Audio-Video Transport WG D. Singer, Y Lim
Internet Draft Apple Computer, net&tv
Document: draft-singer-mpeg4-ip-04 February 2001
Category: Expires July 2002
MPEG reference: N4427

A Framework for the delivery of MPEG-4 over IP-based Protocols

Status of This Memo

Distribution of this document is unlimited.

Abstract

This document forms an umbrella specification for the carriage and operation of MPEG-4 multimedia sessions over IP-based protocols, including RTP, RTSP, and HTTP, among others. It addresses IP Multicast as well.

It also serves to document the standard MIME types associated with MPEG-4 files.

1 Introduction

ISO/IEC 14496 is a 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/IEC14496 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

Glossary of terms and acronyms

AAC - MPEG-4 advanced audio codec

AU - access unit in an ES (the smallest media data unit to which
timing can be attributed).
BIFS - binary format for scenes; the MPEG-4 scene composition system
CELP - MPEG-4 speech codec
CTS - composition time stamp
DTS - decoding time stamp
ES - elementary stream
ESID - elementary stream ID
FCR - flexmux clock reference
FlexMux - a multiplex of several PDUs into a single unit; not used
for multiplexing in RTP
IOD - initial object descriptor; the 'hook' to the MPEG-4 streams
needed to start a session
OCR - object clock reference; an external clock reference for an
MEG-4 stream
OD - object descriptor; declares and defines an MPEG-4 stream
SL - synchronization layer
SL Packet - synchronization layer protocol data unit, in MPEG-4
systems

2 Use of RTP

There are a number of RTP packetization schemes for ISO/IEC 14496
data[5] [6] [9]. 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) [8].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 MUST

have: however, such characteristics depend on the fact that the RTP Session contains a single elementary stream or a flexmux stream.

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

2.1] The RTP timestamp corresponds to the presentation time (e.g. CTS) of the earliest AU within the packet.

2.2] 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.

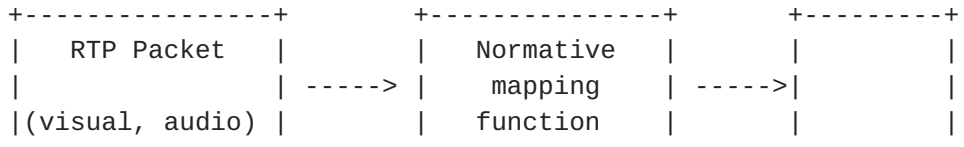
2.3] The ISO/IEC 14496 timescale (clock ticks per second), which is timeStampResolution in the case of ISO/IEC 14496 Systems, MUST be used as the RTP timescale, e.g. as declared in SDP for an RTP stream.

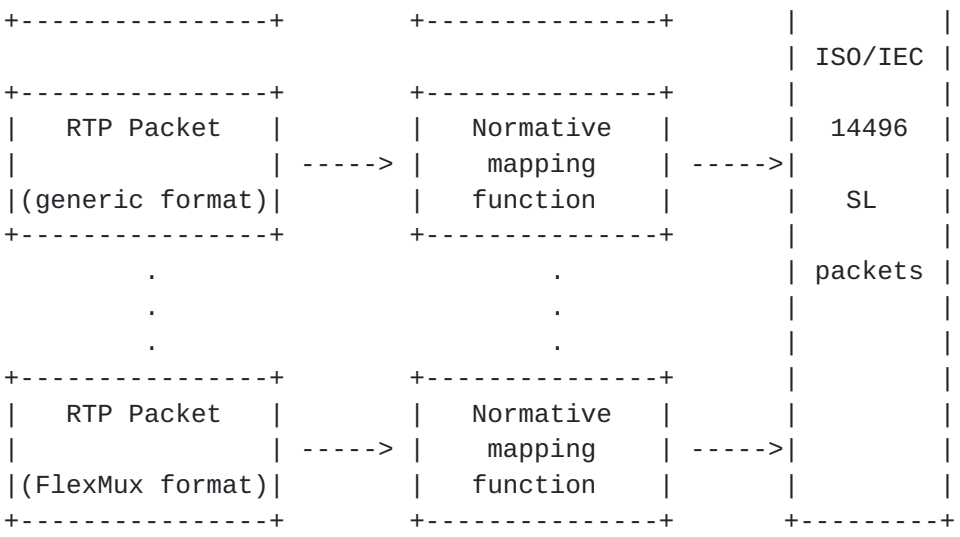
2.4] To achieve a base level of interoperability, and to ensure that any ISO/IEC 14496 stream may be carried, all senders and receivers MUST implement a default RTP payload mapping scheme. It is highly desirable that this default scheme is common for both pure Audio and Visual streams as well as for SL Packetized streams. This default scheme is not yet identified.

