ISMA
Internet Streaming Media Alliance

2570 W. El Camino Real, Suite 304
Mountain View, CA 94040

T 650.949.6780
F 650.949.6785

www.isma.tv

Internet Streaming Media Alliance Implementation Specification

Version 1.0

28 August 2001

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

ISMA SPECIFICATION LIMITATIONS AND CONDITIONS OF USE

LEGAL LIMITATIONS AND CONDITIONS OF USE

USERS OF THE ISMA SPECIFICATION ARE NOT PERMITTED OR AUTHORIZED TO STATE OR CLAIM THAT THEIR PRODUCTS OR APPLICATIONS COMPLY WITH THE SPECIFICATION, PENDING ISMA'S DEVELOPMENT AND IMPLEMENTATION OF A COMPLIANCE OR CERTIFICATION PROGRAM AND USER'S EXPRESS AGREEMENT WITH THE TERMS AND CONDITIONS THEREOF. BY REQUESTING OR USING THE SPECIFICATION, USER AGREES TO THIS LIMITATION AND CONDITION.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

ISMA SPECIFICATION DISCLAIMER

LEGAL DISCLAIMER

THIS SPECIFICATION IS PROVIDED "AS IS" AND WITHOUT ANY WARRANTY OR REPRESENTATION OF ANY KIND, EXPRESS OR IMPLIED. WITHOUT LIMITATION, THERE IS NO WARRANTY OF NON-INFRINGEMENT, NO WARRANTY OF MERCHANTABILITY, AND NO WARRANTY OF FITNESS FOR A PARTICULAR PURPOSE. ALL WARRANTIES ARE EXPRESSLY DISCLAIMED.

USER ASSUMES THE FULL RISK OF USING THE SPECIFICATION. IN NO EVENT SHALL ISMA OR ANY MEMBER OF ISMA BE LIABLE FOR ANY ACTUAL, DIRECT, INDIRECT, PUNITIVE, OR CONSEQUENTIAL DAMAGES ARISING FROM SUCH USE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES."

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Table Of Contents

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

1 Acronyms & Terms

AVP	Audio Visual Profile (IETF RFC 1890)
BIFS	Binary Format for Scene
CIF	Common Intermediate Format (352 x 288)
cRTP	Compressed Real Time Protocol (IETF RFC 2508)
ESD	Elementary Stream Descriptor
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
IOD	Initial Object Descriptor
IP	Internet Protocol
IPv4	Internet Protocol Version 4
IPv6	Internet Protocol Version 6
ISMA	Internet Streaming Media Alliance
ISO	International Organization for Standardization
MP4	MPEG-4 File Format
M4IF	MPEG-4 Industry Forum
OCR	Object Clock Reference
OD	Object Descriptor
QCIF	Quarter Common Intermediate Format (176 x 144)
QoS	Quality of Service
RFC	Request for Comment
RTP	Real Time Protocol (IETF RFC 1889)
RTSP	Real Time Streaming Protocol (IETF RFC 2326)
SDP	Session Description Protocol (IETF RFC 2327)
TCP	Transmission Control Protocol (IETF RFC 793)
UDP	User Datagram Protocol (IETF RFC 768)
WMF	Wireless Multimedia Forum

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

2 Technical Specification:

2.1 Document Status

This version of the document is the final version proposed by the Technical committee and approved by the ISMA Board of Directors as version 1.0 of the ISMA Implementation Specification.

2.2 Scope

For several years, the promise of delivering video and audio over the Internet has been widely promoted. However, the fulfillment of these promises has been delayed due to a number of technical issues. One significant impediment has been the lack of a set of widely adopted technical standards for the transmission of video and audio on the Internet. This has led to a fragmentation of the market and a lack of critical mass necessary to achieve rapid deployment.

This implementation specification addresses this need by setting forth a framework for the use of existing open standards that vendors can use to build interoperable video and audio systems for use on IP networks and the Internet.

Note that this specification assumes use of the existing IPv4 infrastructure and does not require any deployments of new or advanced IP technology. However, this specification will allow conforming implementations to benefit from improvements in the IP infrastructure as they are deployed. Specifically, this implementation specification is independent of the existence of IPv6, IPsec, IP multicast, and IP QoS.

Note that this specification targets broadband quality networks (less than 1.5Mbps in raw bandwidth). The 1.0 Implementation Specification also addresses both narrowband (dialup) and wireless networks.
Further this specification assumes use of existing MPEG technologies and although initially focused on MPEG-4 technologies, future adaptations and revisions may include MPEG-2, 7 and other non-MPEG technologies.

2.3 Architecture

At its simplest, the architecture of ISMA consists of a media server, an IP network, and a media client. Media is transmitted from the media server to the media client in one of two modes: on-demand or broadcast.

In the on-demand mode, the media client requests media content from the media server that after appropriate verifications transmits the media to the client.
In the broadcast mode, the media server begins transmitting the media content to one or more media clients at a specified time. The media client is advised when to receive the media content via an out-of-band mechanism.
This simple model can be easily extended to allow for a number of intermediary systems in the transmission process that perform a range of services. These intermediary systems either work transparently from the viewpoint of both the media server and media client, or

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

the intermediary system presents a media server interface to media clients, and a media client interface to media servers. Examples of the services that can be provided by intermediary systems are: short-term or long-term storage, re-encoding to new bit rates, mixing with other media streams. Streaming media caches/proxies are the current preeminent example of such intermediary systems.

Note this architecture makes no assumptions that the transmission of media across the IP network must occur in “real-time” such that the media can be immediately rendered at the media client. In fact, we wish to enable media to flow in a mixture of real-time, and off-line transmissions, so as to best utilize network resources, and provide the best end-user experience.

Also note that this architecture makes no assumptions regarding the “production” phase of media content where capture, editing, and encoding of the media occurs. How such systems operate and what output they produce is an implementation issue for vendors producing media servers, encoders and the like.

2.4 Technology Selection Criteria

Three criteria are used in selecting the technologies for this implementation specification:

The first criterion is that the technology must be an open standard. The specifications must be available to all who wish to participate and the licensing of any associated intellectual property should be on reasonable terms for vendors of any size.

The second criterion is that the technology should enable interoperable products to be brought to market quickly. This argues for giving preference to technologies that already have multiple implementations and have wide deployment. This criterion also argues that the profile should aim for simplicity; seeking a solution that addresses the core needs of the Internet video and audio applications, and not the specialized needs of every conceivable application.

The third criterion is that the technology be forward-looking, and provide for new network media technologies and emerging video-enabled information appliances.

Note that these criteria may evolve in future versions of this document as more sophisticated standards and implementations evolve in the marketplace.

2.5 Definitions

Note this document uses the following definitions taken from IETF RFC 2119:

1. MUST This word, or the terms "REQUIRED" or "SHALL", mean that the definition is an absolute requirement of the specification.
2. SHOULD This word, or the adjective "RECOMMENDED", mean that there may exist valid reasons in particular circumstances to ignore a particular item, but the full implications must be understood and carefully weighed before choosing a different course.
3. MAY This word, or the adjective "OPTIONAL", mean that an item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because the vendor feels that it enhances the product while another vendor may omit the same item. An implementation that does not include a particular option MUST be

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

prepared to interoperate with another implementation that does include the option, though perhaps with reduced functionality. In the same vein an implementation that does include a particular option MUST be prepared to interoperate with another implementation that does not include the option (except, of course, for the feature the option provides.)

2.6 Functions

The functions of the Internet video and audio ecosystem can be described as follows.

2.6.1 Media Transmission

This is the means by which the media content is transmitted over the network, either in real-time or off-line.

2.6.2 Media Control

This is the means by which a viewer requests transmission of media content to them for rendering and optionally controls the transmission (e.g. pause, rewind, fast-forward).

2.6.3 Media Announcement

This is the means by which a viewer (either human or computer) discovers the existence of available media content and the information necessary to request access to the media content.

Note that a given single product may not implement all of these functions, but will contribute to the ecosystem's functionality in one or more areas. For example, a specialized video encoder product might only address the media transmission function.

2.7 Profiles

The following technical standards are the identified profiles for the ecosystem. To be compliant with this specification a product must completely implement profile 0, and may implement additional profiles. For example, an on-demand video server would likely need to implement all possible profiles to address a wide client base. However a decoder/terminal may only support profile 0.

This approach has been taken to ensure that any product certified as ISMA compliant, has the capability to minimally interoperate with any other ISMA compliant product.

This is a definition of base interoperability. Vendors are still free to add additional functionality beyond that specified in this document. However that said, a conforming product cannot make any additional requirements beyond this specification to interoperate with another conforming system.

2.7.1 All Profile Requirements

The following lists those requirements common to all profiles.

Media Decoding

A conforming implementation MUST implement either Video, or Audio, or both.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Transports

REQUIRED - RTP: A Transport Protocol for Real-Time Applications RTP
IETF RFC 1889

REQUIRED - RTP Profile for Audio and Video Conferences with Minimal Control
IETF RFC 1890

REQUIRED – User Datagram Protocol
IETF RFC 768

OPTIONAL - Interleaved RTSP & RTP/AVP over TCP transport
Real Time Streaming Protocol – IETF RFC 2326, section 10.12
Transmission Control Protocol - IETF RFC 793

Rationale: The RTP/AVP/UDP profile is the simplest and most widely supported option in current Internet streaming media systems. The RTSP/RTP interleave over TCP provides the option of reliable transport. Further RTSP/RTP over TCP will also permit traversal of Network Address Translators and Firewalls.

The media may be transmitted in unicast, multicast or in a broadcast manner. For the transmission of numerous simultaneous sessions it is RECOMMEDED that multicast be used for media transmission. It is further RECOMMENDED that a multicast implementation be conformant to IGMP v3. See section 4 for IGMPv3 references at time of publication.

RTP Payloads

REQUIRED (Video) - RFC 3016 RTP Payload Format for MPEG-4 Audio/Visual Streams
REQUIRED (Audio) - Appendix G

ISMA Restriction: Video decoder specific configuration information MUST be present in the SDP description of the media stream. This is a restriction of RFC 3016 that requires it to be in-band, at least once, and optionally present in the SDP description.

Note: cRTP – IETF RFC 2508 is an optional adjunct to RTP and can be used in conjunction with this specification in a transparent fashion.

Content Distribution

REQUIRED **–** MPEG-4 MP4 Format - ISO/IEC 14496-1:2000(E)

Rationale – In order to encourage content distribution and interoperability between products at the file storage level, the MPEG-4 MP4 Format, based on Apple's QuickTime file format, is the most obvious choice for MPEG-4 content.

ISMA compliant content, when stored in files, will contain the minimal BIFS and OD streams as described in Appendix E. The BIFS and OD streams MUST be stored in their own tracks. The usage of "data:" URLs to carry these streams (as described in Appendix F) is limited to transmission sessions. Of course, files may contain hint tracks to assist the preparation of the IOD for transmission. This restriction is required in order to avoid multiple flavors of same content.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

RECOMMENDED - MPEG MP4 Format:

In order to provide easy access to the sample dimensions, the initial 32 bit reserved field in VisualSampleEntry should be used as a 16 bits width followed by a 16 bit height of the video frame.

To facilitate a simple parsing scheme, the atoms that provide information for other atoms should be placed before the dependent atoms. For example, the header atom for a container should precede other atoms in the container that require information found in the header atom, the handler reference atom should precede the relevant information atom, and the Decoding Time to Sample Atom should precede the Composition Time to Sample Atom.

Media Control

REQUIRED - Real Time Streaming Protocol (RTSP)
IETF RFC 2326

The minimal RTSP implementation described in Appendix D of RFC2326 SHALL be used, with the addition of the DESCRIBE method.

RECOMMENDED – that a conforming RTSP implementation support the DESCIBE method. If the DESCRIBE method is used, it is REQUIRED that SDP is supported as the description format, as specified in Appendix C of RFC2326.

REQUIRED – since the RTP/AVP transport must be supported to conform to this specification. A conforming RTSP implementation MUST support RTP/AVP in the “Transport” header and MUST support the “RTP-Info” header. It SHOULD also support the “client_port”, “server_port”, "source", and “ssrc” parameters of the “Transport” header.

REQUIRED – that a conforming RTSP implementation support the PLAY method.

RECOMMENDED – that a conforming RTSP implementation accept/generate the RTSP headers: “Bandwidth”, “Cache-control”, “Expires”, “If-modified-since”, “Last-modified”, “User-Agent”, “Server”.

Media Announcement

REQUIRED - SDP: Session Description Protocol
IETF RFC 2327

The SDP data should be formatted according to SDP specification [RFC2327] and Appendix C of RFC2326.

Use the ISMA specific parameter:

a=isma-compliance:<profile>,<lowest-spec-version>,<authored-to-version>
where:
profile:

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

an integer specifying the ISMA profile to which the content conforms

lowest-spec-version:

a decimal number, indicating the lowest version number of the ISMA specification to which a client can conform, and still decode the content. Clients MUST not decode content with a lowest-spec-version higher than the highest specification version that they implement. The first published specification was 1.0; therefore this field is 1.0 or greater

authored-to-version:

the version of the specification against which the content was authored. Ideally the client also implements this version, whereupon the user can be more confident that the content is being completely decoded. A content author may choose to allow clients written to earlier versions of the spec achieve partial decode.

Servers MUST generate the SDP iod attribute using a binary iod (not XML); see Appendix E for details. If the URL is a data URL, then the binary IOD is encapsulated directly in the URL using base64 encoding.

Note that the media description in SDP form can be transmitted in a number of ways; examples include HTTP, RTSP, SMTP, SAP, and SIP.

2.7.2 Profile 0

Rationale: This profile was selected to allow for video and audio at bitrates suitable that match capabilities of narrowband and mobile wireless infrastructures and to align with the patent pool work in M4IF.

Video

REQUIRED - MPEG-4 ISO/IEC 14496-2:1999 + Cor 1:2000 + Cor 2:2001

Simple Profile @ Level 1

Typical Visual Session Size is QCIF (176x144)

Maximum bit rate is 64kbit/s

ISMA Restriction: Profile 0 is limited to one (1) video object only

Audio

REQUIRED - MPEG-4 ISO/IEC 14496-3:1999 and AMD1 2000

High Quality Audio Profile @ Level 2

Up to 2 channels

Up to 48000 Hz sampling rate

This profile contains both CELP and Low Complexity AAC

ISMA Restriction: Profile 0 is limited to one (1) audio object only

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Informative Note: Using the High Quality Audio Profile @ L2 may lead to a situation where the audio bit rate is greater than the video bit rate. Although unusual, there are scenarios where this may be desired.

