Society of Cable Telecommunications Engineers

ENGINEERING COMMITTEE
Digital Video Subcommittee

AMERICAN NATIONAL STANDARD

ANSI/SCTE 138 2009

STREAM CONDITIONING FOR SWITCHING OF ADDRESSABLE CONTENT IN DIGITAL TELEVISION RECEIVERS

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) Standards are intended to serve the public interest by providing specifications, test methods and procedures that promote uniformity of product, interchangeability 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 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, whether used domestically or internationally.

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

Attention is called to the possibility that implementation of this standard may require the use of subject matter covered by patent rights. By publication of this standard, no position is taken with respect to the existence or validity of any patent rights in connection therewith. 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 standard 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 http://www.scte.org.

140 Philips Road
Exton, PA 19341

TABLE OF CONTENTS

TABLE OF FIGURES

1.0 SCOPE

This document describes the stream Conditioning required to enable Client-DPI Receivers to implement switching in a both non-seamless fashion ("Level 0", or "L0"), and in a seamless fashion ("Level 1", or "L1").

Note: Client-DPI devices include Receivers conforming to OCAP 1.1 (and later) as well as other cable-compatible Receivers.

2.0 NORMATIVE REFERENCES

The following documents contain provisions, which, through reference in this text, constitute provisions of this standard. At the time of subcommittee approval, the editions indicated were valid. All standards are subject to revision, and parties to agreement based on this standard are encouraged to investigate the possibility of applying the most recent editions of the documents listed below.

2.1 SCTE References

1. ANSI/SCTE 30 2009: Digital Program Insertion Splicing API.
2. ANSI/SCTE 43 2005: Digital Video Systems Characteristics Standard for Cable Television.
3. ANSI/SCTE 54 2009: Digital Video Service Multiplex and Transport System for Cable Television

2.2 Standards from other Organizations

4. ISO/IEC 13818-1 (2007): International Standard, Information Technology – Generic coding of moving pictures and associated audio information: Systems.
5. ISO/IEC 13818-2 (2000): International Standard, Information Technology – Generic coding of moving pictures and associated audio information: Video
6. ATSC A/52B (2005): Digital Audio Compression Standard (AC-3, E-AC-3)
7. ATSC A/53 Part 5 (2007): AC-3 Audio System Characteristics

3.0 INFORMATIVE REFERENCES

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

3.1 SCTE References

8. ANSI/SCTE 35 2007: Digital Program Insertion Cueing Message for Cable.

3.2 Standards from other Organizations

9. SMPTE 312M-2001: Television – Splice Points for MPEG-2 Transport Streams

10. OC-SP-OCAP1.1-I01-061229 – OpenCable Application Platform Specification OCAP1.1 profile

4.0 COMPLIANCE NOTATION

"SHALL"	This word or the adjective "REQUIRED" means that the item is an absolute requirement of this specification.
"SHALL NOT"	This phrase means that the item is an absolute prohibition of this specification.
"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.

5.0 DEFINITIONS AND ACRONYMS

Throughout this standard the terms below have specific meanings. Because some of the terms are defined in other documents having very specific technical meanings, the reader is referred to the original source for their definition. For terms used within this standard, brief definitions are given below.

Addressable Content Set: A collection of simultaneously transmitted elementary streams typically containing video and/or audio that may be selected as the result of a Switch.

Client-DPI Receiver: A Receiver 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 Receivers.

Decision Engine: The algorithm in a Client-DPI Receiver that selects which particular content should be selected from an Addressable Content Set and presented to the viewer.

DPI: Digital Program Insertion. [1] [8]

DTS: Decode Time Stamp. [4]

Filler: Used in Level 0 Streams, this is the region within which the Client-DPI Receiver begins and completes a Switch. The Filler is inserted into all 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 Receiver begins and completes a Switch. The Gap is inserted into all video and audio elementary streams that comprise an Addressable Content Set.

GOP: Group of Pictures. [5]

Insertion Channel: As defined in SCTE 30, section 4.

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. [1]

Level 1 Stream: A Primary Channel or Insertion Channel that has been conditioned in order to enable a Level 1 Switch. [1]

Level 0 Switch: This term refers to a Switch from one elementary stream to another elementary stream of the same stream_type without any perceptible audio or video artifacts, but which may include the use of black frames and audio silence to mask Switch latency.

