INTERNATIONAL STANDARD

ISO/IEC 14496-8

First edition
2004-05-15

Information technology — Coding of audio-visual objects —

Part 8: Carriage of ISO/IEC 14496 contents over IP networks

Technologies de l'information — Codage des objets audiovisuels —

Partie 8: Transport du contenu MPEG-4 sur les réseaux IP

Reference number
ISO/IEC 14496-8:2004(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

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.

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

ISO/IEC 14496-8 was prepared by Joint Technical Committee ISO/TC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

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*
— *Part 8: Carriage of ISO/IEC 14496 contents over IP networks*
— *Part 9: Reference hardware description*
— *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)*

Introduction

ISO/IEC 14496 is an International Standard designed for the representation and delivery of multimedia information over a variety of transport protocols. It includes interactive scene management, visual and audio representations as well as systems functionality like multiplexing, synchronization, and an object descriptor framework. This document provides a framework for the carriage of ISO/IEC 14496 contents over IP networks and guidelines for designing payload format specifications for the detailed mapping of ISO/IEC 14496 content into several IP-based protocols

Information technology — Coding of audio-visual objects —

Part 8:
Carriage of ISO/IEC 14496 contents over IP networks

1 Scope

This part of ISO/IEC 14496 specifies transport level functionalities for the communication of interactive audio-visual scenes. More specifically:

1. Framework for the carriage of ISO/IEC 14496 contents over IP networks;
2. Guidelines to design RTP payload formats for ISO/IEC 14496 contents including fragmentation and concatenation rules;
3. Usage rules of SDP to transport ISO/IEC 14496-1 related information;
4. MIME type definitions for ISO/IEC14496 contents; and
5. Analysis on RTP Security and Multicasting.

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.

IETF RFC 1889, *RTP A Transport Protocol for Real-Time Applications*

IETF RFC 1890, *RTP Profile for Audio and Video Conference with Minimal Control*

IETF RFC 2326, *Real Time Streaming Protocol (RTSP)*

IETF RFC 2327, *SDP: Session description protocol*

IETF RFC 3016, *RTP payload format for MPEG-4 audio/visual streams*

IETF RFC 3640, *Transport of MPEG-4 elementary streams*

3 Terms and definitions

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

3.1
MIME
Multipurpose Internet Mail Extensions, referring to an official Internet standard that identifies the format of the contents exchanged over different systems connected to the network

3.2
RTP
Real Time Transport Protocol, an Internet protocol used for transport of multimedia data

3.3
RTSP
Real Time Streaming Protocol, an application-level Internet protocol for control over the delivery of data with real-time properties

3.4
RTCP
Real Time Transport Control Protocol, an Internet protocol used to configure RTP packets and to control the transport of RTP packets

3.5
SDP
an Internet protocol used for describing multimedia sessions for the purposes of session announcement, session invitation, and other forms of multimedia session initiation

4 Symbols and abbreviated terms

AAC	Advanced Audio Coding
AU	Access Unit
BIFS	Binary Format for Scene
CELP	Code Excited Linear Prediction
CTS	Composition Time Stamp
ES	Elementary Stream
ESI	Elementary Stream Interface
ESID	Elementary Stream Identifier
FCR	FlexMux Clock Reference
IETF	Internet Engineering Task Force
MIME	Multipurpose Internet Mail Extensions
OCR	Object Clock Reference
OD	Object Descriptor
ODID	Object Descriptor Identifier
QoS	Quality of Service
RFC	Request For Comments
SL	Synchronization Layer
SL-Packet	Synchronization Layer Packet
URL	Universal Resource Locator

5 Use of RTP

5.1 Introduction

There are a number of RTP packetization schemes for ISO/IEC 14496 data. Media-aware packetization (e.g. video frames split at recoverable sub-frame boundaries) is a principle in RTP, and thus it is likely that several RTP schemes will be needed, to suit both the different kinds of media — audio, video, etc. — and different encodings (e.g. AAC and CELP audio codecs).This specification does not specify any payload format but do specify a general framework to design and utilize the payload formats in appropriate way.

This specification requires that, no matter what packetization scheme is used, there are a number of common characteristics that all shall have: however, such characteristics depend on the fact that the RTP Session contains a single elementary stream or a flexmux stream.