2.7.3 Profile 1

Rationale: This profile was selected to allow for a richer streaming experience over infrastructures with broadband bit rates. Profile 1 is a superset of Profile 0 and can fully decode any streams generated by Profile 0 encoders.

Video

REQUIRED - MPEG-4 ISO/IEC 14496-2:1999 + Cor 1:2000 + Cor 2:2001
- Advanced Simple Profile @ Level 3
  - Typical Visual Session Size is CIF (352x288)
- ISMA Maximum Bitrate 1.5 Mbps

ISMA Restriction: Profile 1 is limited to one (1) video object only
Note: The specifications above do not detail out the Advanced Simple Profile. This is an approved profile in MPEG detailed in N3904 and will be noted in a future revision of 14496-2.

Audio

REQUIRED - MPEG-4 ISO/IEC 14496-3:1999 and AMD1 2000
- High Quality Audio Profile @ Level 2
  - Up to 2 channels
  - Up to 48000 Hz sampling rate
  - This profile contains both CELP and Low Complexity AAC

ISMA Restriction: Profile 1 is limited to one (1) audio object only

Cumulative Bit Rate Limitation

ISMA Restriction: The combined audio and video bitrates in a Profile 1 session is limited to 1.5 Mbps.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

3 Security Considerations

The current version of this framework does not add nor does it detract from any security that exists with the protocols/specifications it uses/refers to.

Future versions of this framework must address both network level security issues as well as digital rights management issues.

4 Future Work Areas

The following areas are possibilities for inclusion in a future version of this document.

Digital Rights Management:

MPEG-4 IPMP provides a framework for digital rights management (DRM), but today there is no standard that addresses the complete function of a DRM system. The ISO MPEG-4 group is currently soliciting submissions to fully address DRM. We would expect to consider the results of this standards effort for inclusion in ISMA.

Additional MPEG-4 Profiles:

Those that provide scalable encoding (aka temporal or spatial enhancement levels), and a mapping of the encoding levels to RTP streams such that network elements can provide differential QoS and/or rate-adaptation.

Profiles targeted at home entertainment level qualities, i.e. Main Profile @ L3, and Main Profile @ L4 or ACE profile.

Additional MPEG-4 Technologies: MPEG-4 Systems, or Synthetic Video, or Synthetic Audio.

Investigation into various techniques for firewall traversal. These could include using HTTP or RTSP proxies.

An RTSP mechanism for negotiating QoS for video and audio streams – there exist some proprietary schemes today, but no standard yet exists.

An RTP/RTCP retransmission and/or forward-error correction (FEC) mechanism – several proposals have been made within the IETF but it is still early in the standardization process for these capabilities.

Network level Quality of Service.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

5 References

Internet Engineering Task Force: www.ietf.org
Motion Picture Experts Group http://www.cselt.it/mpeg/

Postel, J., "User Datagram Protocol," RFC 768, USC/Information Sciences Institute, August 1980.

Postel, J, “Transmission Control Protocol”, RFC 793, USC/Information Sciences Institute, September 1981.

H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications." RFC 1889, Jan 1996.

H. Schulzrinne, “RTP Profile for Audio and Video Conferences with Minimal Control”, RFC 1890, Jan 1996.

S. Casner, V. Jacobson, “Compressing IP/UDP/RTP Headers for Low Speed Serial Links”, RFC 2508, February 1999.

H. Schulzrinne, A. Rao, R. Lamphier, “Real Time Streaming Protocol (RTSP)”, RFC 2326, April 1998

M. Handley, V. Jacobson, “SDP: Session Description Protocol”, RFC 2327, April 1998.

Y. Kikuchi, T. Nomura, S. Fukunaga, Y. Mutsui, H.Kimata, “RTP Payload Format for MPEG-4 Audio/Visual Streams”,RFC 3016, November 2000.

The following are IGMPv3 References at time of publication:
Internet Group Management Protocol, Version 3, Brad Cain et.al., IETF draft-ietf-idmr-igmp-v3-*.txt, Nortel, Cisco, Ericsson, Work In Progress

Source-Specific Multicast for IP, H. Holbrook, B. Cain, IETF draft-holbrook-ssm-arch-*.txt, Cisco, Cereva Networks, Work In Progress

Socket Interface Extensions for Multicast Source Filters D. Thaler, W. Fenner, B. Quinn, IETF draft-ietf-idmr-msf-api-*.txt ,Microsoft, AT&T Research, Stardust.com, Work In Progress

SDP Source-Filters, B.Quinn, IETF draft-ietf-mmusic-sdp-srcfilter-*.txt, Stardust.com, Work In Progress

Using IGMPv3 For Source-Specific Multicast, H. Holbrook, B. Cain, IETF draft-holbrook-idmr-igmpv3-ssm-*.txt, Cisco, Cereva Networks, Work In Progress

SDP Bandwidth Modifiers for RTCP Bandwidth, S. Casner, IETF draft-ietf-avt-rtcp-bw-*.txt, Packet Design, Work In Progress

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

RTCP Extension for Source Specific Multicast Sessions, J. Chesterfield, IETF draft-chesterfield-avt-rtcpssm-*.txt , AT&T Internet Research, Work In Progress

The following are MPEG references listed within this document:
ISO/IEC JTC1/SC 29/WG11 N3904, “Information technology – Coding of audio-visual objects – Part 2: Visual, Amendment 4 Streaming video profile”, ISO/IEC 14496-2:1999/FDAM 4

ISO/IEC JTC1/SC 29/WG11 N2501 + COR 1 + AMD 1, ISO/IEC 14496-1:2000(E), “Information technology – Coding of audio-visual objects – Part 1: Systems”
ISO/IEC JTC1 SC29 WG11 N4081 – Text of ISO/IEC 14496-8/CD

ISO/IEC 14496-1 Part 8 (Note: Work in progress)

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

6 Modification History

Version	Date	Editor	Changes
1.0.1	August 27, 2001	Technical Committee	Version 1.0 public specification.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

7 Contacts

ISMA Web Site www.ism-alliance.org

Implementation Specific Editor SpecEditor@ISM-Alliance.org
Executive Director of ISMA ExecDirector@ISM-Alliance.org

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

8 Appendix A: RTP/RTCP, RTSP, SDP Field Usage (Normative)

This appendix lists all parameters in RTP/RTCP, RTSP and SDP specifications that have been listed as optional that for ISMA purposes need to be mandatory, or fixed to certain values.

RTP/RTCP:

No parameters to date have been noted as needing to be fixed or mandatory.

RTSP:

No parameters to date have been noted as needing to be fixed or mandatory.

SDP:

a=isma-compliance:<profile>,<lowest-spec-version>,<authored-to-version>

where:

profile:

an integer specifying the ISMA profile to which the content conforms

lowest-spec-version:

a decimal number, indicating the lowest version number of the ISMA specification to which a client can conform, and still decode the content. Clients MUST not decode content with a lowest-spec-version higher than the highest specification version that they implement. The first published specification was 1.0; therefore this field is 1.0 or greater

authored-to-version:

the version of the specification against which the content was authored. Ideally the client also implements this version, whereupon the user can be more confident that the content is being completely decoded. A content author may choose to allow clients written to earlier versions of the spec achieve partial decode.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Appendix B: Hint Track Format (Normative)

8.1 Introduction

RTP is the real-time streaming protocol defined by the IETF (RFC 1889 and 1890) and is currently defined to be able to carry a limited set of media types (principally audio and video) and codings. The packing of MPEG-4 elementary streams into RTP is under discussion in both bodies. However, it is clear that the way the media is packetized does not differ in kind from the existing techniques used for other codecs in RTP, and supported by this scheme.

In standard RTP, each media stream is sent as a separate RTP stream; multiplexing is achieved by using IP's port-level multiplexing, not by interleaving the data from multiple streams into a single RTP session. However, if MPEG is used, it may be necessary to multiplex several media tracks into one RTP track (e.g. when using MPEG-2 transport in RTP, or FlexMux). Each hint track is therefore tied to a set of media tracks by track references. The hint tracks extract data from their media tracks by indexing through this table. Hint track references to media tracks have the reference type 'hint'.

This design decides the packet size at the time the hint track is created; therefore, in the declarations for the hint track, we indicate the chosen packet size. This is in the sample-description. Note that it is valid for there to be several RTP hint tracks for each media track, with different packet size choices. Similarly the time-scale for the RTP clock is provided. The timescale of the hint track is usually chosen to match the timescale of the media tracks, or a suitable value is picked for the server. In some cases, the RTP timescale is different (e.g. 90 kHz for some MPEG payloads), and this permits that variation. Session description (SAP/SDP) information is stored in user-data atoms in the track.

8.2 Sample Description Format

RTP hint tracks are hint tracks (media handler 'hint'), with an entry-format in the sample description of 'rtp ':

```
class RtpHintSampleEntry() extends SampleEntry ('rtp ') {
   uint(16)     hinttrackversion = 1;
   uint(16)     highestcompatibleversion = 1;
   uint(32)     maxpacketsize;
   atom       additionaldata[];
  }
```

The hinttrackversion is currently 1; the highest compatible version field specifies the oldest version with which this track is backward-compatible.

The maxpacketsize indicates the size of the largest packet that this track will generate.

The additional data is a set of atoms, from the following.

```
class timescaleentry() extends atom('tims') {
   unit(32)   timescale;
}

class timeoffset() extends atom('tsro') {
   int(32)    offset;
}
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
class sequenceoffset extends atom('snro') {
   int(32)    offset;
}
```

The timescale entry is required. The other two are optional. The offsets over-ride the default server behavior, which is to choose a random offset. A value of 0, therefore, will cause the server to apply no offset to the timestamp or sequence number respectively.

8.3 Sample Format

Each sample in the hint track will generate one or more RTP packets, whose RTP timestamp is the same as the hint sample time. Therefore, all the packets made by one sample have the same timestamp. However, provision is made to ask the server to 'warp' the actual transmission times, for data-rate smoothing, for example.

Each sample contains two areas: the instructions to compose the packets, and any extra data needed when sending those packets (e.g. an encrypted version of the media data). Note that the size of the sample is known from the sample size table.

```
aligned(8) class RTPsample {
   unsigned int(16)   packetcount;
   unsigned int(16)   reserved;
   RTPpacket  packets[packetcount];
   byte    extradata[];
}
```

8.3.1 Packet Entry format

Each packet in the packet entry table has the following structure:

```
aligned(8) class RTPpacket {
   int(32) relative-time;
   // the next fields form initialization for the RTP
   // header (16 bits), and the bit positions correspond
   bit(2)  reserved;
   bit(1)  P-bit;
   bit(1)  X-bit;
   bit(4)  reserved;
   bit(1)  M-bit;
   bit(7)  payload-type;

   unsigned int(16)   RTPsequenceseed;
   unsigned int(13)   reserved = 0;
   unsigned int(1) extra-flag;
   unsigned int(1) bframe-flag;
   unsigned int(1) repeat-flag;
   unsigned int(16)   entrycount;
   if (extra-flag) {
      unit(32) extra-information-length;
      atom extra-data-tlv[];
   }
   dataentry  constructors[entrycount];
}

class rtpoffsetTLV() extends atom('rtpo') {
   int(32) offset;
}
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

The relative-time field ‘warps’ the actual transmission time away from the sample time. This allows traffic smoothing. The following 2 bytes exactly overlay the RTP header;they assist the server in making the RTP header (the server fills in the remaining fields).

The sequence seed is the basis for the RTP sequence number. If a hint track causes multiple copies of the same RTP packet to be sent, then the seed value would be the same for them all. The server normally adds a random offset to this value (but see above, under ‘sequenceoffset’).

The extra-flag indicates that there is extra information after the constructors, in the form of type-length-value sets. Only one such set is currently defined; ‘rtpo’ gives a 32-bit signed integer offset to the actual RTP time-stamp to place in the packet. This enables packets to be placed in the hint track in decoding order, but have their presentation time-stamp in the transmitted packet be in a different order. This is necessary for some MPEG payloads. Note that the extra-information-length is the length in bytes of this field and all the TLV entries. Note also that the TLV atoms are aligned on 32-bit boundaries;the atom size indicates the actual bytes used, not the padded length. The extra-information-length will be correct.

The bframe-flag indicates a disposable ‘b-frame’. The repeat-flag indicates a ‘repeat packet’, one that is sent as a duplicate of a previous packet. Servers may wish to optimize handling of these packets.

8.3.2 Constructor format

There are various forms of the constructor. Each constructor is 16 bytes, to make iteration easier. The first byte is a union discriminator:

```
aligned(8) class RTPconstructor(type) {
   unsigned int(8) constructor-type = type;
}

aligned(8) class RTPnoopconstructor
   extends RTPconstructor(0)
{
   uint(8) pad[15];                     // 1 byte for type
}

aligned(8) class RTPimmediateconstructor
   extends RTPconstructor(1)
{
   unsigned int(8) count;
   unsigned int(8) data[count];
   unsigned int(8) pad[16-2-count];     // 1 byte for type, 1 byte for length
}

aligned(8) class RTPsampleconstructor
   extends RTPconstructor(2)
{
   unsigned int(8) trackrefindex;
   unsigned int(16)   length;
   unsigned int(32)   samplenumber;
   unsigned int(32)   sampleoffset;
   unsigned int(16)   bytesperblock = 1;
   unsigned int(16)   samplesperblock = 1;
}

aligned(8) class RTPsampledescriptionconstructor
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
   extends RTPconstructor(3)
{
   unsigned int(8) trackrefindex;
   unsigned int(16)   length;
   unsigned int(32)   sampledescriptionindex;
   unsigned int(32)   sampledescriptionoffset;
   unsigned int(32) reserved = 0;
}
```

The immediate mode permits the insertion of payload-specific headers (e.g. the RTP H.261 header). For hint tracks where the media is sent 'in the clear', the `sample` entry then specifies the bytes to copy from the media track, by giving the sample number, data offset, and length to copy. The track reference may index into the table of track references (a strictly positive value), name the hint track itself (-1), or the only associated media track (0). (The value zero is therefore equivalent to the value 1.)

The bytesperblock and samplesperblock concern compressed audio, using a scheme prior to MP4, in which the audio framing was not evident in the file. These fields have the fixed values of 0 for MP4 files.

The sampledescription mode allows sending of sample descriptions (which would contain elementary stream descriptors), by reference, as part of an RTP packet.
For complex cases (e.g. encryption or forward error correction), the transformed data would be placed into the hint samples, in the extradata field, and then sample mode referencing the hint track itself would be used.

Notice that there is no requirement that successive packets transmit successive bytes from the media stream. For example, to conform to RTP-standard packing of H.261, it is sometimes required that a byte be sent at the end of one packet and also at the beginning of the next (when a macroblock boundary falls within a byte).

8.4 SDP Information

Streaming servers using RTSP and SDP usually use SDP as the description format; and there are necessary relationships between the SDP information, and the RTP streams, such as the mapping of payload IDs to mime names. Provision is therefore made for the hinter to leave fragments of SDP information in the file, to assist the server in forming a full SDP description. Note that there are required SDP entries that the server should also generate. The information here is only partial.

SDP information is formatted as a set of atoms within user-data atoms, at both the movie and the track level. The text in the movie-level SDP atom should be placed before any media-specific lines (before the first 'm=' in the SDP file).

8.4.1 Movie SDP information

At the movie level, within the user-data ('udta') atom, a hint information container atom may occur:

```
aligned(8) class moviehintinformation extends atom('hnti') {
}

aligned(8) class rtpmoviehintinformation extends atom('rtp ') {
   uint(32) descriptionformat = 'sdp ';
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
   char  sdptext[];
}
```

The hint information atom may contain information for multiple protocols; only RTP is defined here. The RTP atom may contain information for various description formats; only SDP is defined here. The sdptext is correctly formatted as a series of lines, each terminated by <crlf>, as required by SDP.

8.4.2 Track SDP Information

At the track level, the structure is similar; however, we already know that this track is an RTP hint track, from the sample description. Therefore the child atom merely specifies the description format.