Level 1 Switch: This term refers to a Switch from one elementary stream to another elementary stream of the same stream_type without any perceptible delay or audio or video artifacts.

MPTS: Multiprogram Transport Stream. [4]

PCR: Program Clock Reference. [4]

PES: Packetized Elementary Stream. [4]

PID: Packet Identifier. [4]

PMT: Program Map Table. [4]

Primary Channel: As defined in SCTE 30, section 4.

PSI: Program Specific Information. [4]

PTS: Presentation Time Stamp. [4]

Receiver: A cable television receiving device which is typically a set-top box or a DTV receiver.

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 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 Receiver, which chooses certain elementary streams out of an Addressable Content Set. Note: switch and splice are typically used interchangeably.

Switching Engine: The functionality that executes a Switch in a Client-DPI Receiver.

Trigger Signal: An element within the MPEG-2 transport stream multiplex which indicates the location of a Signaled Switch Point.

6.0 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 Receiver. The Receiver 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.

6.1 Client-DPI Receiver 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.

6.2 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 ISO/IEC 13818-1 [4]. 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 Receiver 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 [4], the transmission of video, audio, and data is timed to insure that data from each stream is available at the Receiver in time for it to be decoded without overflowing that stream's buffer in the Receiver. Null packets may be used in the stream as necessary to satisfy QAM modulator requirements. The result, as seen at the input to the Receiver, is a continuous stream of packets.

6.3 Level 0 Systems

6.3.1 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 Receiver, 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 1 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.

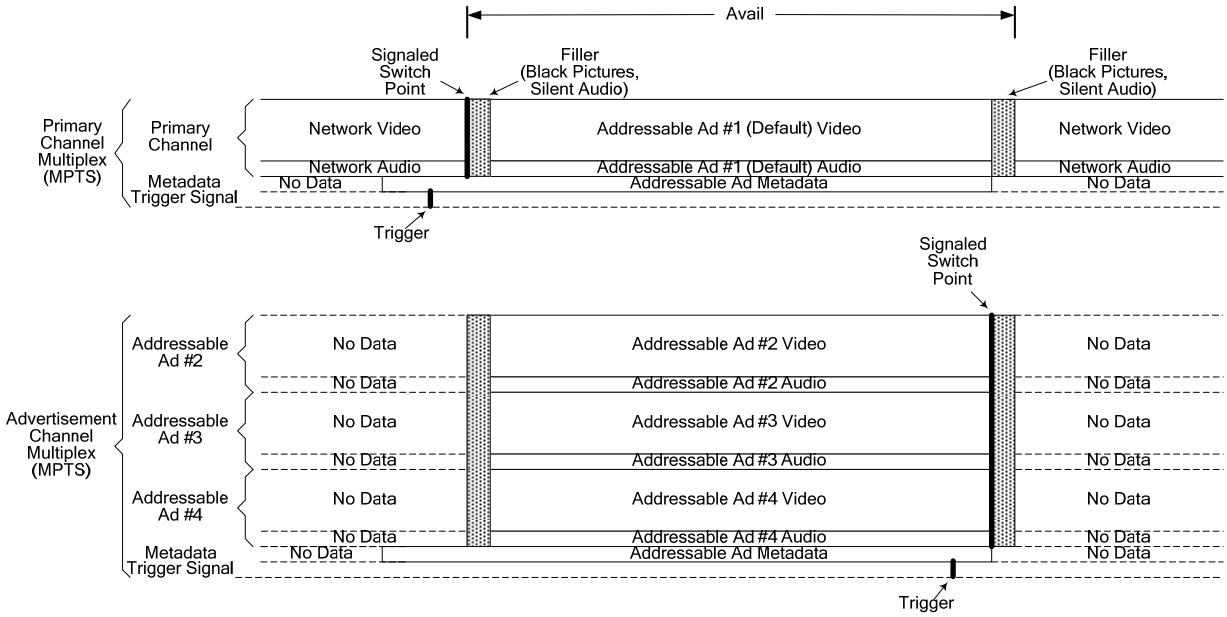

Figure 1 – Example Level 0 Transport Multiplex

In this example, one of the addressable advertisements, designated as the default, 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 Receivers that do not implement Level 0 switching. Also, the default addressable advertisement will be selected by Level 0 capable Receivers 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 Receiver 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 Receiver will be able to acquire and begin decoding the new channel without displaying any artifacts or losing any significant video or audio data. The duration of the Filler is selected to allow for the time required for the slowest Receiver to tune to and begin decoding the new multiplex.

