ISMA™
Internet Streaming Media Alliance

Presidio of San Francisco
P.O. Box 29920
572 B Ruger Street
San Francisco, CA 94129-0920

T 415.561.6276
F 415.561.6120

www.isma.tv

Internet Streaming Media Alliance Implementation Specification

Version 2.0

April 2005

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 PRODUCTDS 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.

Table Of Contents

Document Status 4
1 Acronyms & Terms 4
2 Definitions 4
3 Scope 4
4 Architecture and Functions 5
5 Media Formats 5
5.1 Audio and Video Support 5
5.2 Audio 6
5.3 Video 6
5.4 Profiles 7
5.4.1 Profile 2 7
5.4.2 Profile 3 8
5.4.3 Profile 4 9
6 Media Transport 9
7 Media Control 10
8 Media Description 11
8.1 Additional SDP fields 11
9 Media Storage 12
Annex A: Transmission Offsets and Seeking in Hint Tracks 14
References 16

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 2.0 of the ISMA Implementation Specification.

1 Acronyms & Terms

AAC	Advanced Audio Coding
AAC-LC	Advanced Audio Coding – Low Complexity Object Type
AOT	Audio Object Type
BIFS	Binary Format for Scenes
CIF	Common Intermediate Format (352 x 288)
H.264/AVC	ITU-T Recommendation H.264 | ISO/IEC 14496-10 Advanced Video Coding
HE-AAC	Advanced Audio Coding - High Efficiency AAC Profile
IETF	Internet Engineering Task Force
IOD	Initial Object Descriptor
ISMA	Internet Streaming Media Alliance
kbps	1000 bits per second
Mbps	1000000 bits per second
MP4	MPEG-4 File Format
OD	Object Descriptor
PCE	Program Config Element
RFC	Request For Comments
RTP	Real time Transport Protocol
RTSP	Real time Streaming Protocol
SBR	Spectral Bandwidth Replication
SDP	Session Description Protocol
3GPP	3rd Generation Partnership Project

2 Definitions

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 [8].

Paragraphs introduced by the keyword “Rationale:” provide informative information explaining the design choices that were made. They are not a normative part of this specification.

3 Scope

This implementation specification builds on Version 1.0 of the ISMA specification [16] but can be used independently and stand-alone. It is not intended to replace ISMA 1.0 but is expected to coexist with it. The main motivation for version 2.0 is the availability of advanced audio and video codecs that provide significant improvements in coding efficiency and enable new and improved services. Therefore, these codecs are integrated into the existing ISMA framework, allowing vendors to build more competitive products while still benefiting of the interoperability provided by ISMA 1.0.

This specification particularly targets broadband quality networks, i.e. up to 15 Mbps. However, the profiles selected are generic enough to serve a wide range of applications and bit rates. The increased coding efficiency also permits new applications at the low bit rate range, i.e. below 500 kbps.

4 Architecture and Functions

This specification is based on the same architecture and provides the same functionality as ISMA 1.0 [16]. The basic architecture components are a media server, an IP network, and a media client, while allowing for a number of intermediate systems in the transmission chain (storage, transcoders, caches/proxies). The functional areas addressed in this specification are media formats (audio and video), media transport, media control, media description, and media storage. This architecture is outlined in Fig. 1 and specified further in the following.

