MPEG–4 System
INTRODUCTION

ISO/IEC 14496-1
Final Committee Draft of International Standard

Sunghwan Kang
Contents

1. Preview
2. Introduction
3. Normative References
4. Additional References
5. Definitions
6. Abbreviations and Symbols
7. Conventions
8. System Decoder Model
MPEG-4 System

1. Preview

ISO/IEC 14496-1
Final Committee Draft of International Standard
What’s MPEG

• Moving Picture Experts Group (MPEG)
  – is a working group of ISO/IEC charged with the development of video and audio encoding standards
• Its first meeting was in May of 1988 in Ottawa, Canada
• MPEG has grown to include approximately 350 members per meeting
• MPEG’s official designation is ISO/IEC JCT1/SC29 WG11
Background of MPEG-4

- **MPEG-21**: Future standard as a multimedia framework
- **MPEG-7**: Describing multimedia content
- **MPEG-4**: Multimedia Communication based on ‘Object’
- **MPEG-3**: Standard for Digital Television
- **MPEG-2**: Standard for Digital Television
- **MPEG-1**: Storage of audio-visual information an compact disc

Figure 1. Concept of MPEGs
MPEG-4’s parts

- ISO/IEC 14496-1: Systems
- ISO/IEC 14496-2: Visual
- ISO/IEC 14496-3: Audio
- ISO/IEC 14496-4: Conformance Testing
- ISO/IEC 14496-5: Reference Software
- ISO/IEC 14496-6: DMIF
  - Delivery Multimedia Integration Framework
What’s new MPEG-4

- User Interaction
  - Object-based coding
- Synthetic/Natural Hybrid Coding (SNHC)
- Flexible Decoder Architecture
  - Tool-based
  - S/W-based implementation
- Error Resilience
- Flexible Composition/Rendering
- NOT RESTRICTED TO “very-low bit rate coding”
  - Coding of Audio/Visual Object
- DMIF
Scene Composition and Presentation of AV Object

audiovisual object

3-D world kept within the decoder

Rendered the image on a 2-D display

User interaction possible
MPEG-2 v.s. MPEG-4 System

- Video v.s. Video Object (VO)
- Picture v.s. Video Object Plane (VOP)

- Examples
  - Program Clock Reference (PCR) v.s. Object Clock Reference (OCR)
MPEG-4 Example of Scene (1/3)
MPEG-4 Example of Scene (2/3)

ObjectDescriptor
{
  OD_ID_1
  List of
  {
    Elementary-Stream-Descriptors
  }
}

ObjectDescriptor
{
  OD_ID_2
  List of
  {
    Elementary-Stream-Descriptors
  }
}
MPEG-4 Example of Scene (3/3)

ObjectDescriptor
{
    OD_ID_1
    List of
    {
        Elementary-Stream-Descriptors
    }
}

ObjectDescriptor
{
    OD_ID_2
    List of
    {
        Elementary.....
    }
}

ES_Descriptor
{
    ES_ID_1
    ......
}

ES_Descriptor
{
    ES_ID_2
    ......
}

ES_Descriptor
{
    ES_ID_3
    ......
}
2D Example

Animated Text + Video + Still Images

Video Overlay of Animated Text + Video + Still Images
3D Example

3D Graphic Primitives

Complex 3D Mesh

Scene with Face Object

Animated Face = Face + Location + Spatialized sound

2D Interface

Video Object

3D Object

Interactive Behavior

3D Example 3D Example 3D Example 3D Example

2D Interface

Animated Face = Face + Location + Spatialized sound

3D Graphic Primitives

Complex 3D Mesh

Scene with Face Object

2006-9-14 

Realtime image processing & Telecommunication Lab
2D/3D Example

2D or 3D scene as a texture map on 3D

2D inside a 3D plane
MPEG-4 System

2. Introduction

ISO/IEC 14496-1
Final Committee Draft of International Standard
This Standard describes a system for communicating **interactive audiovisual scenes**. Such scenes consist of: media objects, scene description, synchronization, identification, description, association.
This standard specifies the following tools:

- A terminal model for time and buffer management
  - SDM
- A coded representation of interactive audiovisual scene description information
  - Binary Format for Scenes – BIFS

Figure 2. logical structure of the scene
This standard specifies the following tools (cont’d)

- a coded representation of identification and description of audiovisual streams as well as the logical dependencies between stream information
  - **Object** and other **Descriptors**
- a coded representation of synchronization information
  - **Sync Layer** – **SL**
- a multiplexed representation of individual streams in a single stream
  - **FlexMux**
- a coded representation of descriptive audiovisual content information
  - **Object Content Information** – **OCI**
MPEG-4 Architecture

Figure 3. Processing stages in an audiovisual terminal
Terminal Model: System Decoder Model (SDM)

- SDM is to allow a sender to predict how the receiver will behave in terms of buffer management and synchronization.

