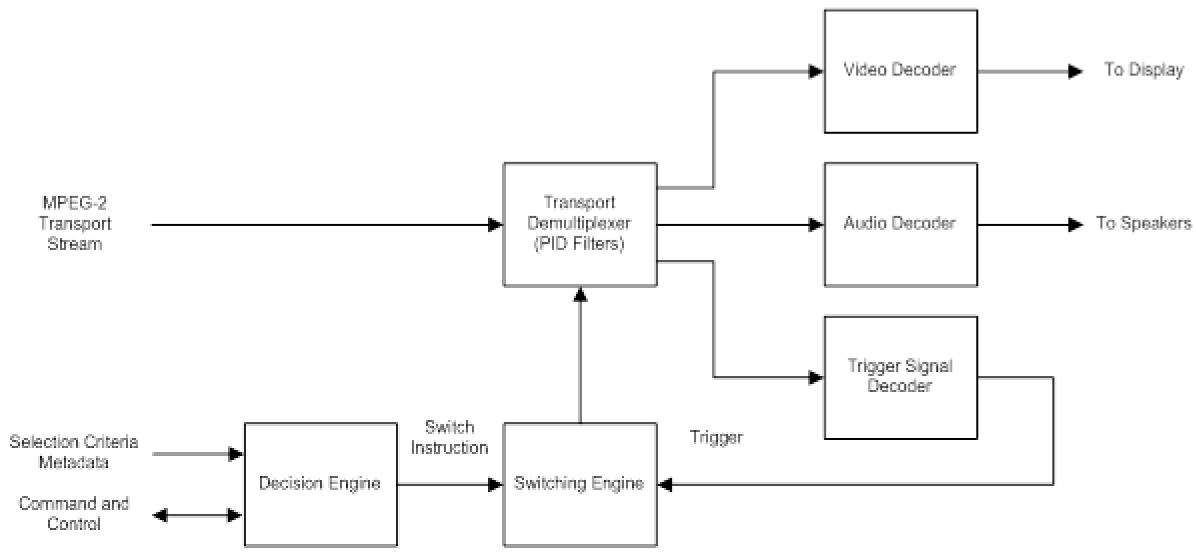


Figure 9 - Example Level 1 Set-Top Decoder Architecture

In Figure 9, the selection of the required addressable advertisement is performed by the demultiplexer. There are three registers within the demultiplexer; one that holds the PID value for transport stream packets containing data to be forwarded to the video decoder, another that holds the PID value for transport stream packets containing data to be forwarded to the audio decoder, and a third that holds the PID value for transport stream packets containing data to be forwarded to the Trigger Signal decoder. The Switching Engine causes the required streams from the Addressable Content Set to be selected by loading the proper PID values into the first two of these registers. The Trigger Signal arrives sufficiently early to permit the Client-DPI Device to process it, and the presence of the Gap surrounding the Switch Point insures that there is time for the Client-DPI Device to perform this operation without disrupting the flow of data into the video and audio decoders.

It can be seen that, in this architecture, the video and audio decoders are not involved in the Switching process. The objective of stream conditioning is to insure that, in ways relevant to decoding the affected streams, the signal supplied to these decoders appears to be continuous, as if no Switch had taken place.

Aspects of Decision Engine operation may be configured via command and control signaling.

To provide a transparent viewing experience, there may be other software in the Client-DPI Device that maintains a state associated with the currently tuned Primary Channel (for example, to display the channel number on the front panel). In such a case, to avoid confusing the viewer, it is necessary that the Switching Engine be able to control the demultiplexer without altering this state.

A.5 Signaling

Three distinct types of signaling may be employed in an addressable advertising system. One type of signaling provides the Client-DPI Device with selection criteria associated with each Addressable Content Set, so that the Client-DPI Device can select the proper content to present to the viewer. A separate signal, the Trigger Signal, indicates the location of each Signaled Switch Point. Finally, there may be command and control signaling used to control the behavior of the Decision Engine.

 © 2024 SCTE

A.5.1 Selection Criteria Data

Selection criteria data is operated on and maintained by the Decision Engine. This data may be transmitted to the Decision Engine entirely in-band, as private data transmitted shortly before or concurrently with the Addressable Content Set, or entirely out-of-band, or a combination of both.