```
aligned(8) class trackhintinformation extends atom('hnti') {
}

aligned(8) class rtptracksdphintinformation extends atom('sdp ') {
   char sdptext[];
}
```

The sdptext is correctly formatted as a series of lines, each terminated by <crlf>, as required by SDP.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

9 Appendix C: Hint Track Example (Informative)

The following is an example of a hint track with 3 samples. Six packets are constructed from these three samples

```
mpeg4video.mov {
  time scale:                                   600
  duration:                                     60
  Descriptions {
    'rtp ' (14 octets):
      sdp b=AS:159
            73 64 70 20 62 3D 41 53 3A 31 35 39 0D 0A
  }
  Hint Tracks {
    Hinted Video Track {
      duration:                                 60
      media time scale:                         600
      Hinted Tracks {
        1:                                      'vide' track ID=1
      }
      Hint Info {
        payload-type:                           96 (X-ISMA)
        Descriptions {
          'sdp ' (73 octets):
           m=video 0 RTP/AVP 96
           b=AS:159
           a=rtpmap:96 X-ISMA
           a=control:trackID=3
        }
      }
      Sample Descriptions {
        Sample Description 1 {
          size:                                 36
          format:                               'rtp '
          data reference index:                 1
          hint track version:                   1
          highest compatible version:           1
          max packet size:                      1450
          Data {
            'tims' (4 octets):
              ....                  00 01 5F 90
          }
        }
      }
      Samples (15) {
        Sample 1 {
          sample time:                          0
          sample description index:             1
          Packets (4) {
            Packet 0 {
              RTP sequence number seed:         1
              relative transmission time:       0
              repeat-flag:                      false
              B-frame-flag:                     false
              RTP Header {
                P-bit:                          false
                X-bit:                          false
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
    M-bit:                          false
    payload-type:                   96
  }
  Payload (1 data entries) {
    data entry 0, constructor type 2 (sample data):
    track reference:                -1 (Hinted Video Track)
    sample number:                  1
    sample offset:                  116
    length:                         1438
    bytes per block:                0
    samples per block:              0
  }
}

Packet 1 {
  RTP sequence number seed:         2
  relative transmission time:       0
  repeat-flag:                      false
  B-frame-flag:                     false
  RTP Header {
    P-bit:                          false
    X-bit:                          false
    M-bit:                          false
    payload-type:                   96
  }
  Payload (1 data entries) {
    data entry 0, constructor type 2 (sample data):
    track reference:                -1 (Hinted Video Track)
    sample number:                  1
    sample offset:                  1554
    length:                         1438
    bytes per block:                0
    samples per block:              0
  }
}

Packet 2 {
  RTP sequence number seed:         3
  relative transmission time:       0
  repeat-flag:                      false
  B-frame-flag:                     false
  RTP Header {
    P-bit:                          false
    X-bit:                          false
    M-bit:                          false
    payload-type:                   96
  }
  Payload (1 data entries) {
    data entry 0, constructor type 2 (sample data):
    track reference:                -1 (Hinted Video Track)
    sample number:                  1
    sample offset:                  2992
    length:                         1438
    bytes per block:                0
    samples per block:              0
  }
}
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
            Packet 3 {
              RTP sequence number seed:          4
              relative transmission time:  0
              repeat-flag:                       false
              B-frame-flag:                      false
              RTP Header {
                P-bit:                           false
                X-bit:                           false
                M-bit:                           true
                payload-type:                    96
              }
              Payload (1 data entries) {
                data entry 0, constructor type 2 (sample data):
                track reference:                 -1 (Hinted Video Track)
                sample number:                   1
                sample offset:                   4430
                length:                          350
                bytes per block:                 0
                samples per block:               0
              }
            }
          }
        }
Sample 2 {
          sample time:                           20
          sample description index:              1
          Packets (1) {
            Packet 0 {
              RTP sequence number seed:          5
              relative transmission time:        0
              repeat-flag:                       false
              B-frame-flag:                      false
              RTP Header {
                P-bit:                           false
                X-bit:                           false
                M-bit:                           true
                payload-type:                    96
              }
              Payload (1 data entries) {
                data entry 0, constructor type 2 (sample data):
                track reference:                 -1 (Hinted Video Track)
                sample number:                   2
                sample offset:                   32
                length:                          428
                bytes per block:                 0
                samples per block:               0

              }
            }
          }
        }
Sample 3 {
          sample time:                           40
          sample description index:              1
          Packets (1) {
            Packet 0 {
              RTP sequence number seed:          6
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
                relative transmission time:        0
                repeat-flag:                       false
                B-frame-flag:                      false
                RTP Header {
                  P-bit:                           false
                  X-bit:                           false
                  M-bit:                           true
                  payload-type:                    96
                }
                Payload (1 data entries) {
                  data entry 0, constructor type 2 (sample data):
                  track reference:                 -1 (Hinted Video Track)
                  sample number:                   3
                  sample offset:                   32
                  length:                          364
                  bytes per block:                 0
                  samples per block:               0
                }
              }
            }
          }
        }
      }
    }
  }
}
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

10 Appendix D: RTSP and SDP Examples (Informative)

The following is an example of an RTSP session.

```
DESCRIBE rtsp://stream.isma.tv/foo.mp4 RTSP/1.0
CSeq: 1
Accept: application/sdp
Bandwidth: 4294967286
Accept-Language: en-US
User-Agent: ISMA-Client


RTSP/1.0 200 OK
Server: ISMA-Server
Cseq: 1
Content-length: 853
Content-Type: application/sdp
Content-Base: rtsp://stream.isma.tv/foo.mp4/

v=0
o=- 2890844256 2890842807 IN IP4 17.202.35.74
s=/foo.mp4
t=0 0
u=http://isma.tv/
e=admin@isma.tv
c=IN IP4 0.0.0.0
a=control:/
a=range:npt=0-   7.20000
a=ISMA-compliance:1,1.0,1
a=mpeg4-iod: "data:application/mpeg4-iod;base64,
AoF/AE8BAQEBAQOBEgABQHRkYXRh
OmFwcGxpY2F0aW9uL21wZWc0LW9kLWF1O2Jhc2U2NCxBVGdCR3dVZkF4Y0F5U1FBWlFRTklCRUFG
M0FBQVBvQUFBRERVQVlCQkFFWkFwOERGUUJsQlFRTlFCVUFCOUFBQUQ2QUFBQStnQVlCQXc9PQQN
/wUAAMgAAAAAAAAAAAYJAQAAAAAAAAAAA2EAAkA+ZGF0YTphcHBsaWNhdGlvbi9tcGVnNC1iaWZz
LWF1O2Jhc2U2NCx3QkFTZ1RBcUJYSmhCSWhRUlFVL0FBPT0EEgINAAAUAAAAAAAAAAAFAwAAQAYJ
AQAAAAAAAAAA"

m=audio 0 RTP/AVP 96
a=rtpmap:96 mpeg4-simple-A2/44100/2
a=control:trackID=5
a=fmtp:96 streamtype=5; profile-level-id=15; bitrate=64000;
config=9122620000; sizelength=13; indexlength=3;
indexdeltalength=3; profile=1
a=mpeg4-esid:101
m=video 0 RTP/AVP 97
a=rtpmap:97 MP4V-ES
a=control:trackID=6
a=fmtp:97 profile-level-id=1;
config=000001B001000001B5090000010000000120008440FA282C2090A21F
a=mpeg4-esid:201

SETUP rtsp://stream.isma.tv/foo.mp4/trackID=6 RTSP/1.0
CSeq: 2
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
Transport: RTP/AVP;unicast;client_port=6970-6971
User-Agent: ISMA-Client
Accept-Language: en-US


RTSP/1.0 200 OK
Server: ISMA-Server
Cseq: 2
Session: 2160059768067155308;timeout=60
Transport: RTP/AVP;unicast;client_port=6970-
6971;source=17.202.35.74;server_port=2000-2001


SETUP rtsp://stream.isma.tv/foo.mp4/trackID=5 RTSP/1.0
CSeq: 3
Transport: RTP/AVP;unicast;client_port=6972-6973
Session: 2160059768067155308
User-Agent: ISMA-Client
Accept-Language: en-US


RTSP/1.0 200 OK
Server: ISMA-Server
Cseq: 3
Session: 2160059768067155308
Transport: RTP/AVP;unicast;client_port=6972-
6973;source=17.202.35.74;server_port=2000-2001


PLAY rtsp://stream.isma.tv/foo.mp4 RTSP/1.0
CSeq: 4
Range: npt=0.000000-7.200000
Session: 2160059768067155308
User-Agent: ISMA-Client


RTSP/1.0 200 OK
Server: ISMA-Server
Cseq: 4
Session: 2160059768067155308
RTP-Info: url=trackID=6;seq=18230;rtptime=1426696138
 ,url=trackID=5;seq=27737;rtptime=1269965233


PAUSE rtsp://stream.isma.tv/foo.mp4 RTSP/1.0
CSeq: 5
Session: 2160059768067155308
User-Agent: ISMA-Client


RTSP/1.0 200 OK
Server: ISMA-Server
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
Cseq: 5
Session: 2160059768067155308


PLAY rtsp://stream.isma.tv/foo.mp4 RTSP/1.0
CSeq: 6
Range: npt=3.676667-7.200000
Session: 2160059768067155308
User-Agent: ISMA-Client


RTSP/1.0 200 OK
Server: ISMA-Server
Cseq: 6
Session: 2160059768067155308
RTP-Info: url=trackID=6;seq=18562;rtptime=1426707168
 ,url=trackID=5;seq=27853;rtptime=1270047053


TEARDOWN rtsp://stream.isma.tv/foo.mp4 RTSP/1.0
CSeq: 7
Session: 2160059768067155308
User-Agent: ISMA-Client

RTSP/1.0 200 OK
Server: ISMA-Server
Cseq: 7
Session: 2160059768067155308
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

11 Appendix E: Minimal MPEG-4 Systems Support in ISMA (Normative)

11.1 Embedding all MPEG-4 Systems data in the IOD

MPEG-4 Systems requires that the scene description and the object description are conveyed in separate BIFS and OD streams. However, in order to avoid the consumption of additional resources in the streaming transmission scenario, ISMA 1.0 requires that these streams will be embedded within the IOD that is then transmitted as part of the SDP description.

This mechanism is based on the following points:

- MPEG-4 allows using URLs in ESDs, as alternative to ESIDs
- A URL can be of type “data:”, i.e. the content pointed by the URL can be embedded in the URL itself.

Using these principles, the IOD of ISMA 1.0 compliant presentation looks as in Table 1:

Table 1 - A self-contained IOD for the simple scene