- Timing Model
  - defines the mechanisms through which a receiver establishes a notion of time and performs time-dependent events.
  - conveys the time from sender to receiver the time
    - such as a desired decoding or composition time.
  - Clock references and time stamps.

- Buffer Model
  - enables the sender to monitor and control the buffer resources that are needed to decode each ES.
  - allows the sender to specify when information is removed from these buffers and schedule data transmission so that overflow does not occur.
Layer of MPEG-4 architecture (1/6)

• Multiplexing of Streams: TransMux Layer
  – This layer is not in the scope of 14496-1
  – Only the interface to this layer is defined
  – This layer is specified in 14496-6
  – These mechanisms serve for transmission as well as storage of streaming data

• Synchronization of Streams: Sync Layer
  – extracts timing information to enable synchronized decoding and, subsequently, composition of the ES data
  – SL-packetized streams provide timing and synchronization information as well as random access information
• Compression Layer
  – recovers data from its encoded format and performs the necessary operations to **reconstruct the original information**
  – Access to the various **Elementary Streams (ESs)** gained through object descriptors
  – An ES may contain one of the following
    • Object descriptors
    • Audio or visual object data for a single object
    • Scene description
    • Object content information
Compression Layer – Object Descriptor Streams

- An object descriptor is a collection of one or more ES descriptors that provide configuration and other information for the streams that relate to a media object or scene description.
- Conveyed in elementary streams.
- Unique (within the current session) identifying number (Object Descriptor ID).
- ES Descriptors include information about the encoding format, configuration information for the decoding process and the Sync Layer packetization, as well as quality of service requirements.

<Object Descriptor>

<Elementary Stream descriptors>

Object Descriptor ID

ES Descriptors

ES ID

ES ID

ES ID

Figure 4. Structure of Object Descriptor
Layer of MPEG-4 architecture (4/6)

- Compression Layer – Scene Description Streams
  - addresses the organization of audiovisual objects in a scene, in terms of both spatial and temporal positioning
  - The Scene description is represented using a parametric methodology (BIFS – Binary Format for Scenes)
  - Consists of an encoded hierarchy (tree) of nodes with attributes and other information (including event sources and targets)
  - Leaf nodes in this tree correspond to particular audio or visual objects (media nodes)
  - Intermediate nodes perform grouping, transformation, and other operations (scene description nodes)
  - Scene description can evolve over time by using scene description updates
Layer of MPEG-4 architecture (5/6)

- Compression Layer – Media Streams
  - The coded representations of audio and visual information are described in 14496-2 and 14496-3, respectively

- Compression Layer – Object Content Information Streams
  - carry descriptive information about audiovisual objects
  - The main content descriptors are:
    - content classification descriptors
    - keyword descriptors
    - rating descriptors
    - language descriptors
    - textual descriptors
    - descriptors about the creation of the content
Layer of MPEG-4 architecture (6/6)

- Compression Layer – Upchannel Streams
  - An ES flowing from receiver to sender is treated the same way as any downstream ES
  - The content of upchannel streams is specified in the same part of the specification of downstream data
  - e.g., upchannel control streams for video downchannel ES
MPEG-4 System

3. Normative References
4. Additional References
5. Definition
6. Abbreviations and Symbols
7. Conventions

ISO/IEC 14496-1
Final Committee Draft of International Standard
3. Normative References

- ITU-T Recommendations and International Standards contain provisions of this Final Committee Draft of International Standard
- At the time of publication, the editions indicated were valid
- All Recommendations and Standards are subject to revision
4. Additional References

