INTERNATIONAL STANDARD

ISO/IEC 14496-20

Second edition
2008-12-01

Information technology — Coding of audio-visual objects —

Part 20: Lightweight Application Scene Representation (LASeR) and Simple Aggregation Format (SAF)

Technologies de l'information — Codage des objets audiovisuels — Partie 20: Représentation de scène d'application allégée (LASeR) et format d'agrégation simple (SAF)

Reference number
ISO/IEC 14496-20:2008(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

ISO/IEC 14496-20 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This second edition cancels and replaces the first edition (ISO/IEC 14496-20:2006) which has been technically revised. It also incorporates the Amendment ISO/IEC 14496-20:2006/Amd.1:2008 and the Technical Corrigenda ISO/IEC 14496-20:2006/Cor.1:2007 and ISO/IEC 14496-20:2006/Cor.2:2008.

ISO/IEC 14496 consists of the following parts, under the general title *Information technology — Coding of audio-visual objects*:

- *Part 1: Systems*
- *Part 2: Visual*
- *Part 3: Audio*
- *Part 4: Conformance testing*
- *Part 5: Reference software*
- *Part 6: Delivery Multimedia Integration Framework (DMIF)*
- *Part 7: Optimized reference software for coding of audio-visual objects* [Technical Report]
- *Part 8: Carriage of ISO/IEC 14496 contents over IP networks*
- *Part 9: Reference hardware description* [Technical Report]
- *Part 10: Advanced Video Coding*
- *Part 11: Scene description and application engine*
- *Part 12: ISO base media file format*
- *Part 13: Intellectual Property Management and Protection (IPMP) extensions*
- *Part 14: MP4 file format*
- *Part 15: Advanced Video Coding (AVC) file format*
- *Part 16: Animation Framework eXtension (AFX)*
- *Part 17: Streaming text format*
- *Part 18: Font compression and streaming*

— *Part 19: Synthesized texture stream*
— *Part 20: Lightweight Application Scene Representation (LASeR) and Simple Aggregation Format (SAF)*
— *Part 21: MPEG-J Graphics Framework eXtensions (GFX)*
— *Part 22: Open Font Format*
— *Part 23: Symbolic Music Representation*
— *Part 24: Audio and systems interaction* [Technical Report]
— *Part 25: 3D Graphics Compression model*

Introduction

ISO/IEC 14496-20 specifies syntax and semantics for the following:

— The Lightweight Application Scene Representation (LASeR), specified in Clause 6, which is a binary format for encoding 2D scenes and updates of scenes. The binary format and the scene representation (based on SVG Tiny) are both designed to be suitable for lightweight embedded devices such as mobile phones.

— A Simple Aggregation Format (SAF), specified in Clause 7, to efficiently and easily transport LASeR data together with audio and/or video content over various delivery channels. This multiplexing scheme is designed to be simple to implement and to allow efficient demultiplexing on low-end devices.

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

ISO and IEC take no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has assured ISO and IEC that he is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with ISO and IEC. Information may be obtained from the companies listed in Annex A.

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

Information technology — Coding of audio-visual objects —

Part 20: Lightweight Application Scene Representation (LASeR) and Simple Aggregation Format (SAF)

1 Scope

This part of ISO/IEC 14496 defines a scene description format (LASeR) and an aggregation format (SAF) respectively suitable for representing and delivering rich-media services to resource-constrained devices such as mobile phones.

LASeR aims at fulfilling all the requirements of rich-media services at the scene description level. LASeR supports

— an optimized set of objects inherited from SVG to describe rich-media scenes,

— a small set of key compatible extensions over SVG,

— the ability to encode and transmit a LASeR stream and then reconstruct SVG content,

— dynamic updating of the scene to achieve a reactive, smooth and continuous service,

— simple yet efficient compression to improve delivery and parsing times, as well as storage size, one of the design goals being to allow both for a direct implementation of the SDL as documented, as well as for a decoder compliant with ISO/IEC 23001-1 to decode the LASeR bitstream,

— an efficient interface with audio and visual streams with frame-accurate synchronization,

— use of any font format, including the OpenType industry standard, and

— easy conversion from other popular rich-media formats in order to leverage existing content and developer communities.

Technology selection criteria for LASeR included compression efficiency, but also code and memory footprint and performance. Other aims included: scalability, adaptability to the user context, extensibility of the format, ability to define small profiles, feasibility of a J2ME implementation, error resilience and safety of implementations.