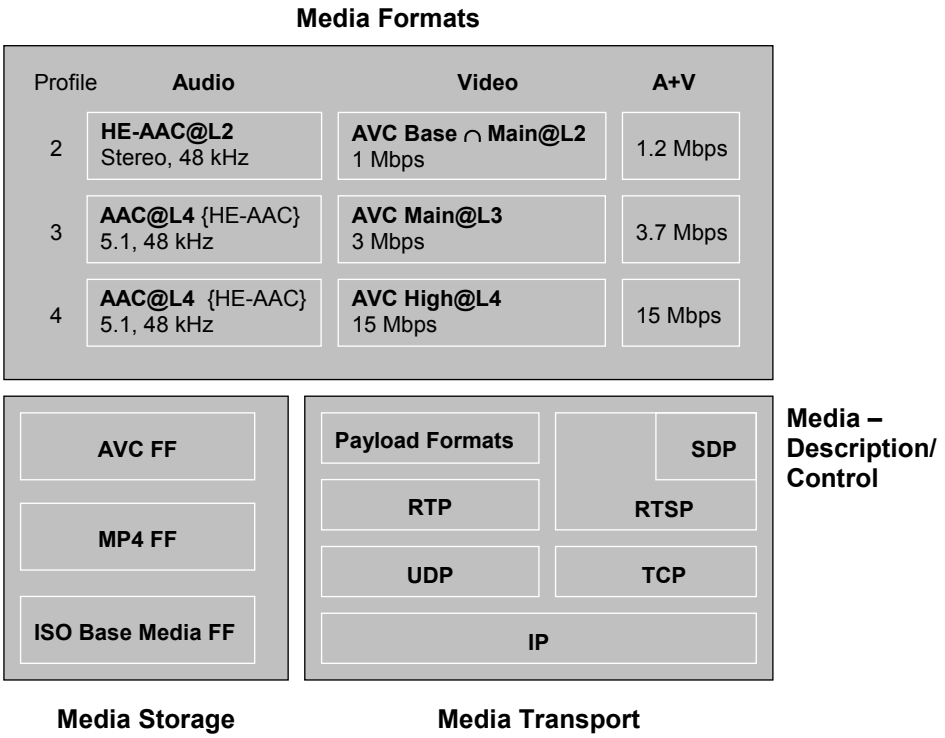

Figure 1: ISMA Architecture Overview

5 Media Formats

5.1 Audio and Video Support

A client capable of decoding and playing audio MUST support audio to this specification. A client capable of decoding and displaying video MUST support video to this specification. To be conformant, at least one of either audio or video MUST be supported.

Rationale: Not all ISMA 2.0 implementations are required to support both video and audio. For example, an audio-only application (e.g. internet radio) is not required to handle video while still being fully compliant to ISMA 2.0. Similarly, a video-only application (e.g. surveillance) is not required to handle audio.

5.2 Audio

The required audio codec in this specification is ISO/IEC 14496-3:2001 / Amd1:2003, MPEG-4 Audio. For details on codec configuration and profiles see Section 5.4.

Channel Configuration and Mixdown:

The channelConfiguration in the AudioSpecificConfig (ASC) SHALL NOT be 0. In the case of channelConfiguration 5 or 6 the mixdown to stereo or mono SHALL be supported as described in the ISO/IEC 14496-3:2001 sub clause 4.5.2.1.3 “Matrix-mixdown method”. The parameters matrix_mixdown_idx and pseudo_surround_enable shall both be set to 0. In the case of channelConfiguration 6 the LFE shall be ignored.

Rationale: Some clients may not be able to render the number of audio channels that are produced by the audio decoder. For example, the audio stream may contain 5.1 surround sound while the client is only capable of stereo output. For this case, a well-defined behavior is defined to ensure predictable user experience.

Support for HE-AAC:

The support for HE-AAC is either optional or mandatory, depending on the ISMA 2.0 profile. In either case, when SBR data is present, it SHALL be signaled explicitly as part of the Audio Specific Configuration.

Rationale: The HE-AAC Profile is a backward compatible superset of the AAC Profile, which allows the extension of the audio bandwidth using Spectral Band Replication (SBR). The required side information is embedded into the audio bit stream and may be skipped by decoders that do not support this extension. The presence of SBR data in the stream does not adversely affect decoders that can only handle AAC-LC. Note that the support for this extension is either optional (backward compatible) or mandatory (non backward compatible) depending on how the audio stream is signaled. In the case that the stream is described as an AAC-LC stream (AOT=2) the encoder may or may not embed SBR data and the decoding of the data is optional. If the stream is described as an HE-AAC stream (AOT=5) the encoder MUST embed SBR data and the decoder MUST decode it. Explicit signaling is required in ISMA 2.0 to allow the decoder to correctly initialize its audio output (sampling rate, buffer size) immediately during setup, and not while decoding the first frame. See also ISO/IEC 14496-3:2001, sub clause 1.6.5 “Signaling of SBR”.