### 5. Definitions (1/4)

<table>
<thead>
<tr>
<th>Terms</th>
<th>descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Unit (AU)</td>
<td>An individually accessible portion of the ES</td>
</tr>
<tr>
<td>Alpha Map</td>
<td>The transparency parameters associated to a texture map</td>
</tr>
<tr>
<td>Audiovisual Scene (AV Scene)</td>
<td>A set of media objects together with scene description information</td>
</tr>
<tr>
<td>Buffer Model</td>
<td>A model that defines how a terminal manages the buffer resources</td>
</tr>
<tr>
<td>Byte Aligned</td>
<td>A position in a coded bit stream</td>
</tr>
<tr>
<td>Clock Reference</td>
<td>A special timestamp that conveys a reading of a time base</td>
</tr>
<tr>
<td>Composition</td>
<td>The process of applying SD information (spatio-temporal)</td>
</tr>
<tr>
<td>Composition Memory (CM)</td>
<td>A random access memory that contains CU</td>
</tr>
<tr>
<td>Composition Time Stamp (CTS)</td>
<td>An indication of the nominal composition time of a CU</td>
</tr>
<tr>
<td>Composition Unit (CU)</td>
<td>An individually accessible portion of the output that a media object decoder</td>
</tr>
<tr>
<td>Compression Layer</td>
<td>The layer is between the coded representation of an ES and its decoded</td>
</tr>
<tr>
<td>Decoding buffer (DB)</td>
<td>A buffer at the input of a media object decoder that contains AU</td>
</tr>
<tr>
<td>Decoder configuration</td>
<td>The configuration of a media object decoder (processing ES data)</td>
</tr>
<tr>
<td>Decoding Time Stamp (DTS)</td>
<td>An indication of the nominal decoding time of an AU</td>
</tr>
<tr>
<td>Descriptor</td>
<td>A data structure that is used to describe ES or media object</td>
</tr>
</tbody>
</table>

Table 1. Definitions of the terms in MPEG-4 System
## 5. Definitions (2/4)

<table>
<thead>
<tr>
<th>Terms</th>
<th>descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary Stream (ES)</td>
<td>A consecutive flow of data from source to destination on compression layer</td>
</tr>
<tr>
<td>Elementary Stream Descriptor</td>
<td>A structure contained in OD that describes information about ES</td>
</tr>
<tr>
<td>Elementary Stream Interface (ESI)</td>
<td>An interface modeling the exchange of ES between composition layer and sync layer</td>
</tr>
<tr>
<td>FlexMux Channel (FMC)</td>
<td>A label to differentiate between data within one FlexMux Stream</td>
</tr>
<tr>
<td>FlexMux Packet</td>
<td>The smallest data entity managed by the FlexMux tool</td>
</tr>
<tr>
<td>FlexMux Stream</td>
<td>A sequence of FlexMux Packets with data</td>
</tr>
<tr>
<td>FlexMux tool</td>
<td>that allows the interleaving of data from multiple data streams</td>
</tr>
<tr>
<td>Graphics Combination Profile</td>
<td>is the required capabilities of a terminal for processing graphical media objects</td>
</tr>
<tr>
<td>Inter</td>
<td>uses previously coded parameters to construct a prediction</td>
</tr>
<tr>
<td>Intra</td>
<td>does not make reference to previously coded parameters to perform the encoding</td>
</tr>
<tr>
<td>Initial Object Descriptor</td>
<td>allows the receiving terminal to gain access to portions of content</td>
</tr>
<tr>
<td>Intellectual Property Identification (IPI)</td>
<td>A unique identification of ES corresponding to media object</td>
</tr>
<tr>
<td>Media Object</td>
<td>A representation of a natural or synthetic object that audiovisual</td>
</tr>
<tr>
<td>Media Object Decoder</td>
<td>translates to decoded representation of an ES</td>
</tr>
<tr>
<td>Native BIFS Node</td>
<td>A BIFS node which is 14496-1 as opposed non-native BIFS (14772-1)</td>
</tr>
</tbody>
</table>

Table 1. Definitions of the terms in MPEG-4 System (cont’d)
### 5. Definitions (3/4)