SAF aims at fulfilling all the requirements of rich-media services at the interface between media/scene description and existing transport protocols:

— simple aggregation of any type of stream,

— signalling of MPEG and non-MPEG streams,

— optimized packet headers for bandwidth-limited networks,

— easy mapping to popular streaming formats,

— cache management capability, and

— extensibility.

SAF has been designed to complement LASeR for simple, interactive services, bringing

— efficient and dynamic packaging to cope with high latency networks,

— media interleaving, and

— synchronization support with a very low overhead.

This part of ISO/IEC 14496 defines the usage of SAF for LASeR content; however, LASeR can be used independently from SAF.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 9899, *Programming languages — C*

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

ISO/IEC 14496-1, *Information technology — Coding of audio-visual objects — Part 1: Systems*

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 14496-18, *Information technology — Coding of audio-visual objects — Part 18: Font compression and streaming*

ISO/IEC 14882:2003, *Programming languages — C++*

ISO/IEC 16262:2002, *Information technology — ECMAScript language specification*

IETF BCP 13, *RFC 4288 on Media Type Specifications and Registration Procedures*, http://www.ietf.org/rfc/rfc4288.txt

IETF RFC 3023, *XML Media Types, M. Murata, S. St. Laurent, D. Kohn, January 2001*, http://www.ietf.org/rfc/rfc3023.txt

IETF RFC 3986, *Uniform Resource Identifiers (URI): Generic Syntax, T. Berners-Lee, R. Fielding, L. Masinter, January 2005*, http://www.ietf.org/rfc/rfc3986.txt

IETF RFC 2045, *MIME formats and encodings*, http://www.ietf.org/rfc/rfc2045.txt

IETF RFC 2326, *Real Time Streaming Protocol*, http://www.ietf.org/rfc/rfc2326.txt

IETF RFC 2965, *HTTP State Management Mechanism, Kristol and Montulli*, http://www.ietf.org/rfc/rfc2965.txt

W3C SVG11, Scalable Vector Graphics (SVG) 1.1 Specification [Recommendation], http://www.w3.org/TR/2003/REC-SVG11-20030114

W3C SMIL2, *Synchronized Multimedia Integration Language (SMIL 2.0) [Second Edition],* J. Ayars, D. Bulterman et. al., 7 January 2005, http://www.w3.org/TR/2005/REC-SMIL2-20050107

W3C CSS, *Cascading Style Sheets, level 2 [Recommendation]*, http://www.w3.org/TR/1998/REC-CSS2-19980512

W3C DOM, *Document* Object *Model Level 2 Events Specification, Version 1.0, W3C Recommendation*, 13 November, 2000, http://www.w3.org/TR/2000/REC-DOM-Level-2-Events-20001113

W3C XML *Events, an Events Syntax for XML, W3C Recommendation,* 14 October 2003, http://www.w3.org/TR/2003/REC-xml-events-20031014

W3C xml:id *Version 1.0, W3C Recommendation,* 9 September 2005, http://www.w3.org/TR/2005/PR-xml-id-20050712

W3C XLink, *XML Linking Language, W3C Recommendation,* 27 June 2001, http://www.w3.org/TR/2001/REC-xlink-20010627

3 Terms, definitions and abbreviations

3.1 Terms and definitions

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

3.1.1
access unit
individually accessible portion of data within a media stream

NOTE An access unit is the smallest data entity to which timing information can be attributed.

3.1.2
media time line
axis on which times are expressed within the transport or system carrying a LASeR or other stream

3.1.3
normal play time
indicates the stream absolute position relative to the beginning of the presentation

[IETF RFC 2326]

3.1.4
packet
smallest data entity managed by SAF consisting of a header and a payload

3.1.5
scene segment
set of access units of a LASeR stream, where only the first access unit contains a LASeRHeader

3.1.6
scene time line
axis on which times are expressed within the SVG/LASeR scene, e.g. begin and end

3.1.7
waiting tree
separate tree defined in addition to the scene tree

NOTE The compositor and renderer have no knowledge of the waiting tree, thus objects in the waiting tree are neither composited nor rendered.

3.2 Abbreviations

CSS Cascading Style Sheets (a W3C standard)

SMIL Synchronized Multimedia Integration Language (a W3C standard)

SVG Scalable Vector Graphics (a W3C standard)

4 Document conventions

This part of ISO/IEC 14496 uses the following styling conventions for various types of information.