```
<InitialObjectDescriptor
        ODProfileLevelIndication= the proper profile indication
        SceneProfileLevelIndication= the proper profile indication
        AudioProfileLevelIndication= the proper profile indication
        VisualProfileLevelIndication= the proper profile indication
        graphicsProfileLevelIndication= the proper profile indication
>
        <esDescr>
                <ES_Descriptor
                        URLString=”data:application/mpeg4-od-au;base64,<Base64 encoded OD
stream >”
                >
                        <decConfigDescr>
                                <DecoderConfigDescriptor streamType="1" bufferSizeDB=”100” />
                        </decConfigDescr>
                        <slConfigDescr>
                                <SLConfigDescriptor predefined="1" />
                        </slConfigDescr>
                </ES_Descriptor>
                <ES_Descriptor
                        URLString=”data:application/mpeg4-bifs-au;base64,<Base64 encoded BIFS
stream >”
                >
                        <decConfigDescr>
                                <DecoderConfigDescriptor
                                        objectTypeIndication="2"
                                        streamType="3"
                                        bufferSizeDB=”100”
                                >
                                        <decSpecificInfo>
                                                <BIFSv2Config pixelMetric=”1”/>
                                        </decSpecificInfo>
                                /DecoderConfigDescriptor>
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
                    </decConfigDescr>
                    <slConfigDescr>
                        <SLConfigDescriptor predefined="1" />
                    </slConfigDescr>
                </ES_Descriptor>
            </esDescr>
</InitialObjectDescriptor>
```

The code in Table 1 assumes that "data:" URLs contain one access unit with null SL packet header.

11.2 Using SDP to Convey MPEG-4 Systems Information

The IOD shall be included in the SDP description, using the *mpeg4-iod* parameter described in ISO/IEC 14496-8 as:

a=mpeg4-iod [<location>*]*

The *location* shall be a URL enclosed in double-quotes, which will supply the IOD in its standard binary format. The IOD may be embedded in a "data:" URL and Base64 encoding of binary data (described in RFC 1341). The media stream will be described in the SDP using the usual parameters. In addition, there is a need to associate ESIDs with the corresponding stream description. This will be done as described in ISO/IEC 14496-9, i.e. a stream-specific attribute shall be present for each MPEG-4 stream. The attribute will take the following form:

a=mpeg4-esid esid

when *esid* is the ESID.

11.3 IOD Binary Syntax

Table 2 specifies the binary syntax of the entire IOD, which includes the embedded OD and BIFS stream. The Value column contains a numerical value when the field has a fixed value, and textual description when the value has to be generated by the module that creates the IOD. Note that some of the fields, like ES_ID, get arbitrary fixed values by ISMA. The rows in the table are grouped according the hierarchical structure of the descriptors. Descriptor names are indented to reflect their nesting.

Table 2 - IOD binary syntax

Descriptor Name				
Field No.	**Size in Bits**	**Field Name**	**Value**	**Comments**
InitialObjectDescriptor				
1	8	InitialObjectDescriptor tag	2	
2	8 or 16	descriptor size	Total size in bytes of fields 3 through 126	If the value is less than 128 it is encoded as a one-byte value. Otherwise 2 bytes are used – the first bit (MSB) of the first byte is 1 to indicate an escape to two bytes. The first bit of

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

				the second byte is left zero. The value is encoded as a 14-bit value, using the last 7 bits of each byte; first byte contains the high-order part
3	10	ObjectDescriptorID	1	
4	1	URL_Flag	0	
5	1	includeInlineProfilesFlag	0	
6	4	Reserved	15	
7	8	ODProfileLevelIndication	P&L value for very simple OD	In standardization process by MPEG
8	8	sceneProfileLevelIndication	P&L value for very simple BIFS	In standardization process by MPEG
9	8	audioProfileLevelIndication	P&L value for audio	
10	8	visualProfileLevelIndication	P&L value for video	
11	8	graphicsProfileLevelIndication	1	This MAY be 255 (no graphics capability required) if the presentation does not include video
	ES_Descriptor (of OD stream)			
12	8	ES_Descriptor tag	3	
13	8 or 16	Descriptor size	Total size in bytes of fields 14 through 91	See comments to field no. 2
14	16	ES_ID	1	
15	1	streamDependenceFlag	0	
16	1	URL_Flag	1	
17	1	OCRstreamFlag	0	
18	5	streamPriority	0	
19	8	string-size	Total size in bytes of fields 20 through 77 after encoding the data portion of the URL in Base 64	
20	36*8	URLstring	"data:application/mpeg4-od-au;base64,"	The binary data in fields 21 through 64 is encoded in Base64. All field lengths, shown in the table before the encoding, shall be calculated accordingly
		ObjectDescriptorUpdate		
21	8	ObjectDescriptorUpdate tag	1	
22	8	Descriptor size	Total size in bytes of fields 23 through 77	
		ObjectDescriptor (of video)		
23	8	ObjectDescriptor tag	1	Fields 23 through 50 exist only if the presentation includes video

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

24	8	descriptor size	Total size in bytes of fields 25 through 50	
25	10	ObjectDescriptorID	20	
26	1	URL_Flag	0	
27	5	reserved	31	
		ES_Descriptor (of video)		
28	8	ES_Descriptor tag	3	
29	8	descriptor size	Total size in bytes of fields 30 through 50	
30	16	ES_ID	201	
31	1	streamDependenceFlag	0	
32	1	URL_Flag	0	
33	1	OCRstreamFlag	1	See notes on synchronization below
34	5	streamPriority	4	May be zero. If used indicates the priority of the stream relative to the audio stream
35	16	OCR_ES_Id	101	See notes on synchronization below
		DecoderConfigDescriptor (of video)		
36	8	DecoderConfigDescriptor tag	4	
37	8	descriptor size	Total size in bytes of fields 37 through 47	
38	8	objectTypeIndication	32	32 indicates MPEG-4 video
39	6	streamType	4	
40	1	upStream	0	
41	1	reserved	1	
42	24	bufferSizeDB	A proper decoding buffer size, depends on stream properties	
43	32	maxBitrate	The maximum bitrate of the video steam	
44	32	avgBitrate	The average bitrate of the video stream	
			DecoderSpecificInfo (of video)	
45	8	DecoderSpecificInfo tag	5	Fields 45 through 47 may be omitted if the decoder can retrieve the DecoderSpecificInfo from another source (e.g. a specific SDP parameter)
46	8	descriptor size	Total size in bytes of field 47	
47	8*size in bytes of this field	DecoderSpecificInfo data	DecoderSpecificInfo for the specific video stream as specified in ISO/IEC 14496-2	

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

		SLConfigDescriptor (for the video stream)		
48	8	SLConfigDescriptor tag	6	
49	8	descriptor size	1	
50	8	predefined	4	A predefined SL-config value for ISMA video needs to be standardized by MPEG
		ObjectDescriptor (of audio)		
51	8	ObjectDescriptor tag	1	Fields 51 through 77 exist only if the presentation includes audio
52	8	descriptor size	Total size in bytes of fields 53 through 77	
53	10	ObjectDescriptorID	10	
54	1	URL_Flag	0	
55	5	reserved	31	
		ES_Descriptor (of audio)		
56	8	ES_Descriptor tag	3	
57	8	descriptor size	Total size in bytes of fields 58 through 77	
58	16	ES_ID	101	
59	1	streamDependenceFlag	0	
60	1	URL_Flag	0	
61	1	OCRstreamFlag	0	See notes on synchronization below
62	5	streamPriority	5	May be zero. If used indicates the priority of the stream relative to the video stream
		DecoderConfigDescriptor (of audio)		
63	8	DecoderConfigDescriptor tag	4	
64	8	descriptor size	Total size in bytes of fields 65 through 74	
65	8	objectTypeIndication	64	64 indicates MPEG-4 audio
66	6	streamType	5	
67	1	upStream	0	
68	1	reserved	1	
69	24	bufferSizeDB	A proper decoding buffer size, depends on stream properties	
70	32	maxBitrate	The maximum bitrate of the audio steam	
71	32	avgBitrate	The average bitrate of the audio stream	
			DecoderSpecificInfo (of audio)	
72	8	DecoderSpecificInfo tag	5	Fields 72 through 74 may be omitted if the decoder

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

				can retrieve the DecoderSpecificInfo from another source (e.g. a specific SDP parameter)
73	8	descriptor size	Total size in bytes of field 74	
74	8*size in bytes of this field	DecoderSpecificInfo data	DecoderSpecificInfo for the specific audio stream as specified in ISO/IEC 14496-3	
		SLConfigDescriptor (of audio)		
75	8	SLConfigDescriptor tag	6	
76	8	descriptor size	1	
77	8	predefined	3	A predefined SL-config value for ISMA audio needs to be standardized by MPEG
		DecoderConfigDescriptor (of OD stream)		
78	8	DecoderConfigDescriptor tag	4	
79	8	descriptor size	Total size in bytes of fields 80 through 86	
80	8	objectTypeIndication	1	
81	6	streamType	1	
82	1	upStream	0	
83	1	reserved	1	
84	24	bufferSizeDB	Large enough to hold the OD stream embedded in the IOD	
85	32	maxBitrate	0	
86	32	avgBitrate	0	
		SLConfigDescriptor (of OD stream)		
87	8	SLConfigDescriptor tag	6	
88	8	descriptor size	9	
89	8	predefined	1	
90	32	startDecodingTimeStamp	0	
91	32	startCompositionTimeStamp	0	
	ES_Descriptor (of BIFS stream)			
92	8	ES_Descriptor tag	3	
93	8	descriptor size	Total size in bytes of fields 94 through 126	
94	16	ES_ID	2	
95	1	streamDependenceFlag	0	
96	1	URL_Flag	1	
97	1	OCRstreamFlag	0	
98	5	StreamPriority	0	
99	8	string-size	Total size in bytes of fields 100 though 101 after	

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

			encoding the data portion of the URL in Base 64	
100	38*8	URLstring	"data:application/mpeg4-bifs-au;base64,"	The binary data in field 99 is encoded in Base64. The field lengths, shown in the table before the encoding, shall be calculated accordingly
101	8*size in bytes of this field	URLstring	For a scene with video+audio: hex "C0 10 12 81 30 2A 05 72 61 04 88 50 45 05 3F 00" For audio-only: hex "C0 10 12 81 30 2A 05 7C" For video-only: hex "C0 10 12 61 04 88 50 45 05 3F 00"	
		DecoderConfigDescriptor (of BIFS)		
102	8	DecoderConfigDescriptor tag	4	
103	8	descriptor size	18	
104	8	ObjectTypeIndication	2	
105	6	StreamType	3	
106	1	Upstream	0	
107	1	Reserved	1	
108	24	BufferSizeDB	20	
109	32	MaxBitrate	0	
110	32	AvgBitrate	0	
		BIFSv2Config		
111	8	BIFSv2Config tag	5	
112	8	descriptor size	3	
113	1	use3DmeshCoding	0	
114	1	UsePredictiveMFField	0	
115	5	NodeIDbits	0	
116	5	RouteIDbits	0	
117	5	ProtoIDbits	0	
118	1	IsCommandStream	1	
119	1	PixelMetric	0	
120	1	HasSize	0	
121	4	byte align	0	
		SLConfigDescriptor (of BIFS)		
122	8	SLConfigDescriptor tag	6	
123	8	descriptor size	9	
124	8	Predefined	1	
125	32	startDecodingTimeStamp	0	
126	32	StartCompositionTimeStamp	0	

Notes on Synchronization:

ISMA presentation will use the RTP mechanism for synchronization. Nevertheless, the description and the table above refer to the MPEG-4 notion of OCR (Object Clock Reference). It is assumed that players that depend on the MPEG-4 Systems model for

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

synchronization, will be able to map the RTP mechanism to the OCR model. Specifically, the server time reported through Sender Report commands in RTCP will be translated into OCRs when the player performs SL mapping. In presentations that include video and audio, it is recommended to assume that the OCRs are conveyed with one of the stream, and the other stream depends on this stream for synchronization. Table 2 follows this model, and assumes that the audio stream is the one that carries the OCRs, while the video stream clock depends on the audio stream. If this pattern does not apply, for instance, if the video stream plays continuously and the audio stream is sometime muted, the values in the table should be changed accordingly. Specifically, *OCRstreamFlag* should be set to 1 if the stream clock depends on another stream, and, in this case, and only in this case, the *ES_Descriptor* of the stream shall include a *OCR_ES_Id* field (as in row 35 in the table) that contains the ES_id of the stream that carries the OCRs.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

12 MP4 BIFS Track

ISMA 1.0 compliant MP4 files must include a BIFS track. ISMA 1.0 content is a presentation with a simple scene that contains at most one audio stream and one video stream. In BIFS terms, the scene contains a node for the audio object, and a rectangle node, whose texture is the video. In case the presentation includes only video or only audio, only one of these nodes would be present.

In terms of BIFS, an audio and video scene would look as in Table 3 (using VRML format for readability):

Table 3 - BIFS representation of a ISMA 1.0 audio + video scene

```
OrderedGroup {
    children [
        Sound2D {source AudioSource {url 10}}
        Shape {
                geometry Bitmap {}
                appearance Appearance {texture MovieTexture {url 20}}
        }
    ]
}
```

Table 4 shows the description of a video-only scene:

Table 4 - BIFS representation of a ISMA 1.0 video-only scene

```
OrderedGroup {
    children [
        Shape {
                geometry Bitmap {}
                appearance Appearance {texture MovieTexture {url 20}}
        }
    ]
}
```

Table 5 shows the description of an audio-only scene:

Table 5 - BIFS representation of an ISMA 1.0 audio-only scene

```
OrderedGroup {
    children [
        Sound2D {source AudioSource {url 10}}
    ]
}
```

12.1 MP4 OD Track

ISMA 1.0 compliant MP4 files must include an OD track. Given the simple BIFS scene of ISMA 1.0 there is one *ObjectDescriptorUpdate* command in the OD track that contains either one or two *ObjectDescriptors*; one for audio, one for video. Each of the ODs contains one ESD.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Table 6 shows the OD stream of an ISMA 1.0 audio and video scene using the XMT format (XML-based format standardized by MPEG for authoring purposes). Note it is included for clarification only, and does not show all the parameters.

Table 6 - OD stream for ISMA 1.0 simple scene

```
< ObjectDescriptorUpdate>
        <OD>
                <ObjectDescriptor objectDescriptorID="od:10">
                        <esDescr>
                                <ES_Descriptor ES_ID="es:101">
                                        <decConfigDescr>
                                                <DecoderConfigDescriptor
                                                        objectTypeIndication="64"
                                                        streamType="5"
                                                        bufferSizeDB="500"
                                                >
                                                        <decSpecificInfo>
                                                                <AudioSpecificInfo audio
decoder config info goes here/>
                                                        </decSpecificInfo>
                                                /DecoderConfigDescriptor>
                                        </decConfigDescr>
                                        <slConfigDescr>
                                                <SLConfigDescriptor predefined="2" />
                                        </slConfigDescr>
                                </ES_Descriptor>
                        </esDescr>
                </ObjectDescriptor>
                <ObjectDescriptor objectDescriptorID="od:20">
                        <esDescr>
                                <ES_Descriptor ES_ID="es:201">
                                        <decConfigDescr>
                                                <DecoderConfigDescriptor
                                                        objectTypeIndication="32"
                                                        streamType="4"
                                                        bufferSizeDB="5000"
                                                >
                                                        <decSpecificInfo>
                                                                <VideoSpecificInfo video
decoder config info goes here/>
                                                        </decSpecificInfo>
                                                </DecoderConfigDescriptor>
                                        </decConfigDescr>
                                        <slConfigDescr>
                                                <SLConfigDescriptor predefined="2" />
                                        </slConfigDescr>
                                </ES_Descriptor>
                        </esDescr>
                </ObjectDescriptor>
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

```
        </OD>
< /ObjectDescriptorUpdate>
```

Explanation: There is one *ObjectDescriptorUpdate* command that contains two *ObjectDescriptors*, one for audio and one for video, with ID 10 and 20 respectively. Each of the ODs contains one ESD, with ID 101 and 201 respectively.

The value of *objectTypeIndication* is, according to the standard, 64 for MPEG-4 audio and 32 for MPEG-4 video. Likewise *streamType* is 4 for video and 5 for audio. The size of the decoding buffer is specified for each stream, in order to comply with MPEG-4 buffer model.

The *decSpecificInfo* field contains decoder specific info. There is Video specific info and audio specific info as defined in 14496-2 and 14496-3 respectively. E.g. for video, this contains the video headers – VOSH, VOL, etc.