The Signaled Switch Points and Fillers 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 the 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 avail 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 1 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.

6.3.2 Level 0 Head-End

Figure 2 and Figure 3 below show hypothetical head-end architectures suitable for inserting Level 0 Addressable Content Sets.

In Figure 2, 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.

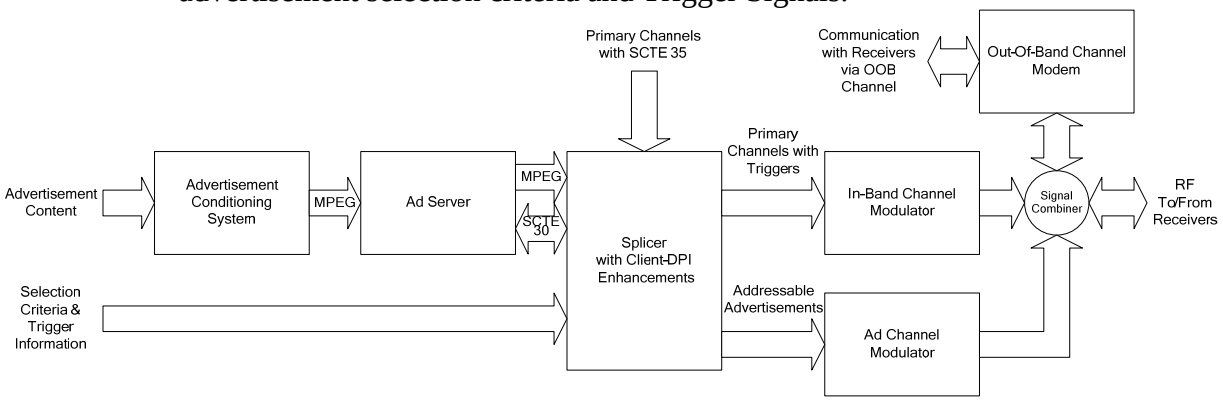

Figure 2 – Example Level 0 Head-End Configuration #1

In the example shown in Figure 3, 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.

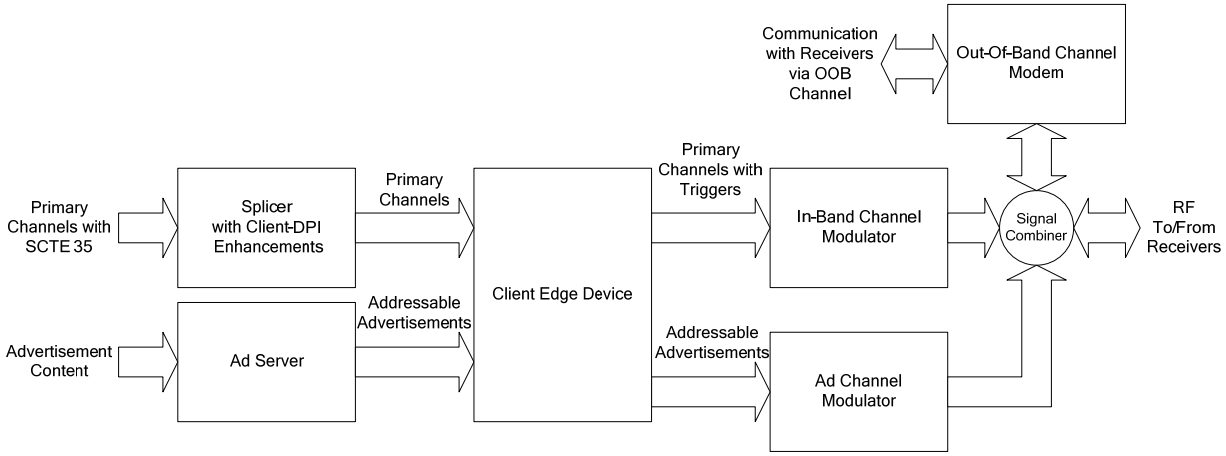

Figure 3 – Example Level 0 Head-End Configuration #2

6.3.3 Level 0 Receiver

Figure 4 below depicts a hypothetical Receiver 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.