Note: The performance improvement of the HE-AAC Profile compared to the AAC Profile is most significant at very low bit rates, i.e. below 48 kbps per channel.

5.3 Video

The required video codec in this specification is MPEG-4 ISO/IEC 14496-10:2003, Advanced Video Coding (also known as AVC and H.264). For details on codec configuration and profiles see Section 5.4.

RECOMMENDED – Video encoders SHOULD insert all the information that is available to them about the nature of the video signal in the VUI parameters, particularly the pixel aspect ratio, full-range, transfer characteristics and the overscan flag.

REQUIRED – If the VUI parameters pixel aspect ratio, full-range, or transfer characteristics are present, they MUST be respected by the decoder and display system. On display systems capable of displaying overscanned video, the overscan flag MUST be respected if present (that is, if the content is marked as being unsuitable for overscanned display, overscan display must not be used).

5.4 Profiles

In order to assure interoperability, the codecs and protocols adopted in this specification MUST be used in particular configurations and combinations. The selected ISMA 2.0 profiles are intended to match certain market requirements and complexity constraints. To avoid any confusion with ISMA 1.0 profiles (0,1) the numbering is continued in ISMA 2.0 with profiles 2-4.

ISMA 2.0 Profile 2 covers low-resolution audio/video; ISMA 2.0 Profile 3 covers standard definition, and ISMA profile 4 covers high definition. Profiles 2-4 are *not* hierarchical. There is a core of intersection between them — AVC main/baseline at L2, and AAC-LC at L2 are decodable in all profiles. However, Profile 2 requires audio support for HE-AAC (SBR) whereas it is optional in the other two profiles, and Profile 4 requires support for AVC High Profile, which is a superset of the AVC Main Profile used in Profiles 2 and 3.

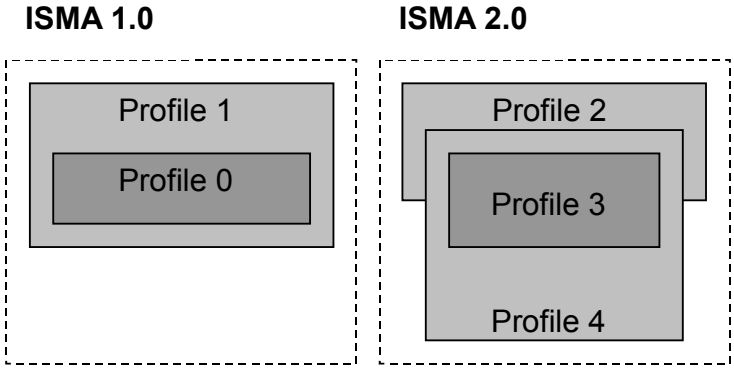

Figure 2: ISMA Profiles

ISMA 2.0 is not backwards compatible with ISMA 1.0, i.e. ISMA 2.0 servers and clients are not required to also support ISMA 1.0.

5.4.1 Profile 2

Rationale: This profile is selected for bit rates that allow low-resolution audio/video as well as audio-only services. A typical maximum video frame size that is supported in this profile is e.g. CIF (352x288). The maximum frame rate supported for a frame size of CIF is 30 fps, and may be higher for smaller frame sizes.

Video:

REQUIRED - MPEG-4 ISO/IEC 14496-10:2003 [1]
Baseline Profile @ Level 2

Restrictions:

The video stream SHALL also be constrained to be a subset of the Main Profile as defined in Annex A of MPEG-4 ISO/IEC 14496-10 (constraint_set1_flag SHALL be set to 1 in all Sequence Parameter Sets).