SLConfigDescriptor contains parameters that are used to interpret the SL packet headers of the data. SL packets are described in MPEG-4 Systems. The various RTP payload format specifications for MPEG-4 streams, explain how the map RTP packets can be mapped to SL packets,

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

13 Appendix F: Example of Minimal MPEG-4 Systems Support in ISMA (Informative)

The example below is based on Table 2 of Appendix E, with specific values assigned to variable fields as in Table 2. Note that The DecoderSpecificInfo descriptors of the audio and video streams are omitted, following the rule specified in Appendix E that DecoderSpecificInfo may be omitted if provided in other means. In this example this information is provided in *config* SDP parameters.

Table F-1 – A sample IOD - binary syntax

Descriptor Name			
Field No.	**Size in Bits**	**Field Name**	**Value**
InitialObjectDescriptor			
1	8	InitialObjectDescriptor tag	2
2	16	descriptor size	0x817f
3	10	ObjectDescriptorID	0
4	1	URL_Flag	0
5	1	includeInlineProfilesFlag	0
6	4	Reserved	15
7	8	ODProfileLevelIndication	1
8	8	sceneProfileLevelIndication	1
9	8	audioProfileLevelIndication	1
10	8	visualProfileLevelIndication	1
11	8	graphicsProfileLevelIndicati on	1
	ES_Descriptor (of OD stream)		
12	8	ES_Descriptor tag	3
13	8	Descriptor size	0x8112
14	16	ES_ID	1
15	1	streamDependenceFlag	0
16	1	URL_Flag	1
17	1	OCRstreamFlag	0
18	5	streamPriority	0
19	8	string-size	110
20	36*8	URLstring	"data:application/mpeg4-od-au;base64,"
		ObjectDescriptorUpdate (Fields 21 through 77 are shown before the encoding in Base64)	
21	8	ObjectDescriptorUpdate tag	1
22	8	Descriptor size	56
		ObjectDescriptor (of video)	
23	8	ObjectDescriptor tag	1
24	8	descriptor size	27
25	10	ObjectDescriptorID	20
26	1	URL_Flag	0
27	5	Reserved	31
		ES_Descriptor (of video)	
28	8	ES_Descriptor tag	3
29	8	descriptor size	23
30	16	ES_ID	201

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

31	1	streamDependenceFlag	0
32	1	URL_Flag	0
33	1	OCRstreamFlag	1
34	5	streamPriority	4
35	16	OCR_ES_Id	101
		DecoderConfigDescriptor (of video)	
36	8	DecoderConfigDescriptor tag	4
37	8	descriptor size	13
38	8	objectTypeIndication	32
39	6	streamType	4
40	1	upStream	0
41	1	Reserved	1
42	24	bufferSizeDB	6000
43	32	maxBitrate	64000
44	32	avgBitrate	50000
		SLConfigDescriptor (for the video stream)	
48	8	SLConfigDescriptor tag	6
49	8	descriptor size	1
50	8	predefined	4
		ObjectDescriptor (of audio)	
51	8	ObjectDescriptor tag	1
52	8	descriptor size	25
53	10	ObjectDescriptorID	10
54	1	URL_Flag	0
55	5	Reserved	31
		ES_Descriptor (of audio)	
56	8	ES_Descriptor tag	3
57	8	descriptor size	21
58	16	ES_ID	101
59	1	streamDependenceFlag	0
60	1	URL_Flag	0
61	1	OCRstreamFlag	0
62	5	streamPriority	5
		DecoderConfigDescriptor (of audio)	
63	8	DecoderConfigDescriptor tag	4
64	8	descriptor size	13
65	8	objectTypeIndication	64
66	6	streamType	5
67	1	upStream	0
68	1	Reserved	1
69	24	bufferSizeDB	2000
70	32	maxBitrate	16000
71	32	avgBitrate	16000
		SLConfigDescriptor (of audio)	
75	8	SLConfigDescriptor tag	6
76	8	descriptor size	1
77	8	predefined	3

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

		DecoderConfigDescriptor (of OD stream)	
78	8	DecoderConfigDescriptor tag	4
79	8	descriptor size	13
80	8	objectTypeIndication	1
81	6	streamType	1
82	1	upStream	0
83	1	Reserved	1
84	24	bufferSizeDB	200
85	32	maxBitrate	0
86	32	avgBitrate	0
		SLConfigDescriptor (of OD stream)	
87	8	SLConfigDescriptor tag	6
88	8	descriptor size	9
89	8	predefined	1
90	32	startDecodingTimeStamp	0
91	32	startCompositionTimeStamp	0
	ES_Descriptor (of BIFS stream)		
92	8	ES_Descriptor tag	3
93	8	descriptor size	97
94	16	ES_ID	2
95	1	streamDependenceFlag	0
96	1	URL_Flag	1
97	1	OCRstreamFlag	0
98	5	StreamPriority	0
99	8	string-size	54
100	36*8	URLstring	"data:application/mpeg4-bifs-au;base64,"
101	24*8	URLstring	Binary data before encoding in Base64: 0x"C0 10 12 81 30 2A 05 72 61 04 88 50 45 05 3F 00"
		DecoderConfigDescriptor (of BIFS)	
102	8	DecoderConfigDescriptor tag	4
103	8	Descriptor size	18
104	8	ObjectTypeIndication	2
105	6	StreamType	3
106	1	Upstream	0
107	1	Reserved	1
108	24	BufferSizeDB	20
109	32	MaxBitrate	0
110	32	AvgBitrate	0
		BIFSv2Config	
111	8	BIFSv2Config tag	5
112	8	Descriptor size	3
113	1	use3DmeshCoding	0
114	1	UsePredictiveMFField	0
115	5	NodeIDbits	0
116	5	RouteIDbits	0
117	5	ProtoIDbits	0

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

118	1	IsCommandStream	1
119	1	PixelMetric	0
120	1	HasSize	0
121	4	byte align	0
		SLConfigDescriptor (of BIFS)	
122	8	SLConfigDescriptor tag	6
123	8	descriptor size	9
124	8	Predefined	1
125	32	startDecodingTimeStamp	0
126	32	startCompositionTimeStamp	0

This table yields the following hexadecimal string:

02 81 7F 00 4F 01 01 01 01 01 03 81 12 00 01 40
74 64 61 74 61 3A 61 70 70 6C 69 63 61 74 69 6F
6E 2F 6D 70 65 67 34 2D 6F 64 2D 61 75 3B 62 61
73 65 36 34 2C 41 54 67 42 47 77 55 66 41 78 63
41 79 53 51 41 5A 51 51 4E 49 42 45 41 46 33 41
41 41 50 6F 41 41 41 44 44 55 41 59 42 42 41 45
5A 41 70 38 44 46 51 42 6C 42 51 51 4E 51 42 55
41 42 39 41 41 41 44 36 41 41 41 41 2B 67 41 59
42 41 77 3D 3D 04 0D 01 05 00 00 C8 00 00 00 00
00 00 00 00 06 09 01 00 00 00 00 00 00 00 00 03
61 00 02 40 3E 64 61 74 61 3A 61 70 70 6C 69 63
61 74 69 6F 6E 2F 6D 70 65 67 34 2D 62 69 66 73
2D 61 75 3B 62 61 73 65 36 34 2C 77 42 41 53 67
54 41 71 42 58 4A 68 42 49 68 51 52 51 55 2F 41
41 3D 3D 04 12 02 0D 00 00 14 00 00 00 00 00 00
00 00 05 03 00 00 40 06 09 01 00 00 00 00 00 00
00 00

Which, when encoded in Base64, generates the following string:

```
AoF/AE8BAQEBAQOBEgABQHRkYXRhOmFwcGxpY2F0aW9uL21wZWc0LW9kLWF1O2Jhc2U2NCxBVGdC
R3dVZkF4Y0F5U1FBWlFRTklCRUFGM0FBQVBvQUFBRERVQVlCQkFFWkFwOERGUUJsQlFRTlFCVUFC
OUFBQUQ2QUFBQStnQVlCQXc9PQQNAQUAAMgAAAAAAAAAAYJAQAAAAAAAAAA2EAAkA+ZGF0YTph
cHBsaWNhdGlvbi9tcGVnNC1iaWZzLWF1O2Jhc2U2NCx3QkFTZ1RBcUJYSmhCSWhRUlFVL0FBPT0E
EgINAAAUAAAAAAAAAAAFAwAAQAYJAQAAAAAAAAA
```

See Appendix D for an example of how the encoded IOD is used in the SDP description of an RTSP session.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

14 Appendix G: RTP Payload Format for AAC and CELP (Normative)

Status

This appendix is the RTP payload format for transport of MPEG-4 AAC and CELP as defined by ISMA. The format is capable of transporting any MPEG-4 elementary stream, but only AAC and CELP configurations are currently defined. In the future, other MPEG-4 elementary streams can be supported by definition of more configurations.

Abstract

The MPEG Committee (ISO/IEC JTC1/SC29 WG11) is a working group in ISO that recently produced the MPEG-4 [1] standard. MPEG defines tools to compress content such as audio-visual information into elementary streams. In [6] a generic RTP payload format is defined for transport of any non-multiplexed MPEG-4 elementary stream. To achieve the generic MPEG-4 functionality, [6] addresses detailed issues related to the MPEG-4 SL layer. However, many initial applications will not use the SL Layer. To facilitate usage of [6] by such applications, this document describes how to use [6] when no SL layer is used.

This specification is a product of the Audio/Video Transport working group within the Internet Engineering Task Force. Comments are solicited and should be addressed to the working group's mailing list at avt@ietf.org and/or the authors.

1. Introduction

The MPEG Committee is Working Group 11 (WG11) in ISO/IEC JTC1 SC29 that specified the MPEG-1, MPEG-2 and, more recently, the MPEG-4 standards [1]. The MPEG-4 standard specifies compression of audio-visual data into for example an audio or video elementary stream. In the MPEG-4 standard, these streams take the form of audiovisual objects that may be arranged into an audio-visual scene by means of a scene description. Each MPEG-4 elementary stream consists of a sequence of Access Units; in case of audio an Access Unit (AU) is an audio frame and in case of video a picture.

The MPEG-4 system specification is a rather abstract specification in the sense that no transport format for MPEG-4 elementary streams is defined. Instead, a conceptual SL layer has been specified to store transport specific information such as time stamps and random access point information. When transporting an MPEG-4 elementary stream, transport information from the SL layer is typically mapped to the actual transport layer. Note however that the SL layer is conceptual and may not exist in practice.

In [6], a general payload format is defined for transport of a single MPEG-4 elementary stream over RTP. The RTP payload format specified in [6] allows for carriage of any information that may be contained in the MPEG-4 SL layer, either by mapping to the RTP header fields or by carriage in specific fields defined in the RTP payload. Consequently, the

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

format defined in [6] is very generic and complete; for example, transcoding issues from and to the SL layer are described in detail.

However, in many initial MPEG-4 applications the SL layer does not exist in practice. Such applications do not require any knowledge of the SL layer. While the use of [6] is highly desirable for all MPEG-4 applications, to understand [6] may be difficult without knowledge of the MPEG-4 SL layer. Therefore in this document the use of [6] is described without requiring knowledge of the SL layer to understand its functionality.

Sophisticated features on interleaving of fragmented Access Units are defined in [6]. Because initial applications do not require these complicated features, these features are not supported in this document. Hence, only a functional subset of [6] is supported.

In [6], a general and configurable payload structure is defined for transport of MPEG-4 streams. This allows for the design of receivers that can be configured to receive any MPEG-4 stream. Configuration of the payload is provided to accommodate transport of any MPEG-4 stream, but for a specific MPEG-4 elementary stream typically only very few configurations are needed. So as to allow for the design of simplified, but dedicated receivers, this specifications requires that specific modes are defined for transport of MPEG-4 streams. In this document only modes are defined for transport of MPEG-4 CELP and AAC streams, but in future new RFCs are expected to specify additional modes for transport of other MPEG-4 streams.

In summary, this document:
- is intended for applications that do not apply the SL layer;
- describes how to use [6] without requiring knowledge of the SL layer;
- defines a functional but true subset of [6];
- defines modes how to use this specification for transport of MPEG-4 CELP and AAC streams.

The use of [6] defined in this document is simple to implement and reasonably efficient. It allows for optional interleaving of Access Units (such as audio frames) to increase error resiliency in packet loss.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [3].

2. Carriage of MPEG-4 elementary streams over RTP

2.1 MPEG-4 stream type identification

Information on the type of MPEG-4 stream that is carried in the payload is conveyed by format parameters in an SDP message or by other means.

2.2 MPEG Access Units

For carriage of compressed audio-visual data MPEG defines Access Units. An MPEG Access Unit (AU) is the smallest data entity to which

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

timing information can be attributed. In case of audio an Access Unit represents an audio frame and in case of video a picture. MPEG Access Units are by definition byte aligned. If for example an audio frame is not byte aligned, up to 7 zero-padding bits MUST be inserted at the end of the frame to achieve a byte-aligned Access Unit.
Decoders MUST be able to decode AUs in which such padding is applied.

Consistent with the MPEG-4 specification, this document requires that each MPEG-4 video Access Unit includes all the coded data of a picture, any video stream headers that may precede the coded picture data, and any video stream stuffing that may follow it, up to, but not including the startcode indicating the start of a new video stream or the next Access Unit.

2.3 Concatenation of Access Units

Frequently it is possible to carry multiple Access Units in one RTP packet. This is particularly useful for audio; for example, when AAC is used for encoding of a stereo signal at 64 kbits/sec, AAC frames contain on average approximately 200 bytes. On a LAN with a 1500 byte MTU this would allow on average 7 complete AAC frames to be carried per AAC packet.

Access Units may have a fixed size in bytes, but a variable size is also possible. To facilitate parsing in case of multiple concatenated AUs in one RTP packet, the size of each AU is made known to the receiver. When concatenating in case of a constant AU size, this size is communicated through a format parameter. When concatenating in case of variable size AUs, the RTP payload carries an AU size field for each contained AU. In combination with the RTP payload length the size information allows the RTP payload to be split by the receiver back into the individual AUs.

To simplify the implementation of [6] defined in this document, it is required that when multiple AUs are carried in an RTP packet, that each AU MUST be complete, i.e. the number of AUs in an RTP packet MUST be integral.

2.4 Fragmentation of Access Units

MPEG allows for very large Access Units. Since most IP networks have significantly smaller MTU's, this payload format allows to fragment the AUs over multiple RTP packets so as to avoid IP layer fragmentation. To simplify the implementation of [6] defined in this document, an RTP packet SHALL either carry one or more complete Access Units or a single fragment of one Access Unit.