Any name of element, attribute, descriptor or command defined in this specification is styled in bold italic, such as ***Add***. Any name of element, attribute, descriptor or command defined in another specification is prefixed with the name of that specification, such as ***SVG animate*** or ***SMIL video***.

XML examples use the following style:

```
<?xml version="1.0" encoding="UTF-8"?>
<svg width="480" height="360" viewBox="0 0 480 360"
     version="1.1" baseProfile="tiny">
        <defs>    …
```

SDL descriptions of binary syntax use the following style:

```
Insert extends LASeRUpdate {
  const bit(UpdateBits) InsertCode;
  uint(idBits) ref;
```


The following is the style used for ECMA Script:

```
function Insert(parentId, field, value) {…
```

5 Architecture

5.1 Overview

LASeR is defined in terms of abstract access units, which may be adapted for transmission over a variety of protocols. LASeR streams may be packaged with some or all of their related media into files of the ISO base media file format family (e.g. MP4) and delivered over reliable protocols. There is also a simple aggregation format (SAF), which aggregates a LASeR stream with some or all of its associated media into stream order. SAF may be delivered over reliable or unreliable protocols. Finally, LASeR streams could be adapted to other delivery protocols such as RTP [RFC 2326] or MPEG-2 transport [ISO/IEC 13818-1]; however, the definition of these mappings is outside the scope of this specification.

Figure 1 presents the LASeR and SAF architecture.

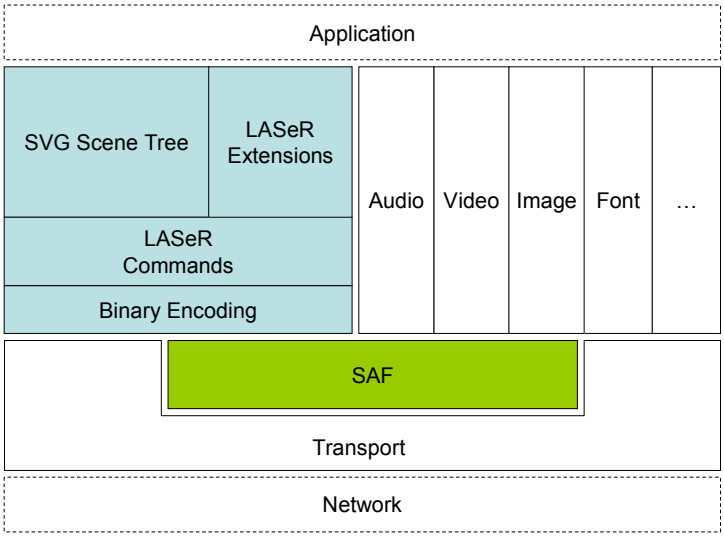

Figure 1 — Architecture of LASeR and SAF

5.2 LASeR systems decoder model

5.2.1 Introduction

The purpose of the LASeR systems decoder model is to provide an abstract view of the behaviour of the terminal complying with ISO/IEC 14496-20. It may be 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 elementary streams. The LASeR systems decoder model includes a timing model and a buffer model. The LASeR systems decoder model specifies

1. the conceptual interface for accessing data streams (Delivery Layer),

2. decoding buffers for coded data for each elementary stream,

3. the behavior of elementary stream decoders,

4. composition memory for decoded data from each decoder, and

5. the output behavior of composition memory towards the compositor.

Each elementary stream is attached to one single decoding buffer.