The maximum bit rate for video in Profile 2 is 1 Mbps.

Audio:

REQUIRED - MPEG-4 ISO/IEC 14496-3:2001 [2]
HE-AAC Profile @ Level 2

- up to 2 channels
- up to 48000 Hz sampling rate

The support of HE-AAC [3] is REQUIRED in Profile 2. The encoder MAY use SBR – but is not required to do so. In this profile decoders MUST decode HE-AAC and therefore encoders MAY use the non-backwards-compatible HE-AAC signaling mode (AOT=5).

Note: Since the signaling of SBR may be non-backward-compatible in this profile, AAC decoders may not be able to decode Profile 2 audio streams. This applies to earlier ISMA 1.0 decoders as well as to ISMA 2.0 decoders compliant to Profiles 3 and 4.

Cumulative Bit Rate Limitation:

The combined audio and video bit rate in Profile 2 is limited to 1.2 Mbps.

5.4.2 Profile 3

Rationale: This profile is selected to allow for video at bit rates that match the capabilities of broadband rates, and is mainly targeted towards qualities suitable for high-resolution applications. Profile 3 intersects Profile 2, i.e. it is possible to make bitstreams decodable in both profiles if HE-AAC is either unused or used in a backward-compatible fashion. A typical maximum video frame size supported in this profile is e.g. D1-PAL (720x576). This profile provides support for both interlaced and progressive mode video.

Video:

REQUIRED - MPEG-4 ISO/IEC 14496-10:2003 [1]
Main Profile @ Level 3

Restriction: The maximum bit rate for video in Profile 3 is 3 Mbps.

Audio:

REQUIRED - MPEG-4 ISO/IEC 14496-3:2001 / Amd1:2003 [2]
AAC Profile @ Level 4

- up to 5.1 channels
- up to 48000 Hz sampling rate

The use of HE-AAC [3] is OPTIONAL in Profile 3. The encoder MAY use SBR – but is not required to do so. If SBR data is present the decoder MAY decode it – but is not required to do so. If SBR data is present, explicit backwards compatible signaling SHALL be used.

Cumulative Bit Rate Limitation:

The combined audio and video bit rate in Profile 3 is limited to 3.7 Mbps.

5.4.3 Profile 4

Rationale: This profile is selected to allow for video at bit rates that match the capabilities of local area rates (Ethernet, 802.11g etc.), and is mainly targeted towards qualities suitable for high-resolution applications. Profile 3 is a subset of Profile 4, in that all High-Profile AVC decoders must also decode Main Profile AVC. A typical maximum video frame size supported in this profile is e.g. HD (720p @ 60 fps, or 2k by 1k @ 30 fps). This profile provides support for both interlaced and progressive mode video.

Video:

REQUIRED - MPEG-4 ISO/IEC 14496-10:2003 [1]
High Profile @ Level 4

Restriction: The maximum bit rate for video in Profile 4 is 15 Mbps.

Audio:

REQUIRED - MPEG-4 ISO/IEC 14496-3:2001 / Amd1:2003 [2, 3]
AAC Profile @ Level 4

- up to 5.1 channels
- up to 48000 Hz sampling rate

Cumulative Bit Rate Limitation:

The combined audio and video bit rate in Profile 4 is limited to 15 Mbps.

6 Media Transport

The following IETF documents specify the transport over IP:

REQUIRED - User Datagram Protocol
IETF RFC 769 [9]

REQUIRED - RTP: A Transport Protocol for Real-Time Applications
IETF RFC 3550 [11]

REQUIRED - RTP Profile for Audio and Video Conferences with Minimal Control
IETF RFC 3551 [12]

OPTIONAL - Interleaved RTSP & RTP/AVP over TCP transport
Real Time Streaming Protocol - IETF RFC 2326, Section 10.12 [13]
Transmission Control Protocol - IETF RFC 793 [10]

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. Furthermore, RTSP/RTP over TCP permits traversal of Network Address Translators and Firewalls.