5.2 Carrying a single elementary stream

In case an RTP Session contains a single elementary stream the following characteristics apply:

- The RTP timestamp corresponds to the presentation time (e.g. CTS) of the earliest AU within the packet.
- RTP packets have sequence numbers in transmission order. The payloads logically or physically have SL Sequence numbers, which are in decoding order, for each elementary stream.
- The ISO/IEC 14496 timescale (clock ticks per second), which is timeStampResolution in the case of ISO/IEC 14496 Systems, shall be used as the RTP timescale, e.g. as declared in SDP for an RTP stream.
- To achieve a base level of interoperability, and to ensure that any ISO/IEC 14496 stream may be carried, all receivers should implement a generic payload format defined in “draft-ietf-avt-mpeg4-multisl-04.txt” as default RTP payload mapping scheme. Any new payload format should be a configurable subset of the generic payload format.
- Streams should be synchronized using RTP techniques (notable RTCP sender reports). When the ISO/IEC 14496 OCR is used, it is logically mapped to the NTP time axis used in RTCP.
- The RTP packetization schemes may be used for ISO/IEC 14496 elementary streams 'standing alone' (e.g. without ISO/IEC 14496 systems, including BIFS); or they may be used within an overall presentation using the object descriptor framework. In the latter case, a SLConfigDescriptor is sent describing the stream. Logically, each RTP stream is passed through a mapping function which is specific to the payload format used; this mapping function yields a SL packetized stream. The SLConfigDescriptor describes this logical stream, not the actual bits in the RTP payload. For example, the RTP sequence number may be used to make the SLPacketHeader sequence number; other SL fields may be set in this way, dynamically, or from static values in the payload specification. For example, as all RTP packets carry a composition time-stamp, the flag in the SL header indicating its presence can normally be statically defined as 'true'. Each payload format for ISO/IEC 14496 content shall specify the mapping function for the formation of the SLConfigDescriptor and the SLPacketHeader. In the case of the RFC 3016, the mapping shall be defined in a seperate document.

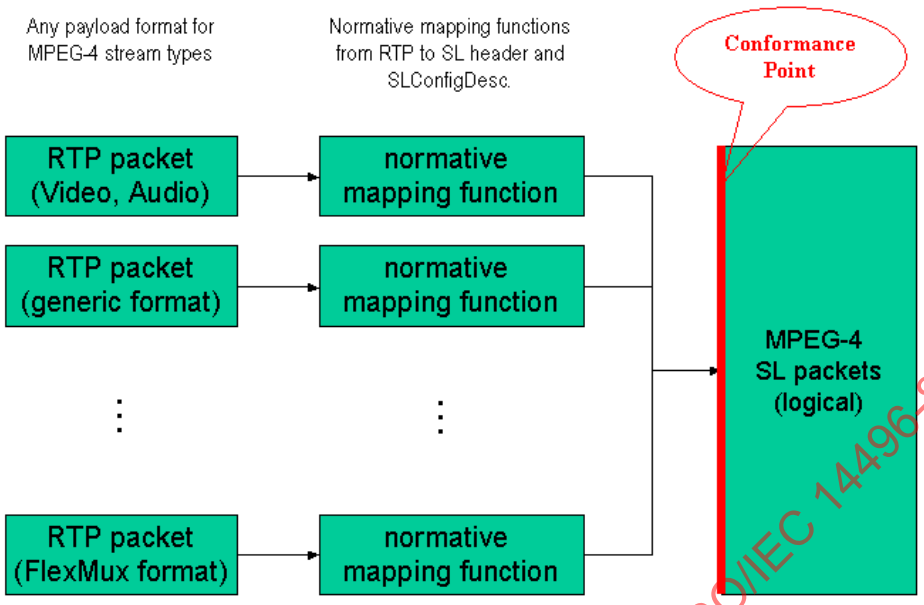

Figure 1 — RTP packet to SL packet mapping

5.3 Payload format selection

Other payload formats may be used. They are signalled as dynamic payload IDs, defined by a suitable name (e.g. a payload name in an SDP RTPMAP attribute). In particular, the development of specialized RTP payloads for video (e.g. respecting video packets) and audio (e.g. providing interleave) is expected. It is possible that these schemes can be compatible with the default scheme required here.

There may be a choice of RTP payload formats for a given stream (e.g. as an elementary stream, an SL-packetized stream, using FlexMux, and so on). It is recommended that

- terminals implementing a given subsystem (e.g. video) accept at least an ES and the default SL packings of that stream; for example, this means accepting the RFC 3016 and also the generic payload format for MPEG-4 Visual;

- terminals implementing a given payload format accept any stream over that format for which they have a decoder, even if that packing is not normally the 'best' packing.