2.5 Interleaving

When an RTP packet carries a contiguous sequence of Access Units, the loss of such packet can result in "decoding gaps" for the user. One method to alleviate this problem is to allow for the Access Units to be interleaved in the RTP packets. For a modest cost in latency and implementation complexity, significant error resiliency to packet loss can be achieved.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

To support optional interleaving of Access Units, this payload format allows for index information to be sent for each Access Unit. The RTP sender is free to choose the interleaving pattern without propagating this information to the receiver(s). Indeed the sender could dynamically adjust the interleaving pattern based on the Access Unit size, error rates, etc. The RTP receiver does not need to know the interleaving pattern used, it only need extract the index information of the Access Unit and insert the Access Unit into the appropriate sequence in the rendering queue. An example of interleaving is given below.

Assume that an RTP packet contains 3 AUs, and that the AUs are numbered 1, 2, 3, 4, etc. If an interleaving group length of 9 is chosen, then RTP packet(i) contain the following AU(n):

```
RTP packet(1):  AU(1),  AU(4),  AU(7)
RTP packet(2):  AU(2),  AU(5),  AU(8)
RTP packet(3):  AU(3),  AU(6),  AU(9)
RTP packet(4):  AU(10), AU(13), AU(16)
RTP packet(5):  AU(11), AU(14), AU(17)
Etc.
```

2.6 Time stamp information

MPEG-4 defines two type of time stamps, the decoding time stamp DTS and the composition time stamp CTS. The RTP timestamp is equivalent to the composition time stamp.

The RTP time stamp MUST carry the sampling instance of the first AU (fragment) in the RTP packet. When multiple AUs are carried within an RTP packet, the time stamps of subsequent AUs can be calculated if the frame period of each AU is known. For audio and video this is possible if the frame rate is constant. However, in some cases it is not possible to make such calculation, for example for variable frame rate video and for MPEG-4 BIFS streams carrying composition information. To support such cases, this payload format can be configured to carry a CTS in the RTP payload for each contained Access Unit. A CTS time stamp MAY be conveyed in the RTP payload only for non-first AUs in the RTP packet, and SHALL NOT be conveyed for the first AU (fragment), as the time stamp for the latter is carried by the RTP time stamp.

The DTS timestamp is applied only in MPEG video streams that use bi-directional coding, i.e. when pictures may be predicted in both forward and backward direction by using either a reference picture in the past, or a reference picture in the future. The DTS cannot be carried in the RTP header. In some cases the DTS can be derived from the RTP time stamp using frame rate information; this requires deep parsing in the video stream, which may be considered objectionable. But if the video frame rate is variable, the required information is not even present in the video stream. For both reasons, the capability has been defined to optionally carry a DTS in the RTP payload for each contained Access Unit.

Since RTP time stamps may be re-stamped by RTP devices, each CTS and DTS contained in the RTP payload is coded differentially from the RTP time stamp, so as to avoid extensive parsing by re-stamping devices.

2.7 Carriage of auxiliary information.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

This payload format defines a specific field to carry auxiliary data on the contained MPEG-4 stream, representing MPEG-4 system information. The auxiliary data corresponds to the RSLH field defined in [6]. Receivers MAY use the auxiliary data to decode the contained stream, but receivers that have no interest in such data MAY skip the auxiliary data field. To facilitate skipping of the data, and to avoid the need for parsing it, the auxiliary data field is preceded by a field that specifies the length of the auxiliary data.

2.8 Format parameters and the conditional presence and length of fields

To support the features described in the previous sections several fields are defined for carriage in the RTP payload. However, their use strongly depends on the type of MPEG-4 elementary stream that is carried. Sometimes a specific field is needed with a certain length, while in other cases such field is not needed at all. To be efficient in either case, the fields needed for these features are configurable by means of format parameters. In general, a format parameter defines the presence and length of associated fields. A length of zero indicates absence of the field. As a consequence, parsing of the payload requires knowledge of format parameters. The format parameters are conveyed to the receiver via SDP messages or through other means.

2.9 Global structure of payload format

The payload structure in [6] is described in terms derived from the SL layer. In this document exactly the same structure is described in more general terms, so as to improve the readability for people with no knowledge of the SL layer. So the payload structure described below corresponds on bit level exactly to the payload structure defined in [6].

The RTP payload following the RTP header, contains three byte aligned data sections, of which the first two MAY be empty. See figure 1.

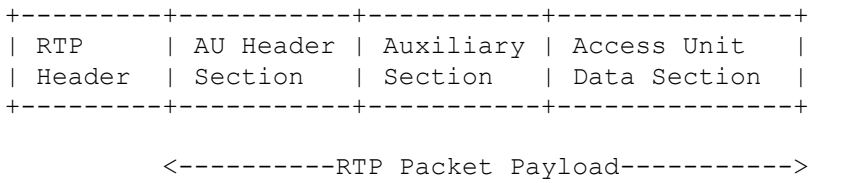


Figure 1: Data sections within an RTP packet

The first data section is the AU (Access Unit) Header Section, that contains one or more AU-headers; however, each AU-header MAY be empty, in which case the entire AU Header Section is empty. The second section is the Auxiliary Section, containing auxiliary data; also this section MAY be configured empty. The third section is the Access Unit Data Section, containing either a single fragment of one Access Unit or one or more complete Access Units. The Access Unit Data Section is never empty.

When compared to the terms used in [6], the AU Header Section exactly corresponds to the MSLHSection, the Auxiliary Section to the RSLHSection, and the Access Unit Data Section to the SLPPSection.

2.10 Modes to transport MPEG-4 streams

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

While it is possible to build fully configurable receivers capable of receiving any MPEG-4 stream, this specification also allows for the design of simplified, but dedicated receivers, that are capable for example to receive only one type of MPEG-4 stream. This is achieved by requiring that specific modes be defined for using this specification. Each mode defines how to transport specific MPEG-4 streams, for example by defining suitable constraints or payload configurations. Modes can be defined as deemed appropriate. However, each mode MUST be in full compliance with this specification.

The applied mode MUST be signalled. Signalling the mode is particularly important for receivers that are only capable of decoding a particular mode. Such receivers need to determine whether that particular mode is applied, so as to avoid problems with processing of payloads that are beyond the capabilities of the receiver.

In this internet draft only modes are defined for transport of MPEG-4 CELP and AAC streams. However, in future new RFCs are expected to specify additional modes of using this specification for transport of other MPEG-4 streams.

2.11 Alignment with "RFC-generic" and RFC 3016

This document defines a subset of the "RTP payload format for MPEG-4 streams" [6]. The main characteristic of this subset is that each RTP payload is only allowed to contain either a single fragment of one Access Unit or one or more complete Access Units. Obviously, RTP payloads that apply this subset in conformance with this document conform also to [6]. Receivers that comply with [6] are able to decode MPEG-4 streams carried in compliance with this document.

Receivers designed to only comply to this document may not be able to decode a RTP payload that conforms to [6] but not to this document. Such receivers may also not be capable of exploiting some of features of the SL layer supported in [6], such as knowledge of AU-start, random access information and other information carried in the SL header, but not described in this document.

Furthermore, this payload can be configured to be identical to the payload format defined in RFC 3016 for the MPEG-4 video configurations recommended in RFC 3016. Hence, receivers that comply with RFC 3016 can decode such RTP payload. Vice versa, receivers that comply with the specification in this document SHOULD be able to decode payloads, names and parameters defined for MPEG-4 video in RFC 3016.

For interoperability reasons, applications that transport MPEG-4 video over RTP SHOULD use the payload format and associated names and parameters defined in RFC 3016 if the functionality provided by RFC 3016 can meet the requirements of that application.

3 Payload Format

3.1 RTP Header Fields Usage

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Payload Type (PT): The assignment of an RTP payload type for this RTP packet format is outside the scope of this document, and will not be specified here. It is expected that the RTP profile for a particular class of applications will assign a payload type for this encoding, or if that is not done, then a payload type in the dynamic range shall be chosen.

Marker (M) bit: The M bit is set to 1 to indicate that the RTP packet payload includes the end of each Access Unit of which data is contained in this RTP packet. As the payload either carries one or more complete Access Units or a single fragment of an Access Unit, the M is always set to set to 1, except when the packet carries a single fragment of an Access Unit that is not the last one.

Extension (X) bit: Defined by the RTP profile used.

Sequence Number: The RTP sequence number SHOULD be generated by the sender with a constant random offset.

Timestamp: Indicates the sampling instance of the first AU contained in the RTP payload. This sampling instance is equivalent to the CTS in the MPEG-4 time domain. The clock rate of the RTP time stamp MAY be expressed as part of the RTPMAP. If an audio or video stream with a fixed frame rate is transported, the rate SHOULD be set to the same value as the sampling frequency of the audio or video frames (number of samples per second).

In all cases, the sender SHALL make sure that RTP time stamps are identical only if the RTP time stamp refers to fragments of the same Access Unit.

According to RFC 1889 [2] (section 5.1), RTP timestamps are recommended to start at a random value for security reasons. However,then a receiver is, in the general case, not able to reconstruct the original MPEG Time Stamps, which creates problems for applications where streams from multiple sources are to be synchronized. Therefore the usage of a random offset SHOULD be avoided.

SSRC: set as described in RFC1889 [2].

CC and CSRC fields are used as described in RFC 1889 [2].

RTCP SHOULD be used as defined in RFC 1889 [2].

3.2 RTP Payload Structure

As already noted in section 2.9 of this document, this document uses more general names to describe exactly the same payload structure as defined in [6]. For mapping between section names in [6] and in this document see section 2.9.

3.2.1 The AU Header Section

When present, the AU Header Section consists of the AU-header-length field, followed by a number of AU-headers. See figure 2.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

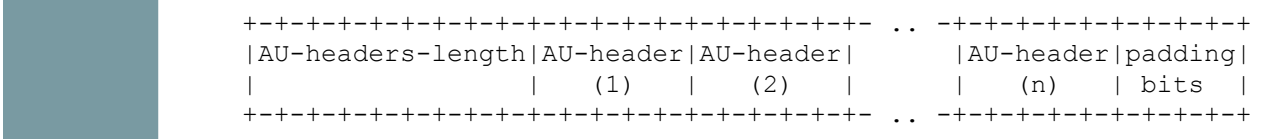


Figure 2: The AU Header Section

The AU-headers are configured using format parameters and MAY be empty. If the AU-header is configured empty, the AU-headers-length field SHALL not be present and consequently the AU Header Section is empty. If the AU-header is not configured empty, then the AU-headers-length is a two byte field that specifies the length in bits of the immediately following AU-headers.

Each AU-header is associated with a single Access Unit (fragment) contained in the Access Unit Data Section in the same RTP packet. For each contained Access Unit (fragment) there is exactly one AU-header. Within the AU Header Section, the AU-headers are bit-wise concatenated in the order in which the Access Units are contained in the Access Unit Data Section. Hence, the n-th AU-header refers to the n-th AU(fragment). If the concatenated AU-headers consume a non-integer number of bytes, up to 7 zero-padding bits MUST be inserted at the end in order to achieve byte-alignment of the AU Header Section.

3.2.1.1 The AU-header

The AU-header contains the fields given in figure 3. The length in bits of the above fields with the exception of the CTS-flag and the DTS-flag fields is defined by format parameters; see section 4.1. If a format parameter has the default value of zero, then the associated field is not present.

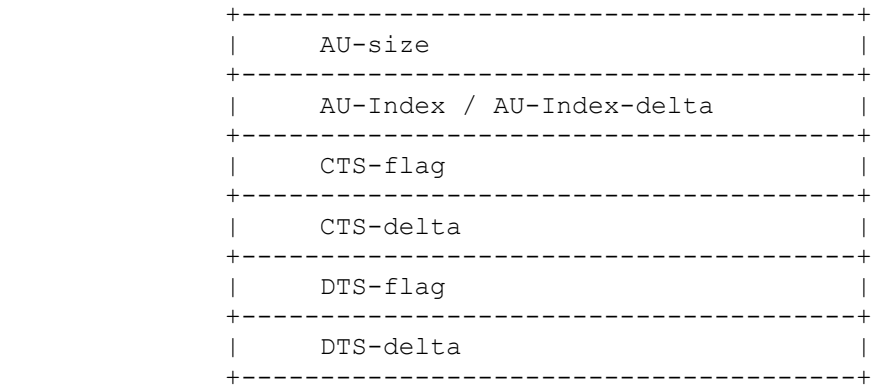


Figure 3: The fields in the AU-header. If used, the AU-Index field only occurs in the first AU-header within an AU Header Section; in any other AU-header the AU-Index-delta field occurs instead.

AU-size: indicates the size in bytes of the associated Access Unit in the Access Unit Data Section in the same RTP packet. When the AU-size is associated to an AU fragment, the AU size indicates the size of the

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

entire AU and not the size of the fragment. This can be exploited to determine whether a packet contains an entire AU or a fragment, which is particularly useful after losing a packet carrying the last fragment of an AU.

AU-Index: indicates the serial number of the associated Access Unit (fragment). For each (in time) consecutive AU or AU fragment, the serial number is incremented with 1. When present, the AU-Index field occurs in the first AU-header in the AU Header Section, but MUST not occur in any subsequent (non-first) AU-header in that Section. To encode the serial number in any such non-first AU-header, the AU-Index-delta field is used. When each AU-Index field is coded with the value 0, the serial number of the AU (fragment) is not specified and in that case receivers MAY ignore the AU-Index field.

AU-Index-delta: The AU-Index-delta field is an unsigned integer that specifies the serial number of the associated AU as the difference with respect to the serial number of the previous Access Unit. Hence, for the n-th (n>1) AU the serial number is found from:

AU-Index(n) = AU-Index(n-1) + AU-Index-delta(n) + 1

If the AU-Index field is present in the first AU-header in the AU Header Section, then the AU-Index-delta field MUST be present in any subsequent (non-first) AU-header. When the AU-Index-delta is coded with the value 0, it indicates that the Access Units are consecutive in time. An AU-Index-delta value larger than 0 signals that interleaving is applied.

CTS-flag: Indicates whether the CTS-delta field is present.
A value of 1 indicates that the field is present, a value of 0 that it is not present. The CTS-flag field MUST be present in each AU-header if the length of the CTS-delta field is signaled to be larger than zero. In that case, the CTS-flag field MUST have the value 0 in the first AU-header and MAY have the value 1 in all non-first AU-headers. The CTS-flag field SHOULD be 0 for any non-first fragment of an Access Unit.