Video:

REQUIRED - RTP Payload Format for H.264 Video
IETF Internet Draft - Version 11 [7]

Restrictions:

- The interleaved mode (packetization-mode=2) SHALL NOT be used.
- The parameters that are defined for interleaved mode (packetization-mode=2) SHALL NOT be present in the “a=fmtp” line in the SDP
- The parameters “max-mbps”, “max-fs”, “max-dpb”, “max-br” SHALL NOT be present in the “a=fmtp” line of the SDP.
- The format parameters line (“a=fmtp”) in the SDP SHALL include the following parameters: sprop-parameters-sets, profile-level-id.
- Sequence Parameters Set NALUs and Picture Parameter Set NALUs SHALL NOT be present in RTP packets (the parameter sets are sent out-of-band via SDP signaling).

Audio:

REQUIRED - RTP Payload Format for MPEG-4 Streams
IETF RFC 3460 [14]

REQUIRED - When transmitting MPEG-4 AAC streams over RTP the High Bit-rate AAC mode must be used (in SDP, “mode=AAC-hbr” must be used on the “a=fmtp” line).

7 Media Control

REQUIRED - Real Time Streaming Protocol (RTSP)
IETF RFC 2326 [13]

RTSP clients and servers SHALL implement all required features of the minimal RTSP implementation described in Appendix D of RFC 2326.

REQUIRED – RTSP clients and servers SHALL implement the PLAY method.

RECOMMENDED – RTSP clients and servers SHOULD implement the DESCRIBE method. If the DESCRIBE method is implemented, it is REQUIRED that SDP is supported as the description format, as specified in Appendix C of RFC 2326.

REQUIRED – RTSP clients and servers MUST support RTP/AVP transport in the “Transport” header. When the RTP/AVP transport is used for a unicast session, clients SHOULD include the “client_port” parameter in the “Transport” header and servers SHOULD include the “server_port”, “source”, and “ssrc” parameters in the “Transport” header.

REQUIRED – RTSP servers SHALL send the “RTP-Info” header for unicast sessions.

RECOMMENDED – RTSP clients SHOULD generate the following RTSP headers when appropriate: “Bandwidth”, “Cache-Control”, “If-Modified-Since”, “User-Agent”. RTSP servers SHOULD correctly interpret these headers when present.

RECOMMENDED – RTSP servers SHOULD generate the following RTSP headers when appropriate: “Cache-Control”, “Expires”, “Last-Modified”, “Server”. RTSP clients SHOULD correctly interpret these headers when present.

REQUIRED – RTSP servers and clients SHALL support aggregated control of presentations.

REQUIRED – At most one RTSP session shall be “active” on a connection between an RTSP client and an RTSP server at any one time. An RTSP session becomes “active” when it is first referenced in a “Session” header. An RTSP session is no longer “active” after a TEARDOWN request has been issued for that session.

8 Media Description

REQUIRED - SDP: Session Description Protocol
IETF RFC 2327 [15]

The SDP data should be formatted according to SDP [15] and Appendix C of RTSP [13].

If an IOD is generated in the SDP, servers MUST generate the SDP IOD attribute using a binary IOD (not XML). If the URL is a data URL, then the binary IOD is encapsulated directly in the URL using base64 encoding. Since IOD support in this specification is optional, detailed examples and encodings are not given.

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

8.1 Additional SDP fields

The following SDP field SHALL be present to indicate ISMA compliance:

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

where:

<profile> is an integer specifying the ISMA profile to which the content conforms.

<lowest-spec-version> is 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> is 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.

The normal usage when conformant to this specification is to use “2.0” as both the lowest-spec-version and the authored-to-version. However, for some content (e.g. audio-only content using AAC or backwards-compatible HE-AAC) it is possible to include the IOD, OD and BIFS information and indicate that the file was authored to ISMA 2.0 but is playable using an ISMA 1.0 client.