5.4 ISO/IEC 14496 QoS considerations

For those streams requiring a certain Quality of Service (specifiable appropriately), the recommendation is to further investigate possible solutions such as the leverage of existing work in the IETF in this area (including, but not limited to FEC, re-transmission, or repetition). However, techniques in data-dependent error correction, or combined source/channel coding solutions make other schemes attractive. Also, it is recommended that requirement such as efficient grouping mechanisms (i.e. the ability to send in a single RTP packet multiple consecutive AUs, each with its own SL information) and low overhead is also taken into account.

6 Usage of SDP attributes

This specification considers only ISO/IEC 14496 Systems related issues. Usage of SDP information for specific payload format shall be specified in each RTP payload format RFCs. The usage of elementary streams in other contexts is not addressed here: codepoints for this case are specified in, and in other places.

This specification currently assumes that any session described by SDP (e.g. in SAP, as a file download, as a DESCRIBE over RTSP) has at most one ISO/IEC 14496 session. It is desirable that this restriction be lifted.

- Senders should alert receivers that an ISO/IEC 14496 session is included, by means of an SDP attribute that is general (i.e. before any "media" lines). This takes the form of the following attribute line:

  a=mpeg4-iod [<location>]

  location: In an RTSP session, this is an optional attribute. If not supplied, the IOD is retrieved over the RTSP session by using DESCRIBE with an accept of type application/mpeg4-iod or application/mpeg4-iod-xmt. Where the SDP information is supplied by some other means (e.g. as a file, in SAP), the location is obligatory. The location should be a URL enclosed in double-quotes, which will supply the IOD (e.g. small ones may be encoded using "data:", otherwise "http:" or other suitable file-access URL). When application/mpeg4-iod-xmt type is used, the IOD in XMT format shall be supplied. The InitialObjectDescriptor is defined in subclause 8.6.3.1 of ISO/IEC 14496-1:2001 and its XMT format is defined in subclause 15.8.3. of ISO/IEC 14496-1:2001 AMD2. Any terminals using IOD shall understand binary IOD and may understand textual IOD.

- New encoding names for the a = rtpmap attribute

  a = rtpmap:<payload> <name>/<time scale>/<parameters>

  payload is the dynamic payload number. The <name> is defined and documented in the IETF specification for the payload format.

- It is recommended that, no matter what payload format is used, each media stream be placed in a media section that is appropriate. For example, a payload format which can carry both video and audio streams may be used in sections of SDP starting both with "m=video" and "m=audio". The MIME name for the payload format is thus registered under all applicable branches.

- In case of a single elementary stream, the following attribute is defined:

  a=mpeg4-esid : a

  a is the ESID.

- Other SDP attributes should, if used, carry values consistent with those carried in ISO/IEC 14496 systems (for example, bit rate).

7 MIME Types

7.1 Top level MIME Types

- "video" shall be used for MPEG-4 Visual streams (i.e. video as defined in ISO/IEC 14496-2 (Streamtype = 4) and/or graphics as defined in ISO/IEC 14496-1 (Streamtype = 3)) or MPEG-4 Systems streams that convey information needed for an audio/visual presentation.

- "audio" shall be used for MPEG-4 Audio streams (ISO/IEC 14496-3) (Streamtype = 5)) or MPEG-4 Systems streams that convey information needed for an audio only presentation.

- "application" shall be used for MPEG-4 Systems streams (ISO/IEC14496-1 (all other StreamType values)) that serve other purposes than audio/visual presentation, e.g. in some cases when MPEG-J streams are transmitted.

7.2 MIME Types for elementary streams

- When a visual ISO/IEC 14496 ES is served (e.g. over HTTP or otherwise) and shall be identified by a MIME type, the type "video/MPEG4-visual" shall be used. This MIME type may require optional parameters to carry all necessary information to configure a receiver: therefore no further meta-information (such as that defined by the MP4 file format or by the ISO/IEC 14496 Object Descriptor framework) has to be provided in the data, and the data itself merely represents the media content. The format of the bit-stream, including timing etc., is defined in ISO/IEC 14496-2.