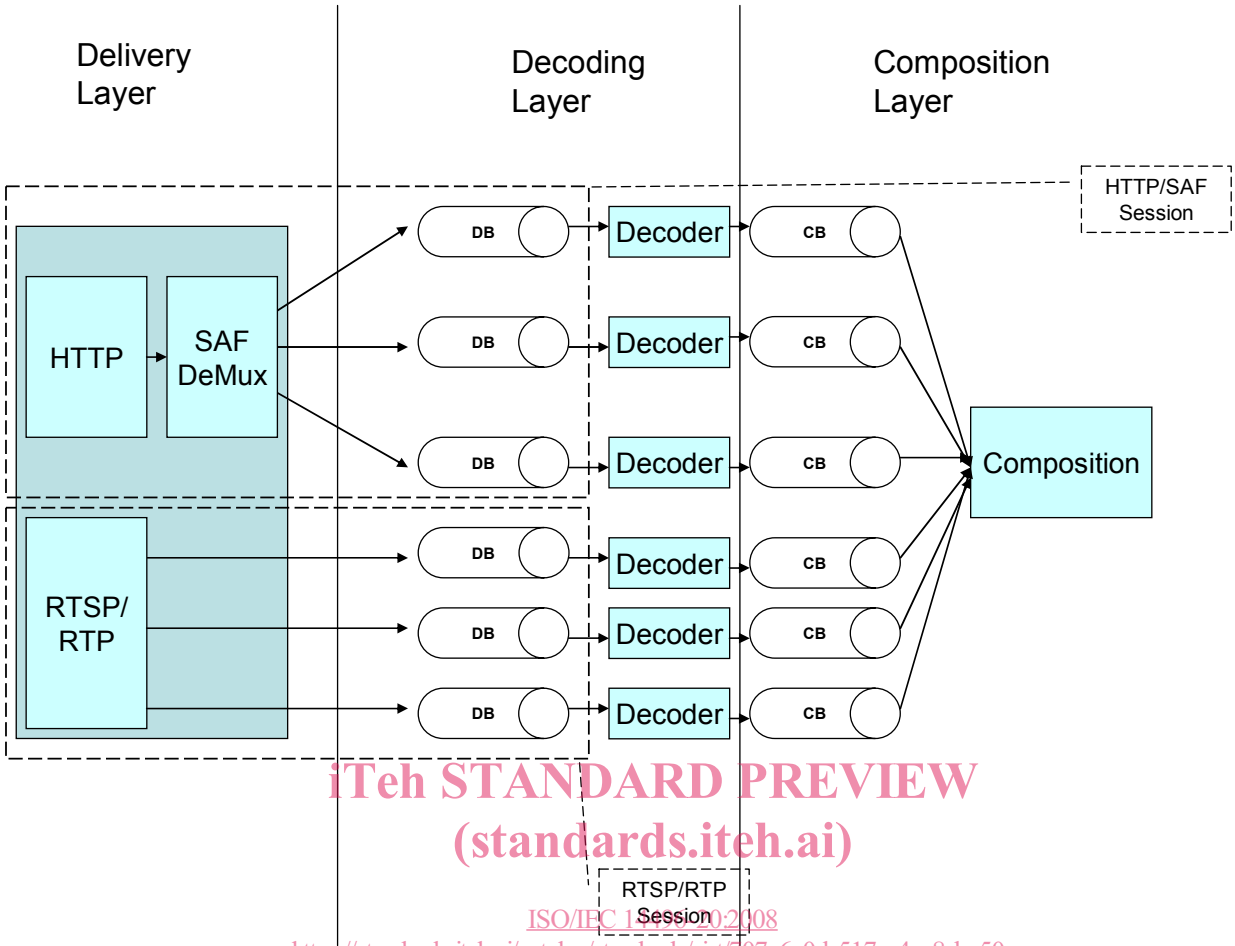

Figure 2 — LASeR systems decoder model

The definition in ISO/IEC 14496-1 of Access Unit, Decoding Buffer(DB), elementary stream (ES), Decoder (CU) and Composition Unit apply.

5.2.2 Decoder model

The decoder model as specified in 7.4.1 of ISO/IEC 14496-1:2004 applies.

5.2.2.1 Decoding buffer

The needed decoding buffer size is known by the sending terminal and conveyed to the receiving terminal as specified in 7.6. The size of the decoding buffer is measured in bytes. The decoding buffer is filled at the rate given by the maximum bit rate for this elementary stream while data is available and with a zero rate otherwise. The maximum bit rate is conveyed by the sending terminal as a part of the decoder configuration information during the set up phase for each elementary stream (see 7.6).

5.2.2.2 Decoder model with grouped streams

This decoder model may be enhanced when used for group of multiple elementary streams.

In such case, only one composition buffer for the group of streams is used for composition.

When such streams are grouped, and when the setup of multiple decoding chains are available, it is possible, although not mandatory, not to decode all streams at a time.