The Decision Engine could operate in a stand-alone fashion within a Client-DPI Device, or it could be implemented to utilize information from the head-end or programming source.

This standard does not specify the operation of the Decision Engine, or the format or syntax of selection criteria data.

A.5.2 Trigger Signals

A Trigger Signal is employed to indicate the location of a Signaled Switch Point. The Trigger Signal may also carry identifying information that the Client-DPI Device can use to match each occurrence of a Trigger Signal with a specific Addressable Content Set, which would, for example, allow the Client-DPI Device to determine if it has missed a Trigger Signal, or if it has received an unexpected one.

One possible Trigger Signal would be a single transport packet using a PID set aside for the purpose. This packet could contain private data which indicates the system clock value corresponding to the location of the next Signaled Switch Point. Additional elements of the private data could convey other vendor-specific information.

Another possible implementation of the Trigger Signal would be to use a sequence of packets which employs the MPEG-2 adaptation field constructs splicing_point_flag and splice_countdown, in each of the video and audio streams where a Signaled Switch Point is located. Using this method, the Signaled Switch Point is located at the end of the packet which causes splice_countdown to be decremented to zero.

Multiple implementations of the Trigger Signal may be present concurrently. In such a case, any particular Client-DPI Device implementation would not be expected to process more than one implementation of the Trigger Signal.

This standard does not specify the format or syntax of the Trigger Signal.

Note: OCAP devices will use a different trigger than legacy devices.

A.5.3 Command and Control

Aspects of Decision Engine operation may be configurable through a signaling means used for the purpose. Typically, this will be through a full-time connection to the Client-DPI Device, as is provided for by an out-of-band communications channel.

This standard does not specify the format or syntax of command and control signaling.

Appendix B: Bound on A/V Presentation Time Mismatch Due to Splicing

The requirements in section 8.2 indicate that the presentation time mismatch introduced between audio and video of an addressable ad by splicing the content to the end of Filler is bounded. The bound on the presentation time mismatch is one-half of the audio frame duration.

Figure 10 shows that while both the Filler and addressable content are perfectly synchronized before splicing, time mismatch is introduced in the addressable content after splicing.