- The payload names used in an RTPMAP attribute within SDP, to specify the mapping of payload number to its definition, also come from the MIME namespace. Each of the RTP payload mappings defined above has a distinct name. It is recommended that visual streams be identified under "video", and audio streams be identified under "audio", and otherwise "application" be used.

- In some cases, the initial object descriptor needs to be identified with a MIME type. In this case, the type “applications/mpeg4-iod” shall be supported, and the type “application/mpeg4-iod-xmt” may be supported. In the latter case, the IOD will be described in an XMT textual format. The InitialObjectDescriptor is defined in subclause 8.6.3.1 of ISO/IEC 14496-1, and its XMT format is defined in subclause 15.8.3. of ISO/IEC 14496-1:2001 Amendment 2.

7.3 MIME Types for MP4 file

When an MP4 file is served (e.g. over HTTP) or otherwise shall be identified by a MIME type, the type "video/mp4" should be used. The types "audio/mp4" may be used when the ISO/IEC 14496 presentation contained within the MP4 file has no visual presentation and refers to a pure audio presentation.

7.4 MIME Type registration information

- MIME media type name:video, and audio
- MIME subtype name:mp4
- MIME media type name:application
- MIME subtype name:mpeg4-iod
- Required parameters:none
- Optional parameters:none
- Encoding considerations:base64 generally preferred; files are binary and should be transmitted without CR/LF conversion, 7-bit stripping etc.
- Security considerations:See below
- Interoperability considerations:A number of interoperating implementations exist within the ISO/IEC 14496 community; and that community has reference software for reading and writing the file format.
- Published specification:ISO/IEC 14496-1:2001
- Applications:Multimedia
- Additional information:

- Magic number(s):none
- File extension(s):mp4 and mpg4 are both declared at <http://pitch.nist.gov/nics/>
- Macintosh File Type Code(s):mpg4 is registered with Apple
- Person to contact for info:David Singer, singer@apple.com
- Intended usage:Common
- Author/Change controller: David Singer, ISO/IEC 14496 file format chair.

8 RTSP usage

This specification considers only ISO/IEC 14496 Systems related issues. The usage of elementary audio or visual streams in other context does not require any specific statement about RTSP. RTSP may be used as a session control protocol for sessions which carry ISO/IEC 14496 information. When RTSP is used as a session-control protocol:

- RTP should be used as the transport protocol.
- The initial DESCRIBE format should be SDP. If the SDP information reveals that an IOD is needed, and the terminal does not already have it, then a second DESCRIBE accepting an IOD should be performed (see above).
- Note that if all ISO/IEC 14496 streams are closed (TEARDOWN) then the RTSP session ID will be lost. The next (re-)opened stream will supply a new session ID. Care should be taken that the target of the URL has not changed in the interval; new DESCRIBEs may be needed.

9 Use of URLs in ES_Descriptors

When it is necessary to reference an RTP stream directly from an ES_Descriptor, the URL field of the descriptor can be used. For example, the URL could contain the SDP description of the stream using the “data:application/sdp” scheme.

When it is necessary to embed stream data directly inside an ES_Descriptor, the URL field of the descriptor can be used. For example, the URL could contain the data using the correct MIME type. In this case, the data consists of one SL packet that contains one access unit.

10 Security Considerations

RTP packets using the payload formats referred to in this specification are subject to the security considerations discussed in the RTP specification. This implies that confidentiality of the media streams is achieved by encryption. Because the data compression used with this payload format is applied end-to-end, encryption may be performed on the compressed data so there is no conflict between the two operations. The packet processing complexity of this payload type does not exhibit any significant non-uniformity in the receiver side to cause a denial-of-service threat. However, it is possible to inject non-compliant MPEG streams (Audio, Video, and Systems) to overload the receiver/decoder's buffers which might compromise the functionality of the receiver or even crash it. This is especially true for end-to-end systems like MPEG where the buffer models are precisely defined. ISO/IEC 14496 Systems supports stream types including commands that are executed on the terminal like OD commands, BIFS commands, etc. and programmatic content like MPEG-J (Java(TM) Byte Code) and ECMASCRIPT. It is possible to use one or more of the above in a manner non-compliant to MPEG to crash or temporarily make the receiver unavailable. Authentication mechanisms can be used to validate of the sender and the data to prevent security problems due to non-compliant malignant ISO/IEC 14496 streams. A security model is defined in ISO/IEC 14496 Systems streams carrying MPEG-J access units which comprises Java(TM) classes and objects. MPEG-J defines a set of Java APIs