CTS-delta: Encodes the CTS by specifying the value of CTS as a 2’s complement offset (delta) from the timestamp in the RTP header of this RTP packet. The CTS MUST use the same clock rate as the time stamp in the RTP header.

DTS-flag: Indicates whether the DTS-delta field is present. A value of 1 indicates that DTS-delta is present, a value of 0 that it is not present. The DTS-flag field MUST be present in each AU-header if the length of the DTS-delta field is signalled to be larger than zero. The DTS-flag field SHOULD be 0 for any non-first fragment of an Access Unit.

DTS-delta: specifies the value of the DTS as a 2’s complement offset (delta) from the CTS timestamp. The DTS MUST use the same clock rate as the time stamp in the RTP header.

If present, the fields MUST occur in the mutual order given in figure 3. In the general case a receiver can only discover the size of an AU-header by parsing it since the presence of the CTS-delta and DTS-delta fields is signalled by the value of the CTS-flag and DTS-flag, respectively.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

3.2.2 The Auxiliary Section

The Auxiliary Section consists of the auxiliary-data-size field followed by the auxiliary-data field. Receivers MAY (but are not required to) parse the auxiliary-data field; to facilitate skipping of the auxiliary-data field by receivers, the auxiliary-data-size field indicates the length in bits of the auxiliary-data. If the concatenation of the auxiliary-data-size and the auxiliary-data fields consume a non-integer number of bytes, up to 7 zero padding bits MUST be inserted immediately after the auxiliary data in order to achieve byte-alignment. See figure 4.

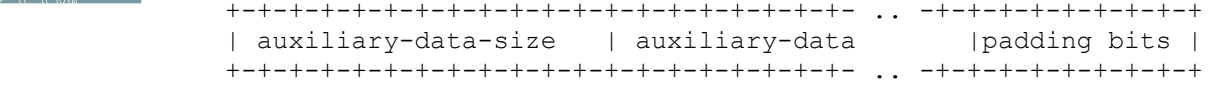


Figure 4: The fields in the Auxiliary Section

The length in bits of the auxiliary-data-size field is configurable by a format parameter; see section 4.1. The default length of zero indicates that the entire Auxiliary Section is absent.

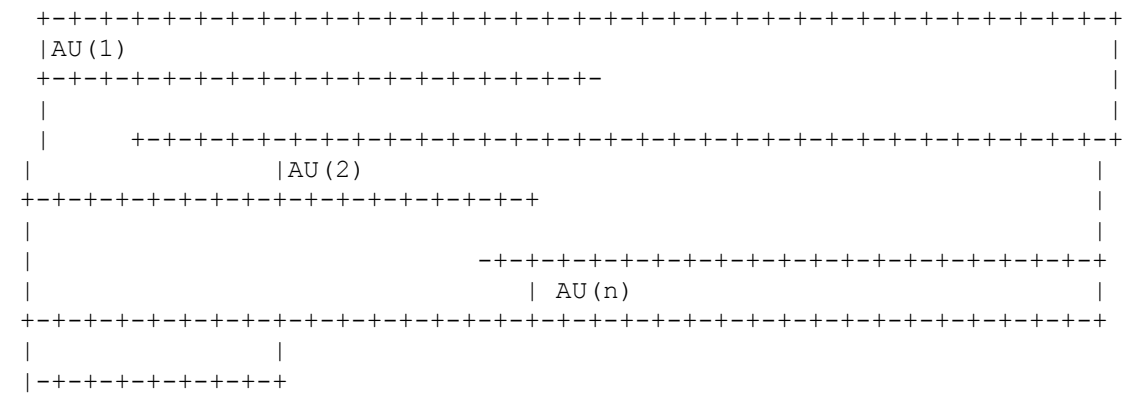
auxiliary-data-size; specifies the length in bits of the immediately following auxiliary-data field;

auxiliary-data; the auxiliary-data field contains the Remaining SL headers (RSLHs) as defined in [6].

3.2.3 The Access Unit Data Section

The Access Unit Data Section contains an integer number of complete Access Units or a single fragment of one AU. The Access Unit Data Section is never empty. If data of more than one Access Units is contained, then the AUs are concatenated into a contiguous string of bytes. See figure 5. The AUs inside the Access Unit Data Section MUST be in decoding order.

The size and number of Access Units SHOULD be adjusted such that the resulting RTP packet is not larger than the path-MTU. To handle larger packets, this payload format relies on lower layers for fragmentation, which may not be desirable.



Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Figure 5: Access Unit Data Section; each AU is byte aligned.

When multiple Access Units are carried, the size of each AU MUST be made available to the receiver. If the AU size is variable then the size of each AU MUST be indicated in the AU-size field of the corresponding AU-header. However, if the AU size is constant for a stream, this mechanism SHOULD NOT be used, but instead the fixed size SHOULD be signalled by the format parameter "ConstantSize", see section 4.1.

The absence of both AU-size in the AU-header and the ConstantSize format parameter indicates carriage of a single AU (fragment), i.e. that a single Access Unit (fragment) is transported in each RTP packet for that stream.

3.2.3.1 Fragmentation

A packet SHALL carry either one or more Access Units, or a single fragment of an Access Unit. Fragments of the same Access Unit have the same time stamp but differing RTP sequence numbers. The marker bit in the RTP header is 1 on the last fragment of an Access Unit, and 0 on all other fragments.

3.2.3.2 Interleaving

Access Units MAY be interleaved. Senders MAY perform interleaving. Receivers MUST support interleaving.

When interleaving of Access Units is used it SHALL be implemented using the AU-Index and AU-Index-delta fields in the AU-header.

Based on the RTP sequence number, the RTP time stamp, the AU-Index and the AU-Index-delta, a receiver can unambiguously reconstruct the original order even in case of out-of-order packets, packet loss or duplication. Note that for this purpose the AU-Index is redundant when the RTP time stamp and the AU-Index-delta values are sufficient for placing the AUs correctly in time. In such cases receivers MAY ignore the AU-Index value and senders MAY code the AU-Index field with the value 0, but only if they code each AU-Index field with that value. When interleaving is applied, a de-interleave buffer is needed in receivers to put the Access Units in their correct logical consecutive order in time. This requires the computation of the time stamp for each Access Unit. In case of a fixed time duration per Access Unit, the time-stamp of each access unit i in an RTP packet with RTP time-stamp This calculated as follows:

```
Timestamp[0] = T
Timestamp[i, i > 0] = T +(Sum(for k=1 to i of (AU-Index-delta[k]
                      + 1))) * access-unit-duration
```

When AU-Index-delta is always 0, this reduces to T + I * (access-unit-duration). This is the non-interleaved case, the frames are consecutive in time. Note that the AU-Index field (present for the first Access Unit) is not needed in this calculation. Hence in cases where the Access-unit-duration has a fixed and known value, the AU-Index does not need to provide index information and can be coded with the value 0. See also the semantics of the AU-Index field in 3.2.1.1.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

When an RTP packet arrives (after any re-ordering has been done), receivers may 'flush' all Access Units from the interleave buffer which have a time-stamp strictly less than the time-stamp of the arriving packet. Similarly the first Access Unit of every arriving packet can always be flushed (as no following packet can provide an earlier Access Unit), and any Access Units which are consecutive with it which have already been received. Access Units should also be flushed in time to be played; this can be important if there is loss before end-of-stream, before a silence interval, or before a large drop-out.

3.2.3.2.1 Constraints for interleaving

The size of the packets should be suitably chosen to be appropriate to both the path MTU and the duration and capacity of the receiver's de-interleave buffer. The maximum packet size for a session should be chosen not to exceed the path MTU.

In order to control receiver latency and mitigate the effects of loss, there are profile-based limits on the size of the packet. This is expressed as a duration: it is calculated from the duration of the Access Units contained within a packet. It is NOT the difference in time-stamp between the first and last Access Unit in a packet.

No matter what interleaving scheme is used, the scheme must be analyzed to calculate the minimum number of frames a receiver has to buffer in order to de-interleave.

The maximum packet duration in milliseconds, and the maximum de-interleave buffer required at the receiver, for the two profiles, shall not exceed:

RTP transport profile 0 -- 200 milliseconds
RTP transport profile 1 -- 500 milliseconds

When interleaving is applied, the applied RTP transport profile MUST be signalled by the profile parameter; see section 4.1.

Note that for low bit-rate material, the duration limit may make packets shorter than the MTU size.

3.3 Usage of this specification

3.3.1 General

Usage of this specification requires definition of a mode. A mode defines how use this specification for transport of one or more types of MPEG-4 streams. Each mode may specify constraints and payload configurations as deemed appropriate.

Senders MUST signal the mode that they use by the format parameter Mode. In this document only modes are defined for transport of MPEG-4 CELP and AAC streams, but more modes are expected to be defined in future RFCs.

3.3.2 Modes for MPEG-4 CELP and AAC streams

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Four modes are defined for transport of MPEG-4 CELP and AAC streams. In each of these modes, the same requirements apply for the rtpmap attributes. The general form of an rtpmap attribute is:

a=rtpmap:<payload type><encoding name>/<clock rate>[/<encoding parameters>]

For audio streams, <encoding parameters> specifies the number of audio channels. This parameter may be omitted if the number of channels is one, provided no additional parameters are needed.

In all four modes, the following attributes are REQUIRED:

a) The encoding name

b) The RTP clock rate MUST be expressed. It is recommended that this be the sampling rate of the audio, to give sample-accurate timing. However, other rates MAY be used (e.g. 90 kHz).

c) The number of audio channels MUST be specified, for example as 2 for stereo material (see RFC 2327) and MAY be specified as 1 for mono material; 1 is the default.

3.3.3 Constant bit-rate CELP.

This mode is signalled by mode=CELP-cbr. In this mode one or more fixed size CELP frames can be transported in one RTP packet; there is no support for interleaving. The RTP payload consist of one or more concatenated CELP frames, each of the same size. Both the AU Header Section and the Auxiliary Section are empty.

The format parameter ConstantSize MUST be provided to specify the length of each CELP frame.

For an example see below.

```
m=audio 49230 RTP/AVP 96
a=rtpmap:96 mpeg-generic/44100/2
a=fmtp:96 streamtype=5; profile-level-id=15; mode=CELP-cbr; config=
AudioSpecificConfig(); ConstantSize=xxx;
```

The AudioSpecificConfig() specifies that the audio stream type is CELP.

3.3.4 Variable bit-rate CELP

This mode is signalled by mode=CELP-vbr. With this mode in one RTP packet one or more variable size CELP frames can be transported with optional interleaving. As the largest possible frame size in this mode is greater than the maximum CELP frames size, there is no support for fragmentation on the CELP frames.

In this mode the RTP payload consists of the AU Header Section,
followed by one or more concatenated CELP frames. The Auxiliary Section is empty. For each CELP frame contained in the payload there is a one byte AU-header in the AU Header Section to provide :

(a) the size of each CELP frame in the payload and

(b) index information for computing the sequence (and hence timing) of each CELP frame.

Transport of CELP frames requires that the AU-size field is coded with 6 bits. In this mode therefore 6 bits are allocated to the AU-size field,

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

and 2 bits to the AU-Index(-delta) field. Each AU-Index field MUST be coded with the value 0. In the AU Header Section, the concatenated AU-headers are preceded by the 16-bit AU-headers-length field, as specified in 3.2.1.

Next to the required format parameters, the following parameters MUST be present:
SizeLength, IndexLength, and IndexDeltaLength.
When interleaving is applied (AU-Index-delta coded with a value larger than 0), also the parameter Profile MUST be present.

Example :

```
m=audio 49230 RTP/AVP 96
a=rtpmap:96 mpeg4-generic/44100/2
a=fmtp:96 streamtype=5; profile-level-id=15; mode=CELP-vbr; config=
AudioSpecificConfig(); SizeLength=6; IndexLength=2;
IndexDeltaLength=2;
Profile=1
```

The AudioSpecificConfig() specifies that audio stream type is CELP.

3.3.5 Low bit-rate AAC

This mode is signalled by AAC-lbr. This mode supports transport of one or more variable size AAC frames with optional support for interleaving and fragmenting. The maximum size of an AAC frame(fragment) in this mode is 63 bytes.

The payload configuration in this mode is the same as in the variable bit-rate CELP mode as defined in 3.3.4. The RTP payload consists of the AU Header Section, followed by concatenated AAC frames. The Auxiliary Section is empty. For each AAC frame contained in the payload the one byte AU-header provides :
(a) the size of each AAC frame in the payload and
(b) index information for computing the sequence (and hence timing) of each AAC frame.
In the AU-header, the AU-size is coded with 6 and the AU-Index(-delta)with 2 bits; the AU-Index field MUST have the value 0 in each AU header.
In the AU-header Section, the concatenated AU-headers are preceded by the 16-bit AU-headers-length field, as specified in 3.2.1.

Next to the required format parameters, the following parameters MUSTbe present:
SizeLength, IndexLength, and IndexDeltaLength.
When interleaving is applied (AU-Index-delta coded with a value larger than 0), also the parameter Profile MUST be present.

Example :

```
m=audio 49230 RTP/AVP 96
a=rtpmap:96 mpeg4-generic/44100/2
a=fmtp:96 streamtype=5; profile-level-id=15; mode=AAC-lbr; config=
AudioSpecificConfig(); SizeLength=6; IndexLength=2;
IndexDeltaLength=2; Profile=1
```


Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

The AudioSpecificConfig() specifies that audio stream type is AAC.

3.3.6 High bit-rate AAC

This mode is signalled by mode=AAC-hbr. This mode supports transport of one or more large variable size AAC frames in one RTP packet with optional support for interleaving and fragmenting. The maximum size of AAC frame (fragment) in this mode is 8191 bytes.

In this mode the RTP payload consists of the AU Header Section, followed by one or more concatenated AAC frames. The Auxiliary Section is empty. For each AAC frame contained in the payload there is an AU-header in the AU Header Section to provide :

(a) the size of each AAC frame in the payload and

(b) index information for computing the sequence (and hence timing) of each AAC frame.

To code the maximum size of an AAC frame requires 13 bits. Therefore in this configuration 13 bits are allocated to the AU-size, and 3 bits to the AU-Index(-delta) field. Thus each AU-header has a size of 2 bytes. Each AU-Index field MUST be coded with the value 0. In the AU Header Section, the concatenated AU-headers are preceded by the 16-bit AU-headers-length field, as specified in 3.2.1.

Next to the required format parameters, the following parameters MUST be present:

SizeLength, IndexLength, and IndexDeltaLength.