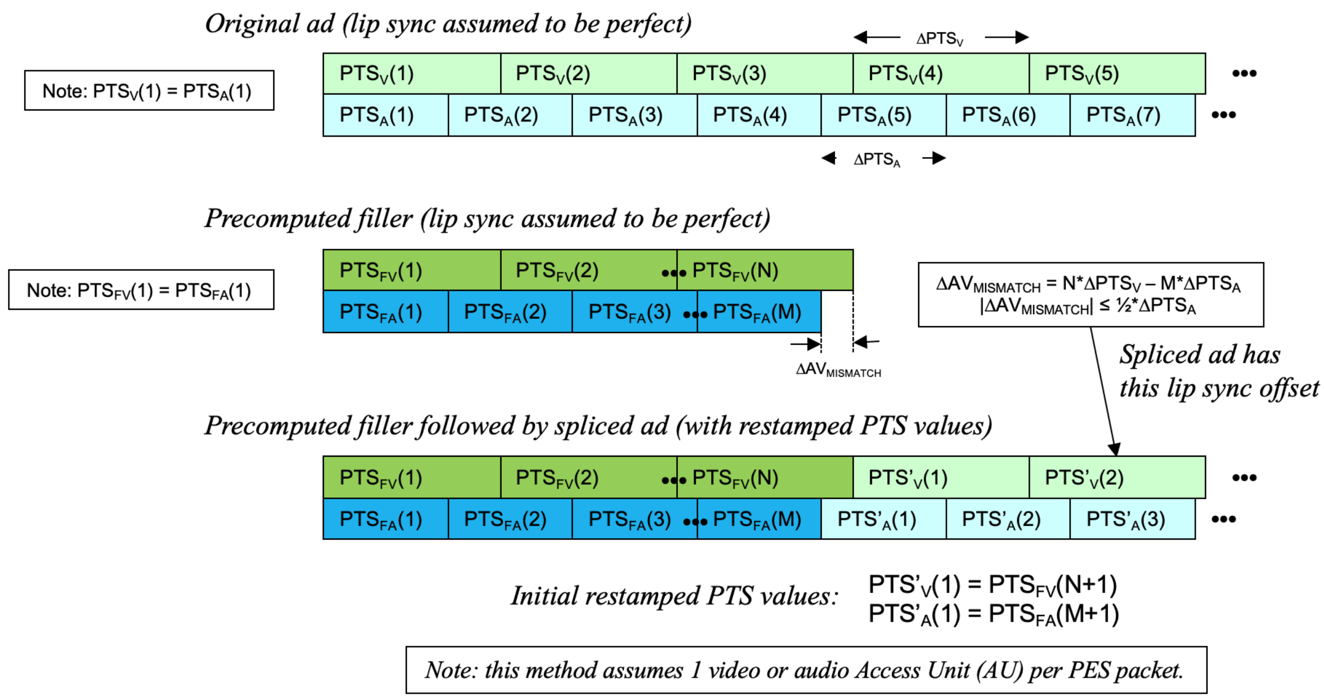


Figure 10 Illustration of A/V presentation time mismatch introduced by splicing a perfectly synchronized ad after a perfectly synchronized Filler.

Figure 11 shows different cases of presentation time mismatch introduced based on different audio and video frame durations. Although the frame durations are fixed within both audio and video components, the frame duration can be different between the audio and video components.