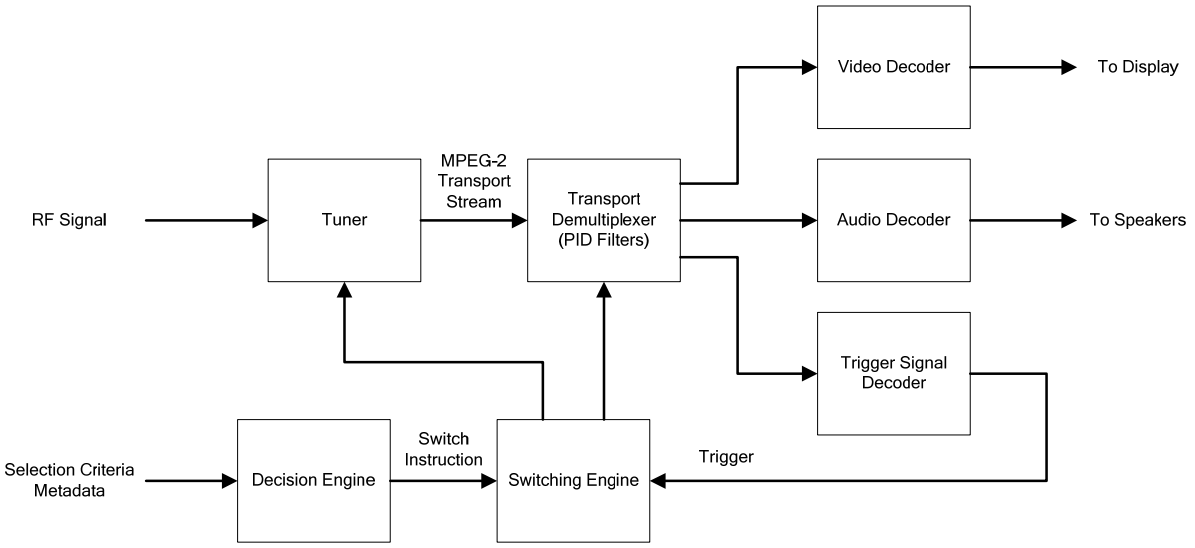

Figure 4 – Example Level 0 Set-Top Decoder Architecture

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 Receiver to process it, and the presence of the Filler following the Switch Point ensures that there is time for the Receiver 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 in a manner that eliminates the need to parse the PSI.

It may be that other software in the Receiver maintains a state associated with the currently tuned channel (for example, to display the channel number on the front panel). In such a case, to maintain a seamless experience for the viewer, it is necessary that the Switching Engine be able to control the tuner and the demultiplexer without altering this state.

6.4 Level 1 Systems

6.4.1 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 Receiver to switch to the selected advertisement seamlessly.