2.5] 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.

2.6] 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, an 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 an 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 MUST specify the mapping function for the formation of the SLConfigDescriptor and the SLPacketHeader.

In the case of RFC 3016, the mapping will be defined in a new section.





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

2.7] There is a single payload format for the carriage of Flexmux Streams over RTP [5]. Senders and receivers MAY implement this scheme.

2.8] The RTP timestamp corresponds to the FCR if present at the Flexmux level.

2.9] The ISO/IEC 14496 Flexmux timescale (FCR resolution in ticks per second) SHOULD be used as the RTP timescale (as can be declared in SDP).

2.10] the ISO/IEC 14496 FCR is logically mapped to the NTP time axis used in RTCP.

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 sub-system (e.g. video) accept at least an ES and the default SL packings of that stream; for example, this means accepting the draft by RFC 3016. and also the generic payload format for ISO/IEC 14496 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.

Future versions of this specification will identify the single standard RTP packing format for each ISO/IEC 14496 stream type. However, at the time of writing the RTP payload format specifications are still being defined, and the set is incomplete. These recommendations will form the basis for improved interoperability.

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 are also taken into account.

3 SDP Information

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 [6], 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.

3.1] 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 an 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. 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). The InitialObjectDescriptor is defined in sub-clause 8.6.3.1 of ISO/IEC 14496-1.

or:

a=mpeg4-iod-xmt:[<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-xmt. Where the SDP information is supplied by some other means (e.g. as a file, in SAP), the location is obligatory. The location shall be a URL enclosed in double-quotes, which will supply the IOD in XMT format (e.g. small ones may be encoded using "data:", otherwise "http:" or other suitable file-access URL). The InitialObjectDescriptor is defined in sub-clause 8.6.3.1 of ISO/IEC 14496-1, and its XMT format is defined in ISO/IEC 14496-1 2001 PDAM 2.

Any receivers using IOD shall understand binary IOD and may understand textual IOD.

3.2] New encoding names for the a = rtpmap attribute 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.

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 forma
time scale is the time scale of the RTP time stamps

parameters if used, is defined in the RTP payload format

3.3] The mapping of RTP streams to elementary streams needs to cover the Flexmux case as well as the single stream. Within the SDP information, a stream-specific attribute SHOULD be present for each ISO/IEC 14496 stream. It takes one of two forms, depending on whether a single elementary stream, or a flexmux, is carried.

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

a=mpeg4-esid:a

a is the ESID.

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

4 MIME Types

4.1] The historical approach for MPEG data is to declare it under "video", and this approach is followed for ISO/IEC 14496. For presentations with audio information and no visual aspect, the "audio" top-level mime type may be used; otherwise, "video" is used.

4.2] Amendment 1 of the ISO/IEC 14496 standard (also known as version 2) includes a standard file type for encapsulating ISO/IEC 14496 data. This file type can be used in a number of ways: perhaps the most

important are its use as an interchange format for ISO/IEC 14496 data, its use as a content-download format, and as the format read by streaming media servers.

These first two uses will be greatly facilitated if there is a standard MIME type for serving these files (e.g. over HTTP).

The ISO/IEC 14496 standard is broad, and therefore the type of data that may be in such a file can vary. In brief, simple compressed video and audio (using a number of different compression algorithms) can be included; interactive scene information; meta-data about the presentation; references to ISO/IEC 14496 media streams outside the file and so on.

The MIME types to be assigned to MP4 files SHOULD be "audio/mp4", and "video/mp4" , based on the criteria in 4.1. In either case, these indicate files conforming to the "MP4" specification (ISO/IEC 14496-1:2000, systems file format).

4.3] When an MP4 file is served (e.g. over HTTP) or otherwise must 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.

4.4] When a visual ISO/IEC 14496 ES is served (e.g. over HTTP or otherwise) and must 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.