<table>
<thead>
<tr>
<th>Terms</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Clock Reference (OCR)</td>
<td>is used by a media object decoder to recover the encoder’s time base</td>
</tr>
<tr>
<td>Object Content Information (OCI)</td>
<td>Additional information about content conveyed through one or more ESs</td>
</tr>
<tr>
<td>Object Descriptor (OD)</td>
<td>associates ES by means of their ES descriptors and defines logical dependencies</td>
</tr>
<tr>
<td>Object Descriptor Message</td>
<td>identifies the action to be taken on a list of OD (e.g. update or remove)</td>
</tr>
<tr>
<td>Object Descriptor Stream</td>
<td>conveys ODs encapsulated in OD message</td>
</tr>
<tr>
<td>Object Time Base (OTB)</td>
<td>A time base valid for a given object, and hence for its media object decoder</td>
</tr>
<tr>
<td>Parametric Audio Decoder</td>
<td>A set of tools for representing and decoding audio signals (14496-3)</td>
</tr>
<tr>
<td>Quality of Service (QoS)</td>
<td>The performance that an ES requests from the delivery channel</td>
</tr>
<tr>
<td>Random Access</td>
<td>The process of beginning to read and decode at an arbitrary point</td>
</tr>
<tr>
<td>Reference Point</td>
<td>A location in the data or control flow of a system</td>
</tr>
<tr>
<td>Rendering</td>
<td>The action of transforming a SD</td>
</tr>
<tr>
<td>Rendering Area</td>
<td>The portion of the display device’s screen</td>
</tr>
<tr>
<td>Scene Description (SD)</td>
<td>Information that describes the spatio-temporal positioning of media object</td>
</tr>
<tr>
<td>Scene Description Profile</td>
<td>A profile that defines the permissible set of SD elements</td>
</tr>
<tr>
<td>Scene Description Stream</td>
<td>An ES that conveys BIFS SD information</td>
</tr>
</tbody>
</table>

Table 1. Definitions of the terms in MPEG-4 System (cont’d)
<table>
<thead>
<tr>
<th>Terms</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session</td>
<td>The communication of the coded representation of an audiovisual scene between two terminals</td>
</tr>
<tr>
<td>SL-Packetized Stream (SPS)</td>
<td>A sequence of SL-Packets that encapsulate one ES</td>
</tr>
<tr>
<td>Stream Multiplex Interface (SMI)</td>
<td>Exchange of SL-packetized stream data between sync layer and the TransMux layer</td>
</tr>
<tr>
<td>Structured Audio</td>
<td>A method of describing sound effects and music</td>
</tr>
<tr>
<td>Sync Layer (SL)</td>
<td>Adapt ES data for communication across the Stream Multiplex Interface</td>
</tr>
<tr>
<td>Sync Layer Configuration</td>
<td>A configuration of the sync layer syntax for a particular ES</td>
</tr>
<tr>
<td>Sync Layer Packet (SL-Packet)</td>
<td>The smallest data entity managed by the sync layer (header, payload)</td>
</tr>
<tr>
<td>Syntactic Description Language (SDL)</td>
<td>A language defined by this specification</td>
</tr>
<tr>
<td>Systems Decoder Model (SDM)</td>
<td>Provides an abstract view of the behavior of a terminal</td>
</tr>
<tr>
<td>System Time Base (STB)</td>
<td>The time base of the terminal</td>
</tr>
<tr>
<td>Terminal</td>
<td>Receives and presents the interactive audiovisual scene</td>
</tr>
<tr>
<td>Time Base</td>
<td>The notion of a clock</td>
</tr>
<tr>
<td>Timing Model</td>
<td>Specifies the semantic meaning of timing information</td>
</tr>
<tr>
<td>Time Stamp</td>
<td>An indication of a particular time instant relative to a time base</td>
</tr>
<tr>
<td>TransMux</td>
<td>A generic abstraction for delivery mechanisms able to store or transmit</td>
</tr>
<tr>
<td>Universal Resource Locator</td>
<td>A unique identification of the location of an ES or an OD</td>
</tr>
</tbody>
</table>

Table 1. Definitions of the terms in MPEG-4 System (cont’d)
### 6. Abbreviations and Symbols (1/3)

<table>
<thead>
<tr>
<th>Abbs.</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU</td>
<td>Access Unit</td>
</tr>
<tr>
<td>AV</td>
<td>audiovisual</td>
</tr>
<tr>
<td>BIFS</td>
<td>Binary Format for Scene</td>
</tr>
<tr>
<td>CM</td>
<td>Composition Memory</td>
</tr>
<tr>
<td>CTS</td>
<td>Composition Time Stamp</td>
</tr>
<tr>
<td>CU</td>
<td>Composition Unit</td>
</tr>
<tr>
<td>DAI</td>
<td>DMIF Application Interface (see Part 6 of this Final Committee Draft of International Standard)</td>
</tr>
<tr>
<td>DB</td>
<td>Decoding Buffer</td>
</tr>
<tr>
<td>DTS</td>
<td>Decoding Time Stamp</td>
</tr>
<tr>
<td>ES</td>
<td>Elementary Stream</td>
</tr>
<tr>
<td>ESI</td>
<td>Elementary Stream Interface</td>
</tr>
<tr>
<td>ESID</td>
<td>Elementary Stream Identifier</td>
</tr>
<tr>
<td>FAP</td>
<td>Facial Animation Parameters</td>
</tr>
<tr>
<td>FAPU</td>
<td>FAP Units</td>
</tr>
<tr>
<td>FDP</td>
<td>Facial Definition Parameters</td>
</tr>
</tbody>
</table>