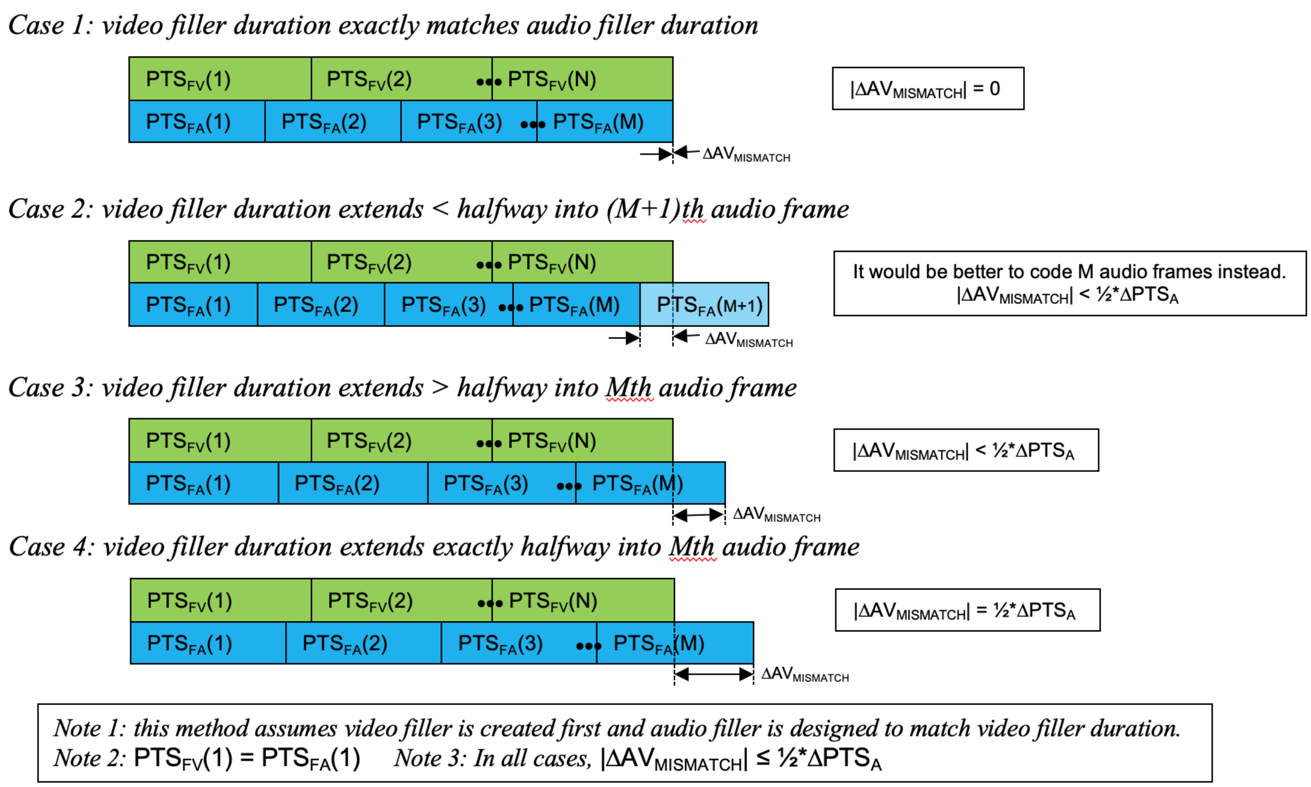


Figure 11 : A/V presentation time mismatch introduced under different audio and video frame durations.

Figure 12 illustrates the bound on the presentation time mismatch introduced when ad and Filler are created with AC-3 audio. Between MPEG-2 video and AVC video, the set of frame rates supported are the same ([SCTE 43], [SCTE 128-1]) and hence the set of video frame durations to be considered for both codecs is the same. The frame duration for AC-3 audio sampled at 48 KHz is fixed at 32 ms [A/52], [SCTE 54]. The bound on the absolute value of the presentation time mismatch is one-half the audio frame duration = 16 ms.

Video Frame Rate	ΔPTS_V	ΔPTS_A	MAX $\lvert \Delta AV_{MISMATCH} \rvert$
23.976 Hz	41.7083 ms	32 ms	16 ms
24 Hz	41.6667 ms	32 ms	16 ms

29.97 Hz	33.3667 ms	32 ms	16 ms
30 Hz	33.3333 ms	32 ms	16 ms
59.94 Hz	16.6833 ms	32 ms	16 ms
60 Hz	16.6667 ms	32 ms	16 ms

Figure 12 : Bound on presentation time mismatch introduced by splicing ad with Filler in the front when Filler and ad are created with either MPEG-2 or AVC video and AC-3 audio.

Figure 13 illustrates the bound when ad and Filler are both encoded with HEVC video codec and AC3 audio codec. Again, the bound on the absolute value of the presentation time difference is the same as the previous case (see Figure 12).

Video Frame Rate	**ΔPTS_V**	**ΔPTS_A**	**MAX $\lvert \Delta AV_{MISMATCH} \rvert$**
23.976 Hz	41.7083 ms	32 ms	16 ms
24 Hz	41.6667 ms	32 ms	16 ms
29.97 Hz	33.3667 ms	32 ms	16 ms
30 Hz	33.3333 ms	32 ms	16 ms
59.94 Hz	16.6833 ms	32 ms	16 ms
60 Hz	16.6667 ms	32 ms	16 ms
119.88 Hz	8.3417 ms	32 ms	16 ms
120 Hz	8.3333 ms	32 ms	16 ms

Figure 13 : Bound on presentation time mismatch introduced by splicing ad with Filler in the front when Filler and ad are created with HEVC video and AC-3 audio.

Figure 14 illustrates the audio/video mismatch region shown in green due to splicing (with Filler and ad encoded with AC-3 and a chosen video codec) overlayed on the detectability and acceptability threshold graph from [ITU-R BT.1359-1]. Observe that the mismatch falls within the undetectability plateau.

 © 2024 SCTE

When the range of presentation time mismatch introduced by splicing is overlayed (Figure 14) on the detectability and acceptability threshold graph [ITU-R BT.1359-1] as shown in green, one can see that with AC-3 codec and a chosen video codec for both Filler and ad content, the mismatch falls within the undetectability threshold (C-C'). In other words, the presentation time mismatch introduced by splicing should be undetectable to the viewer.