When interleaving is applied (AU-Index-delta coded with a value larger than 0), also the parameter Profile MUST be present.

Example :

```
m=audio 49230 RTP/AVP 96
a=rtpmap:96 mpeg4-generic/44100/2
a=fmtp:96 streamtype=5; profile-level-id=15; mode=AAC-hbr; config=
AudioSpecificConfig(); SizeLength=13;
IndexLength=3;IndexDeltaLength=3;Profile=1
```

The AudioSpecificConfig() specifies that the audio stream type is AAC.

4. Types and names

This section describes the MIME types and names associated with this payload format.

Depending on the required payload configuration, format parameters May need to be available to the receiver. This is done using the parameters described in the next section. The absence of any of these parameters is equivalent to the associated field set to its default value, which is always zero. The absence of any such parameters resolves into a default "basic" configuration.

In the MPEG-4 framework the SL stream configuration information is carried using the Object Descriptor. When such information is present both in an Object Descriptor and as a parameter of this payload format it MUST be exactly the same.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

4.1 MIME types

This specification uses exactly the same MIME types as [6], and hence no further MIME type registration is required. In [6] uses the MIME media type names: "video" or "audio" or "application".

"video" SHOULD be used for any MPEG Video stream or any MPEG-4 System (ISO/IEC 14496-1) stream that conveys information needed for an audio-visual presentation.
"audio" SHOULD be used for any MPEG Audio streams and any MPEG-4 System (ISO/IEC 14496-1) stream that conveys information needed for an audio-only presentation.

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

MIME subtype name: mpeg4-generic

Required parameters:

StreamType:
The integer value that indicates the type of MPEG-4 stream that is carried; its coding corresponds to the values of the streamType as defined for the DecoderConfigDescriptor in ISO/IEC 14496-1.

Profile-level-id:
A decimal representation of the MPEG-4 Profile Level indication. This parameter MUST be used in the capability exchange or session set-up procedure to indicate the MPEG-4 Profile and Level combination of which the relevant MPEG-4 media codec is capable of.

For audio streams, this parameter is the decimal value from Table 5(audioProfileLevelIndicationValues) in ISO/IEC 14496-1, indicating which MPEG-4 Audio tool subsets are applied to encode the audio stream.

For visual streams, this parameter is the decimal value from Table G-1 (FLC table for profile and level indication of ISO/IEC 14496-2,indicating which MPEG-4 Visual tool subsets are applied to encode the visual stream.

Config:
A hexadecimal representation of an octet string that expresses the media payload configuration. Configuration data is mapped onto the octet string in an MSB-first basis. The first bit of the configuration data SHALL be located at the MSB of the first octet. In the last octet, if necessary to achieve byte alignment, up to 7 zero-valued padding bits shall follow the configuration data.
For audio streams, config is the audio object type specific decoder configuration data AudioSpecificConfig() as defined in ISO/IEC 14496-3.

For visual streams, config is the MPEG-4 Visual configuration information, as defined in subclause 6.2.1 Start codes of ISO/IEC14496-2. The configuration information indicated by this parameter SHALL be the same as the configuration information in the corresponding MPEG-4 Visual

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

stream, except for first-half-vbv-occupancy and latter-half-vbv-occupancy, if it exists, which may vary in the repeated configuration information inside an MPEG-4 Visual stream (See 6.2.1 Start codes of ISO/IEC14496-2).

Mode:
The mode in which this specification is used. The following modes can be signalled :
mode=CELP-cbr,
mode=CELP-vbr,
mode=AAC-lbr and
mode=AAC-hbr.
Other modes are expected to be defined in future RFCs. When defining a new mode care MUST be taken that an implementation of all features of this specification can decode the payload format corresponding to this new mode. For this reason a mode MUST NOT specify new default values for MIME parameters; in particular, MIME parameters MUST be present (unless they have the default value), even if it is redundant in case the mode assigns fixed values. A mode may define additionally that some MIME parameters are required instead of optional, that some MIME parameters have fixed values (or ranges), and that there are rules restricting the usage.

Optional parameters:

ConstantSize:
The constant size in bytes of each Access Unit for this stream. Simultaneous presence of ConstantSize and the SizeLength parameters is not permitted.

SizeLength:
The number of bits on which the AU-size field is encoded in the AU-header. Simultaneous presence of SizeLength and the ConstantSize parameter is not permitted.

IndexLength:
The number of bits on which the AU-Index is encoded in the first AU-header. The default value of zero indicates the absence of the AU-Index and AU-Index-delta fields in each AU-header.

IndexDeltaLength:
The number of bits on which the AU-Index-delta field is encoded in any non-first AU-header.

CTSDeltaLength:
The number of bits on which the CTS-delta field is encoded in the AU-header.

DTSDeltaLength:
The number of bits on which the DTS-delta field is encoded in the AU-header.

AuxiliaryDataSizeLength:
The number of bits that is used to encode the auxiliary-data-size field.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Profile:
The decimal representation of the RTP transport profile.

Applications MAY use more parameters, in addition to those defined above. Receivers MUST tolerate the presence of such additional parameters, but these parameters SHALL not impact the decoding of receivers that comply to this specification.

Encoding considerations:
System bitstreams MUST be generated according to MPEG-4 System specifications (ISO/IEC 14496-1). Video bitstreams MUST be generated according to MPEG-4 Visual specifications (ISO/IEC 14496-2). Audio bitstreams MUST be generated according to MPEG-4 Visual specifications ISO/IEC 14496-3). The RTP packets MUST be packetized according to the RTP payload format defined in RFC <self-reference-to-this>.

Security considerations:
As in RFC <self-reference-to-this>.

Interoperability considerations:
MPEG-4 provides a large and rich set of tools for the coding of visual objects. For effective implementation of the standard, ubsets of the MPEG-4 tool sets have been provided for use in specific applications. These subsets, called 'Profiles', limit the size of the tool set a decoder is required to implement. In order to restrict computational complexity, one or more 'Levels' are set for each Profile. A Profile@Level combination allows:
. a codec builder to implement only the subset of the standard he needs, while maintaining interworking with other MPEG-4 devices included in the same combination, and
. checking whether MPEG-4 devices comply with the standard ('conformance testing').
A stream SHALL be compliant with the MPEG-4 Profile@Level specified by the parameter "profile-level-id". Interoperability between a sender and a receiver may be achieved by specifying the parameter "profile-level-id" in MIME content, or by arranging in the capability exchange/announcement procedure to set this parameter mutually to the same value.

Published specification:
The specifications for MPEG-4 streams are presented in ISO/IEC 14469-1, 14469-2, and 14469-3. The RTP payload format is described in RFC <self-reference-to-this>.

Applications which use this media type:
Multimedia streaming and conferencing tools, Internet messaging and Email applications.

Additional information: none

Magic number(s): none

File extension(s):
None. A file format with the extension .mp4 has been defined for MPEG-4 content but is not directly correlated with this MIME type which sole purpose is RTP transport.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Macintosh File Type Code(s): none

Person & email address to contact for further information:
Authors of RFC <self-reference-to-this>.

Intended usage: COMMON

Author/Change controller:
Authors of RFC <self-reference-to-this>.

4.2 Concatenation of parameters

Multiple parameters SHOULD be expressed as a MIME media type string, in the form of a semicolon-separated list of parameter=value pairs (for parameter usage examples see Appendix A).

4.3 Usage of SDP

4.3.1 The a=fmtp keyword

It is assumed that one typical way to transport the above-described parameters associated with this payload format is via a SDP message for example transported to the client in reply to a RTSP DESCRIBE of via SAP. In that case the (a=fmtp) keyword MUST be used as described in RFC 2327 [8, section 6]. The syntax being then:

a=fmtp:<format> <parameter name>=<value>[; <parameter name>=<value>]

5. Security Considerations

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification [2]. 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 (i.e. excluding media data processing) 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.

MPEG-4 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.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Authentication mechanisms can be used to validate of the sender and the data to prevent security problems due to non-compliant malignant MPEG-4 streams.

A security model is defined in MPEG-4 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.

6. References

[1] ISO/IEC International Standard 14496 (MPEG-4); "Information technology - Coding of audio-visual objects", January 2000

[2] Schulzrinne, Casner, Frederick, Jacobson RTP: A Transport Protocol for Real Time Applications RFC 1889, Internet Engineering Task Force, January 1996.

[3] S. Bradner, Key words for use in RFCs to Indicate Requirement Levels, RFC 2119, March 1997.

[4] D. Hoffman, G. Fernando, V. Goyal, M. Civanlar, RTP payload format for MPEG1/MPEG2 Video, RFC 2250, January 1998.

[5] Y. Kikuchi, T. Nomura, S. Fukunaga, Y. Matsui, H. Kimata, RTP payload format for MPEG-4 Audio/Visual streams, RFC 3016.

[6] Avaro, Basso, Casner, Civanlar, Gentric, Herpel, Lim, Perkins, van der Meer, RTP payload format for MPEG-4 streams, work in progress, draft-gentric-avt-mpeg4-multiSL-01.txt, January 2001.

[7] D. Singer, Y Lim, A Framework for the delivery of MPEG-4 over IP-based Protocols, work in progress, draft-singer-mpeg4-ip-01.txt,October 2000.

[8] Handley, Jacobson, SDP: Session Description Protocol, RFC 2327, Internet Engineering Task Force, April 1998.

7. Author Adresses

Jan van der Meer
Philips Digital Networks
Cederlaan 4
5600 JB Eindhoven
Netherlands

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

Email : jan.vandermeer@philips.com

David Mackie
Cisco Systems Inc.
170 West Tasman Dr.
San Jose, CA 95034
Email: dmackie@cisco.com

Viswanathan Swaminathan
Sun Microsystems Inc.
901 San Antonio Road, M/S UMPK15-214
Palo Alto, CA 94303
Email: viswanathan.swaminathan@sun.com

David Singer
Apple Computer, Inc.
One Infinite Loop, MS:302-3MT
Cupertino CA 95014
Email: singer@apple.com

Full Copyright Statement

"Copyright (C) The Internet Society (date). All Rights Reserved. This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process MUST be followed, or as required to translate it into.

APPENDIX: Usage of this payload format

Appendix A. Examples

A.1 Examples of delay analysis with interleave

A.1.1 Group interleave

An example of regular interleave is when packets are formed into groups. If the number of packets in a group is N, packet 0 contains frame 0, frame N, frame 2N, and so on; packet 1 contains frame 1, frame 1+N, 1+2N, and so on. The AU-Index field is used to document the sequence of the packet within the group (or the first frame in the packet, which is the same thing in this scheme), and all the AU-Index-delta fields contain N-1.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

Receivers can tell when a new interleave group is starting, by noting that the computed time-stamp of the first frame in a packet is later than any previously computed time-stamp. This is because no following packet can contain an earlier RTP timestamp (RTP rules),and the second and subsequent frames in a packet have larger time-stamps(the frames in a packet are also in time-order).

If the group size is 3, then packets are formed as follows:

Packet	Time-stamp	Frame Numbers	AU-Index, AU-Index-delta
0	T[0]	0, 3, 6	0, 2, 2
1	T[1]	1, 4, 7	0, 2, 2
2	T[2]	2, 5, 8	0, 2, 2
3	T[9]	9,12,15	0, 2, 2

In this case, the receiver would have to buffer 4 frames at least from packets 0 and 1, and can flush all frames when packet 2 arrives. (Frame 0 can be flushed as packet 0 arrives, since it is the earliest frame we hold, and likewise frame 1 from packet 1; we are therefore holding 3,4,6,7 until packet 2 arrives).

If there is loss, then the receiver may wait longer than is strictly necessary before it emits frames. For example, say packet 1 is lost from the above example. Packet 0 allows frame 0 to be emitted, and then packet 2 arrives, allowing us to notice the loss of frame 1, and emit frame 2 and 3. Then it is not until the arrival of packet 3 (which has a time-stamp beyond the times of all the frames seen so far), that we can finish dealing with the loss, even though the first group has, in fact, ended. (This is in contrast to schemes which signal the group size explicitly; if the receiver knows that this is packet 3 of 3, then even if 2 of 3 is missing, it can de-interleave this group without waiting for the next one to start).

In the above example the AU-Index is coded with the value 0, as required for the modes defined in this document. To reconstruct the original order, the RTP time stamp and the AU-Index-delta are used. See also 3.2.3.2.

A.1.2 Continuous interleave

In continuous interleave, once the scheme is 'primed', the number of frames in a packet exceeds the 'stride' (the distance between them). This shortens the buffering needed, smooths the data-flow, and gives slightly larger packets -- and thus lower overhead -- for the same interleave. For example, here is a continuous interleave also over a stride of 3 frames, but with 4 frames per packet, for a run of 20 frames. This shows both how the scheme 'starts up' and how it finishes.

Packet	Time-stamp	Frame Numbers	AU-Index, AU-Index-delta
0	T[0]	0	0
1	T[1]	1 4	0 2
2	T[2]	2 5 8	0 2 2
3	T[3]	3 6 9 12	0 2 2 2

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL

T 650.949.6780
F 650.949.6785
www.isma.tv

4	T[7]	7	10	13	16	0	2	2	2
5	T[11]	11	14	17	20	0	2	2	2
6	T[15]	15	18			0	2		
7	T[19]	19				0			

In this case, the receiver has to buffer only 3 frames, not 4. Say we are waiting for packet 4. We can flush frames 0, 1, 2, 3, 4, 5,6; we are holding therefore 8, 9, 12. Packet 4 arrives, allowing us to emit 7,8,9,10, and we are holding 12,13,16. Each arriving packet contains 4 frames, and allows 4 frames to be flushed.

In the above example the AU-Index is coded with the value 0, as required for the modes defined in this document. To reconstruct the original order, the RTP time stamp and the AU-Index-delta are used. See also 3.2.3.2.

If there is loss, again the receiver has to wait to emit the erasure frames. In this case, say packet 3 is lost. We were holding frames 4,5, and 8. On the arrival of packet 4, (time-stamp of frame 7), we now know frame 3 was lost, we can emit frames 4,5, and we know 6 must be lost, and emit 7, which is in the packet that arrived. Then on the arrival of packet 5 (time-stamp 11) we can emit 8, indicate loss of 9, and emit 10 and 11. Finally, the arrival of packet 6(time-stamp 15) indicates that 12 must be lost; we have now detected all the lost frames.

Copyright © 2001 Internet Streaming Media Alliance. All Rights Reserved.
ISMA is a Trademark and Service Mark of Internet Streaming Media Alliance

CONFIDENTIAL