Table 2. The symbols and abbreviations are used in this specification.
6. Abbreviations and Symbols (2/3)

<table>
<thead>
<tr>
<th>Abbs.</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIG</td>
<td>FAP Interpolation Graph</td>
</tr>
<tr>
<td>FIT</td>
<td>FAP Interpolation Table</td>
</tr>
<tr>
<td>FMC</td>
<td>FlexMux Channel</td>
</tr>
<tr>
<td>FMOD</td>
<td>The floating point modulo (remainder) operator which returns the remainder of x/y such that: fmod(x/y) = x – k*y, where k is an integer. sgn( fmod(x/y) ) = sgn(x) abs( fmod(x/y) ) &lt; abs(y)</td>
</tr>
<tr>
<td>IP</td>
<td>Intellectual Property</td>
</tr>
<tr>
<td>IPI</td>
<td>Intellectual Property Identification</td>
</tr>
<tr>
<td>NCT</td>
<td>Node Coding Tables</td>
</tr>
<tr>
<td>NDT</td>
<td>Node Data Type</td>
</tr>
<tr>
<td>OCI</td>
<td>Object Content Information</td>
</tr>
<tr>
<td>OCR</td>
<td>Object Clock Reference</td>
</tr>
<tr>
<td>OD</td>
<td>Object Descriptor</td>
</tr>
<tr>
<td>ODID</td>
<td>Object Descriptor Identifier</td>
</tr>
<tr>
<td>OTB</td>
<td>Object Time Base</td>
</tr>
</tbody>
</table>

Table 2. The symbols and abbreviations are used in this specification (cont’d)
### Abbreviations and Symbols (3/3)

<table>
<thead>
<tr>
<th>Abbs.</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLL</td>
<td>Phase locked loop</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>SAOL</td>
<td>Structure Audio Orchestra Language</td>
</tr>
<tr>
<td>SASL</td>
<td>Structured Audio Score Language</td>
</tr>
<tr>
<td>SDL</td>
<td>Syntactic Description Language</td>
</tr>
<tr>
<td>SDM</td>
<td>Systems Decoder Model</td>
</tr>
<tr>
<td>SL</td>
<td>Synchronization Layer</td>
</tr>
<tr>
<td>SL-Packet</td>
<td>Synchronization Layer Packet</td>
</tr>
<tr>
<td>SMI</td>
<td>Stream Multiplex Interface</td>
</tr>
<tr>
<td>SPS</td>
<td>SL-packetized Stream</td>
</tr>
<tr>
<td>STB</td>
<td>System Time Base</td>
</tr>
<tr>
<td>TTS</td>
<td>Text-To-Speech</td>
</tr>
<tr>
<td>URL</td>
<td>Universal Resource Locator</td>
</tr>
<tr>
<td>VOP</td>
<td>Video Object Plane</td>
</tr>
<tr>
<td>VRML</td>
<td>Virtual Reality Modelling Language</td>
</tr>
</tbody>
</table>

Table 2. The symbols and abbreviations are used in this specification (cont’d)
7. Conventions

- Syntax Description
  - Syntactic Description Language is used
    - This language allows the specification of the mapping of the various parameters in a binary format
MPEG–4 System

8. SDM (System Decoder Model)

ISO/IEC 14496-1
Final Committee Draft of International Standard
Introduction of SDM

- The purpose of the SDM is to provide an abstract view of the behavior of a terminal
- SDM used by the sender to predict how the receiving terminal will behave in terms of buffer management and synchronization when decoding data received in the form of ES
- The Systems Decoder Model specifies
  - The interface for accessing demultiplexed data streams (Stream Multiplex Interface)
  - Decoding buffers for compressed data for each ES
  - The behavior of media object decoders
  - Composition memory for decompressed data for each media object and the output behavior towards the compositor
Figure 5. Systems Decoder Model

AU: access unit
CU: composition unit
OTB: object time base
STB: system time base
DTS: decoding time stamp
CTS: composition time stamp
OCR: object clock reference
Concepts of the Systems Decoder Model (1/2)

- **Stream Multiplex Interface (SMI)**
  - provides access to streaming data and fills up decoding buffers with the data

- **SL-Packetized Stream (SPS)**
  - The packet contain ES data partitioned in access units as well as timing and access unit labeling