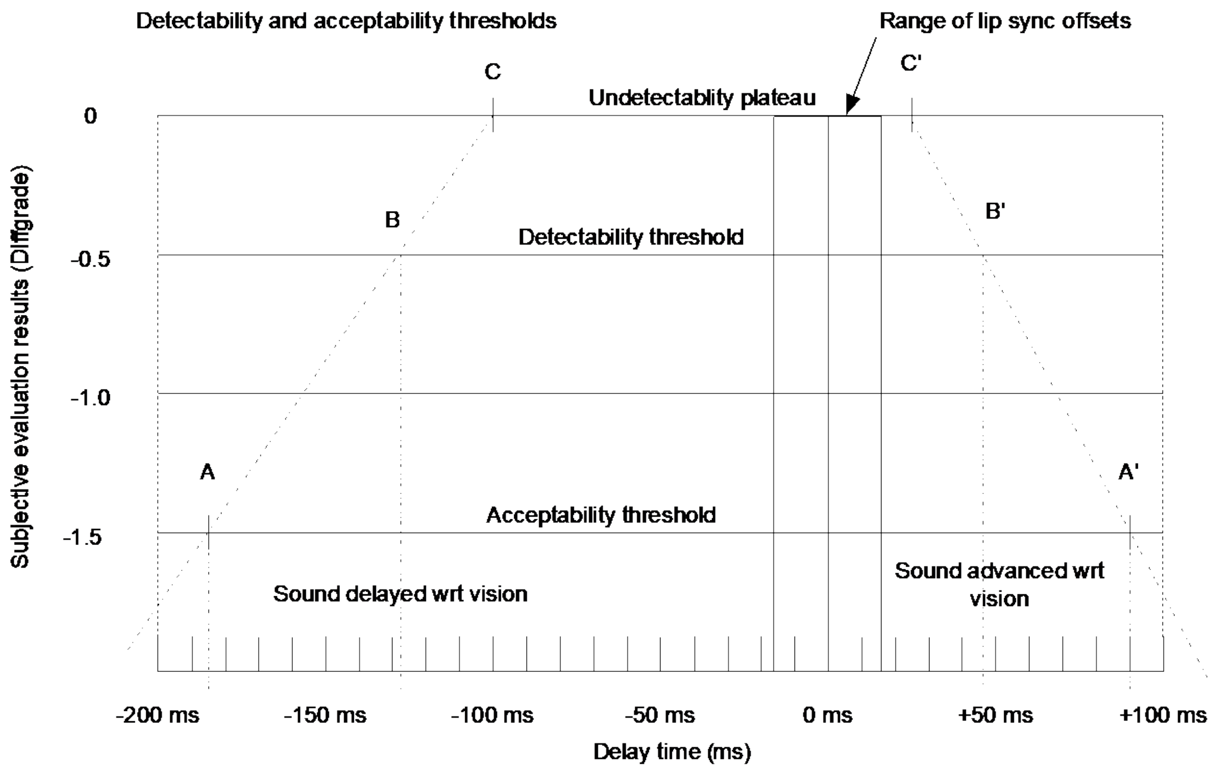


Figure 14: Lip Sync Detectability Limits.

Similarly, other audio codecs can be considered, for instance, if the ad content and Filler are encoded with AVC video and DTS-HD audio codec. For DTS-HD core audio (i.e. no extensions) encoded with 512 samples/frame and 48 KHz sampling rate, the audio frame duration is 10.67 ms. As described above, the maximum bound on the presentation time mismatch between audio and video after splicing is one half of (10.67 ms) = 5.33 ms. The key to this computation is the audio frame duration.

One of the assumptions made in the computation was the video in the Filler was created first and the audio was designed to match the video duration. We could relax that assumption and do the same computation assuming that the Filler was created with audio first and the video was added to match the audio duration. In that case, it can be shown that the presentation time mismatch ($|\Delta AV_{MISMATCH}|$) is:

$$|\Delta AV_{MISMATCH}| = ½*\Delta PTS_V$$

where ΔPTS_V is the fixed video frame duration Therefore, the preferred way to create ad content with Filler is to try both approaches, namely, create audio Filler first and match its duration with video, and vice versa and choose the result which creates least audio/video presentation time mismatch.

Another case of interest could be creating ad and Filler content with AC-4 audio and any video codec like HEVC. AC-4 supports the generation of coded audio frames that represent the same time interval as the associated video frame [A/342-2]. All fractional video frame rate and corresponding video frame durations are supported. The implication is that if the frame alignment option of AC-4 is used in creating spliced Filler + ad content, no presentation time stamp mismatch is introduced. This scenario corresponds to Case 1 in Figure 11.