4.5] 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 sub-clause 8.6.3.1 of ISO/IEC 14496-1, and its XMT format is defined in ISO/IEC 14496-1:2001 PDAM 2.

4.6] When a flexmux stream is served (e.g. over HTTP) or otherwise must be identified by a MIME type, the type "application/mpeg4-flexmux" SHALL be used. These files consist of concatenated flexmux PDUs in transmission order.

4.7] In some cases, the information needed by a flexmux decoder needs to be identified with a MIME type. In this case, the type "application/mpeg4-flexmuxinfo" SHOULD be used.

4.8] 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.

MIME media type name:video, and audio

MIME subtype name: mp4

MIME media type name:application
MIME subtype name: mpeg4-iod, mpeg4-flexmux, mpeg4-flexmuxinfo
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: Pending (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

5 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:

5.1] RTP SHOULD be used as the transport protocol.

5.2] 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).

5.3] 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.

6 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.

7 Security Considerations

RTP packets using the payload formats referred to in this specification are subject to the security considerations discussed in the RTP specification [1]. 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 and a secure execution model. MPEG-J content can call this set of APIs and Java(TM) methods from a set of Java packages supported in the receiver within the defined security model. According to this security model, downloaded byte code is forbidden to load libraries, define native methods, start programs, read or write files, or read system properties.

Receivers can implement intelligent filters to validate the buffer requirements or parametric (OD, BIFS, etc.) or programmatic (MPEG-J, ECMAScript) commands in the streams. However, this can increase the complexity significantly.

8 Multicast considerations

When using IP Multicast, the SDP information describing the ISO/IEC 14496 Session should be made available to the terminal. In addition, elementary stream descriptors may use URLs to directly address ESs. The goal of such URL would be to convey information to enable the terminal to directly connect to the RTP channel carrying the ES. The URL may contain the SDP information required to access the stream as described in section 10 above.

Acknowledgments

This draft has benefited greatly by contributions from many people, including Mike Coleman, Jean-Claude Duford, Viswanathan Swaminathan, Peter Westerink, Carsten Herpel, Olivier Avaro, Paul Christ, Zvi Lifshitz, and many others. Their insight, foresight, and contribution is gratefully acknowledged. Little has been invented here by the author; this is mostly a collation of greatness that has

gone before.

References

[1] H. Schulzrinne, et. al., "RTP : A Transport Protocol for Real-Time Applications", IETF RFC 1889, January 1996.

[2] H. Schulzrinne, et. al., "RTP Profile for Audio and Video Conference with Minimal Control", IETF RFC 1890, January 1996.

[3] H. Schulzrinne, et. al., " Real Time Streaming Protocol (RTSP)", IETF RFC 2326, April 1998.

[4] M. Handley, " SDP: Session Description Protocol", IETF RFC 2327, April 1998.

[5] D.Curet et al., " RTP Payload Format for MPEG-4 FlexMultiplexed Streams", IETF Draft, draft-curet-avt-rtp-mpeg4-flexmux-00.txt, Feburary 2001, expires August 2001.

[6] Yoshihiro Kikuchi et al., " RTP Payload Format for MPEG-4 Audio/Visual Streams", IETF RFC 3016, November 2000

[7] R. Finlayson, "A More Loss-Tolerant RTP Payload Format for MP3 Audio", IETF Draft, draft-ietf-avt-rtp-mp3-03.txt, Aug 3 2000, expires Feb 3 2001

[8] Kretschmer et al., "RTP Payload Format for MPEG-2 AAC Streams", IETF Draft, draft-ietf-avt-rtp-mpeg2aac-00.txt, June 25, 1999, expired December 25, 1999

[9] P. Gentric et al., "RTP Payload Format for MPEG-4

Streams", IETF
Draft, draf, draft-ietf-avt-mpeg4-multiSL-01.txt, July 20, 2001,
expires January 20, 2002

Authors' Contact Information

David Singer
Email: singer@apple.com
Tel: +1 408 974 3162

Apple Computer, Inc.
One Infinite Loop, MS:302-3MT
Cupertino CA 95014
USA

Young-Kwon LIM
E-mail : young@netntv.co.kr
TEL : +82-2-581-2305

net&tv Co., Ltd.
5th Floor Himart Building
1007-46 Sadang-Dong Dongjak-Gu
Seoul, 156-090, Korea