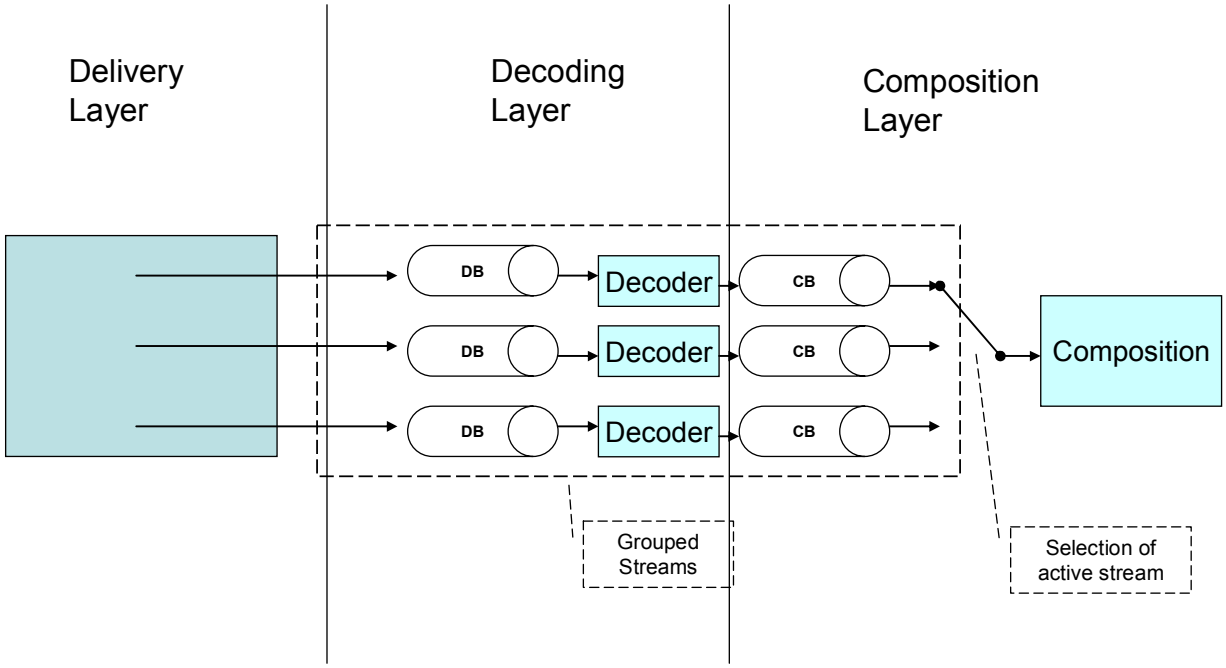

Figure 3 — Stream grouping with specified System Decoder Model (multiple decoders)

It is indeed expected that multiple decoders may not be available in lightweight terminals or that some delivery scenarios do not allow for having all streams available at the same time (e.g. in broadcast scenarios, the delivery layer could only tune in to one of the streams). The usage of new information about this grouping enables a smart usage of buffers and decoders.

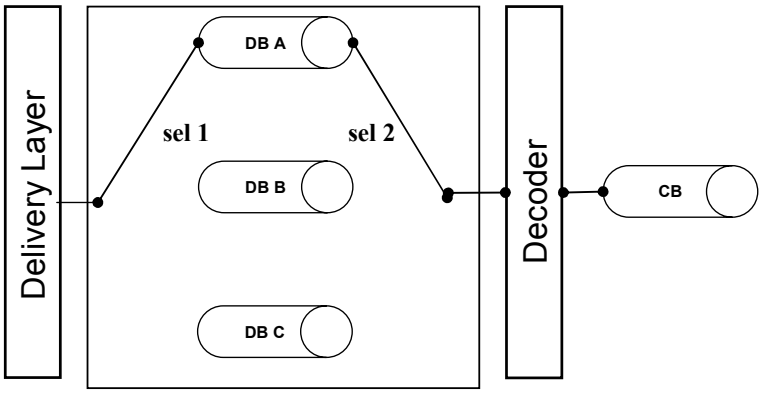

Figure 4 — Broadcast example of streams grouping, showing a potential optimization using a single decoder.

When only a subset of the group of streams can be accessed at a time (e.g. broadcast scenario depicted above), the selection of the active stream corresponds to a request for the corresponding streams. Nevertheless, the buffer model for stream grouping does not assume immediate reception of data after such request and therefore the active decoding buffer may continue to be used by the decoder up to the moment at which data is available for the newly available stream. In this case the decoding buffer associated to the newly