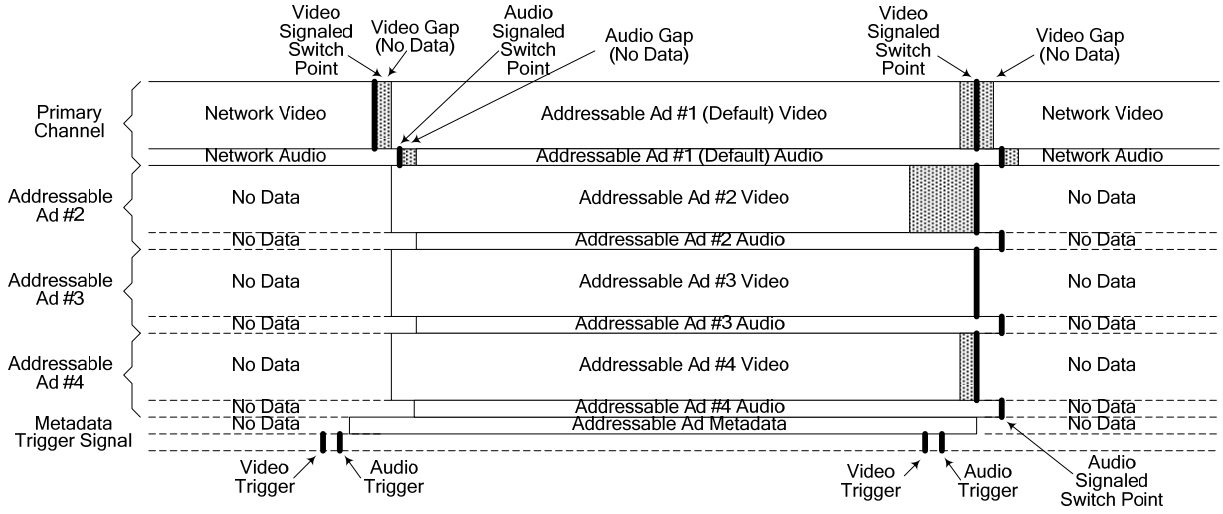

Figure 5 – Example Level 1 Transport Multiplex

Figure 5 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 Receivers that do not implement Level 1 switching. Also, the default addressable advertisement will be selected by Level 1 capable Receivers 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. 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 Receiver 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 Receiver. 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 Receiver 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. In Figure 5, 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.

6.4.2 Level 1 Head-End

Figure 6 shows an example of a head-end architecture suitable for inserting Level 1 Addressable Content Sets.

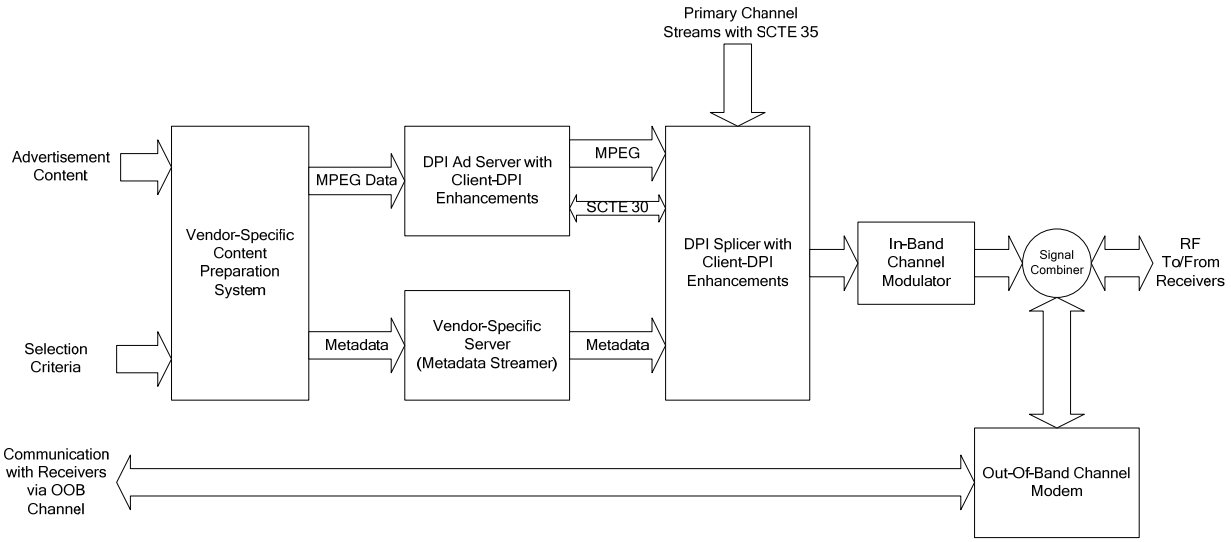

Figure 6 – Example Level 1 Head-End Configuration

In this Figure, 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 Receiver. Some of the features of this conditioning are:

- Conformance with MPEG-2, ATSC and SCTE standards with respect to transport, video, AC-3 audio, buffer management and system clock is maintained across the splice.
- Vertical resolution and frame rate 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 avail 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 Receiver 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 frame transmitted before a splice overlaps in presentation time with audio frames transmitted after that splice.

Additional conditioning for addressable advertisements within Level 1 Addressable Content Sets is required to insure that the Receiver 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.

6.4.3 Level 1 Receiver

Figure 7 depicts a hypothetical Client-DPI Receiver 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.