- **Access Units (AU)**
  - is the smallest data entity to which timing information can be attributed

- **Decoding Buffer (DB)**
  - is a receiver buffer that contains access units

- **Elementary Streams (ES)**
  - Streaming data received at the output
Concepts of the Systems Decoder Model (2/2)

- **Elementary Stream Interface (ESI)**
  - models the exchange of ES data and associated control information between the Compression Layer and the Sync Layer

- **Media Object Decoder**
  - extracts access units from the decoding buffer at precisely defined points in time places composition units in the composition memory

- **Composition Units (CU)**
  - An AU corresponds to an integer number of CUs

- **Composition Memory (CM)**
  - contains composition units

- **Compositor**
  - Takes CUs out of the CM and either composes and presents them or skips them
Timing Model Specification (1/3)

The timing model relies on clock references and time stamps to synchronize media objects.
- Clock references is used to convey the notion of time to a receiver
- Time stamps are used to indicate the precise time instant (event)
- These time events are attached to AUs and CUs

System Time Base (STB)
- defines the receiving terminal’s notion of time e.g., counter

Object Time Base (OTB)
- defines the notion of time for a given media object

Object Clock Reference (OCR)
- corresponds to the value of the OTB at the time the transmitting terminal generates the OCR time stamp
Timing Model Specification (2/3)

- Decoding Time Stamp (DTS)
  - Each AU has an associated nominal decoding time, the time at which it must be available in the decoding buffer for decoding

- Composition Time Stamp (CTS)
  - Each CU has an associated nominal composition time, the time at which it must be available in composition memory for composition

- Occurrence and Precision of Timing Information in ESs
  - DTS, CTS, OCR values are to be inserted in the bitstream as well as the prediction, jitter and drift are application and profile dependent

- Time Stamps for Dependent Elementary Streams
  - A media object may be represented in a scalable manner by multiple ES
  - AUs are identified by identical DTS or CTS values
• Example
  - The example below illustrates the arrival of two access units at the Systems Decoder

Figure 6. example of Systems Decoder
Buffer Model Specification (1/3)

- Elementary Decoder Model

Figure 7. Flow diagram for the Systems Decoder Model

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>Index to the different ESs</td>
</tr>
<tr>
<td>DB_r</td>
<td>Decoding buffer for ES r</td>
</tr>
<tr>
<td>CM_r</td>
<td>Composition memory for ES r</td>
</tr>
<tr>
<td>a/c</td>
<td>Index to AUs of one ES / Index to CUs of one ES</td>
</tr>
<tr>
<td>AU_r(a)</td>
<td>The a&lt;sup&gt;th&lt;/sup&gt; AU in ES r. AU_r(a) is indexed in decoding order</td>
</tr>
<tr>
<td>td_r(a)</td>
<td>The decoding time, measured in seconds, of the a&lt;sup&gt;th&lt;/sup&gt; AU in the ES ‘r’</td>
</tr>
<tr>
<td>CU_r(c)</td>
<td>The c&lt;sup&gt;th&lt;/sup&gt; CU in ES ‘r’. CU_r(c) is indexed in composition order. CU_r(c) results from decoding AU_r(a). There may be several CUs resulting from decoding one AU</td>
</tr>
<tr>
<td>tc_r(c)</td>
<td>The composition time, measured in seconds, of the c&lt;sup&gt;th&lt;/sup&gt; CU in the ES ‘r’</td>
</tr>
</tbody>
</table>
• Assumptions
  – Constant end-to-end delay
    • = Encoding + encoder buffering + multiplexing + communication or storage + demultiplexing + decoder buffering + decoding
  – Demultiplexeser
    • The end-to-end delay between multiplexer output and demultiplexer input is constant
  – Decoding Buffer
    • The needed decoding buffer size is known by the sender and conveyed to the receiver
  – Decoder
    • The decoding time is assumed to be zero
  – Compositor
    • The composition time is assumed to be zero
Managing Buffers: A Walkthrough

- The model is used in a “Push” scenario
- In application where non-real time content is to be transmitted
- The behavior of the various SDM elements is modeled as follows
  - The sender signals the required buffer resources to the receiver before starting the transmission
  - The decoding buffer is filled at the maximum bitrate for this ES if data is available
  - At DTS, an AU is instantaneously decoded and removed from the DB
  - At DTS, a known amount of CUs corresponding to the AU are put in the composition memory
  - The current CU is available to the compositor between its composition time and the composition time of the subsequent CU