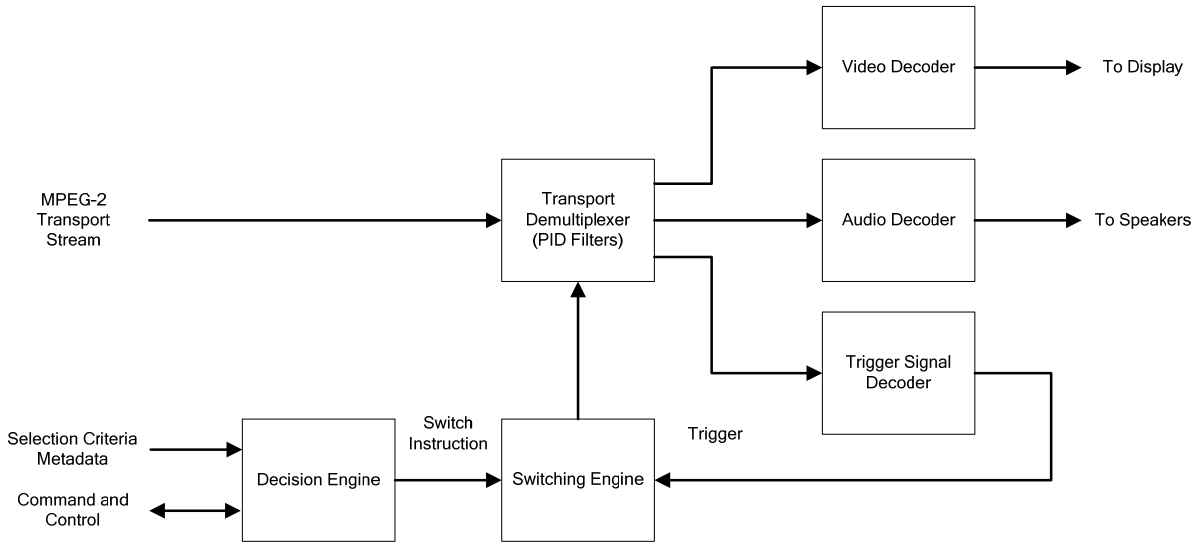

Figure 7 – Example Level 1 Set-Top Decoder Architecture

In this Figure, 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 Receiver to process it, and the presence of the Gap surrounding the Switch Point insures that there is time for the Receiver 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.

It may be that other software in the Receiver maintains a state associated with the currently tuned viewer channel (for example, to display the channel number on the front panel). In such a case, to maintain a seamless experience for the viewer, it is necessary that the Switching Engine be able to control the demultiplexer without altering this state.

6.5 Signaling

Three distinct types of signaling may be employed in an addressable advertising system. One type of signaling provides the Client-DPI Receiver with selection criteria associated with each Addressable Content Set, so that the Receiver 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.

6.5.1 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 Receiver, 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.

6.5.2 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 Receiver can use to match each occurrence of a Trigger Signal with a specific Addressable Content Set, which would, for example, allow the Receiver 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 Receiver 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.

6.5.3 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 Receiver, 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.

7.0 STANDARDS COMPLIANCE

All transport streams (with video and audio components) delivered to Client-DPI Receivers shall be compliant with SCTE 54. [3] In addition, video shall be compliant with SCTE 43 and audio shall be compliant with ATSC A/52B as constrained by A/53 Part 5 (2007). [2][6]][7]

8.0 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. [3]

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.

Note: 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.

Note: 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.

9.0 LEVEL 0 REQUIREMENTS

Level 0 Conditioning consists of a Filler, as defined in section 9.3, and MPEG coding requirements defined in section 0. The Conditioning described in this section shall be present in the transport stream at the input to a Client-DPI Receiver when a Level 0 Switch is enabled. In this context, “enabled” means that a Client-DPI Receiver 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 MPEG-2 video streams.

9.1 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.

9.2 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 9 of this standard, when a Level 0 Switch is enabled.

9.3 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.

In video streams, the 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. The 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.

The audio Filler shall be comprised of complete AC-3 frames. Each AC-3 frame shall be in its own PES packet. The transition between Filler and non-Filler content shall be compliant with AC-3 standards.

The transition between Filler and non-Filler content in both video and audio streams shall not introduce any decoding discontinuities. This includes T-STD, time base and continuity counter discontinuities.

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.

Note: 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.

Note: Client-DPI Receivers 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.

9.4 Filler Alignment

For a given Switch, all Fillers 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 either the video or audio Filler transmission. Across all services involved in a given Switch, all alignment points shall occur within 33 milliseconds.

Following the alignment point, all Fillers, including both video and audio, shall have a transmission duration of at least 533 milliseconds.

Note: The alignment point establishes the earliest point at which a Client-DPI Receiver can initiate a Switch. The start of Filler is aligned to guarantee that a Client-DPI Receiver which completes a Switch very quickly does not join the destination service before the transmission of the Filler on that service has started. The transmission duration of the Filler is chosen to guarantee that a Client-DPI Receiver which completes a Switch slowly does not join the destination service after the transmission of the Filler on that service has ended. The minimum Filler transmission duration is chosen to accommodate a Client-DPI Receiver that can complete a tune in 500 milliseconds or less. In certain implementations, the 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.

9.5 Filler Presentation Duration

Across all video streams involved in a given Switch, the presentation duration of all video Fillers (defined as the difference between the PTS of the last presented video frame in the Filler and the PTS of the first presented video frame in the Filler) shall be the same.

Across all audio streams involved in a given Switch, the presentation duration of all audio Fillers (defined as the difference between the PTS of the last audio frame in the Filler and the PTS of the first audio frame in the Filler) shall be the same.

The presentation duration of audio Filler shall not differ from the presentation duration of video Filler by more than one-half of the video frame presentation duration.

Note: The presentation duration of video Filler and audio Filler cannot be constrained to be identical due to the fact that the frame rate of video is not the same as the frame rate of audio.

9.6 Addressable Content Set Duration

Within an Addressable Content Set, the presentation duration of all video streams (defined as the difference between the PTS of the last presented video frame and the PTS of the first presented video frame) shall be the same.

Within an Addressable Content Set, the presentation duration of all audio streams (defined as the difference between the PTS of the last audio frame and the PTS of the first audio frame) shall be the same.

Within an Addressable Content Set, the presentation duration of audio shall not differ from the presentation duration of video by more than one-half of the video frame presentation duration.

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 9.5.

Note: The presentation duration of video and audio in an Addressable Content Set cannot be constrained to be identical due to the fact that the frame rate of video is not the same as the frame rate of audio.

9.7 MPEG 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.

10.0 LEVEL 1 REQUIREMENTS

Level 1 Conditioning consists of a Gap, as defined in section 10.3, and specific MPEG coding requirements defined in sections 10.4 and 10.5. This Conditioning shall be present in the transport stream at the input to a Client-DPI Receiver when a Level 1 Switch is enabled. In this context, "enabled" means that a Client-DPI Receiver 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 MPEG-2 video streams.

10.1 Use of Transport Streams

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

10.2 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.

10.3 The Gap

When a Level 1 Switch is to be enabled, a Gap shall be simultaneously inserted into all coded video elementary streams that may be switched away from or switched to.

A separate Gap shall be simultaneously inserted into all coded audio elementary streams that may be switched away from or switched to.

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.

The duration of the Gap shall be greater than or equal to 10 milliseconds in transmission time.

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.

10.4 MPEG Coding Requirements

10.4.1 The System Clock and Program Clock Reference

A switch from one stream to another (of the same type) during the Gap is expected to maintain the continuity of the system clock in a Client-DPI Receiver. All of the services comprising an Addressable Content Set shall reference a common PCR_PID.

10.4.2 MPEG Syntax

A switch from one stream to another (having the same stream_type) during the Gap in a Client-DPI Receiver 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.

10.4.3 Constraints on 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 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. Switching between progressive and non-progressive video content is not permitted.

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.

The first picture in presentation order following a Gap shall have a PTS such that the presentation of this picture follows the presentation of the last picture presented prior to the Gap at the proper time as determined by the frame duration.

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.

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 receiver.

10.4.4 Constraints on 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 frame 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 frame. The values of PTS of the first audio frame in all of the audio streams that may be selected as the result of a Switch shall be the same, and the value of that PTS shall be such that the presentation of this frame follows the presentation of the last frame transmitted prior to the Gap at the proper time as determined by the frame duration.

10.5 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.