The following media level SDP field is defined in ISMA 2.0:

a=X-initpredecbufperiod:<initial pre-decoder buffering period>

The field MAY be present in audio and/or video streams and indicates the amount of buffering before playback, i.e. the amount of time in between the reception of the first packet of the media stream and removing it from the buffer for decoding. Values are given in clock ticks based on the RTP time scale of the corresponding media stream, i.e., a 90-kHz clock for H.264/AVC video and typically the sampling rate for audio.

Note that this SDP field is aligned with 3GPP PSS Rel-6 [17] and in case it is part of an H.264/AVC stream it SHALL be interpreted identically, i.e.:

> *If the field is an attribute for an H.264 (AVC) stream, the H.264 (AVC) bitstream is constrained by the value of the nominal removal time of the first access unit from the coded picture buffer (CPB), tr,n(0), equal to "X-initpredecbufperiod" as specified in [1]. If "X-initpredecbufperiod" is not present for an H.264 (AVC) stream, tr,n(0) shall be equal to the earliest time when the first access unit in decoding order has been completely received.*

9 Media Storage

In order to encourage content distribution and interoperability between products at the storage level, the ISO base media file format (including extensions to support MP4 and AVC), is the obvious choice for the storage of media content. Hence, file storage for ISMA 2.0 is based on the following specifications:

REQUIRED - ISO base media file format
ISO/IEC 14496-12:2003 | 15444-12:2003 [4]

REQUIRED - MPEG-4 MP4 file format
ISO/IEC 14496-14:2003 [5]

REQUIRED - MPEG-4 AVC File Format
ISO/IEC 14496-15:2003(E) [6]

Restrictions:

Files compliant with ISMA 2.0 shall also adhere to the following restrictions:

- AVC Parameter sets SHALL be stored within the AVCDecoderConfiguration only. The use of Parameter set elementary streams is disallowed;
- Compact Sample Size boxes ('stz2') SHALL NOT be used;
- Padding Bits boxes ('padb') SHALL NOT be used;
- Movie Fragments boxes ('moof') SHALL NOT be used;
- In the Data Reference box ('dref') of files which are to be placed on servers (hinted or unhinted), the data sources are restricted to being from the following set:
    - the same file as the Movie box meta-data ('self-contained')
    - a relative URL (interpreted as relative to the file containing the movie box meta-data)
    - an absolute URL using the "file:" 'protocol';
- Sample Description boxes ('stsd') SHALL contain exactly one sample entry (the entry_count field SHALL be equal to 1).

Recommendations:

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

Progressive Download Guidelines:

To enable progressive download, session information should always be put at the beginning of the file and the media tracks should be interleaved within the file. The following guidelines should be followed during the creation of ISMA 2.0 files for progressive download:

- The Movie box ('moov') SHOULD be placed at the start of the file, right after the File Type box ('ftyp').
- All media data SHOULD be present within the file (the data source of every Data Reference box should be 'self-contained');
- The media tracks SHOULD be interleaved inside the file. The typical interleaving depth is 1 second.

The meaning of interleaving depth is given as follows: Each chunk contains one or more samples, with the total duration of the samples being either: no greater than 1 second, or the duration of a single sample if that sample's duration is greater than 1 second. Within a track, chunks must be in decoding time order within the Media Data box ('mdat').

A client that does not support progressive download can always play the file once it has been completely downloaded. A progressive download client can start playing a ISMA 2.0 file that has been created along the progressive download guidelines once it has received a first chunk of all media in the session.

File Identification:

ISMA 2.0 files SHOULD be given the file extension '.mp4'.

If the extensions as documented in Clause 4 of the AVC File Format specification [6] are used, then the major brand SHALL be 'avc1', otherwise the major brand SHOULD be 'mp41' or 'mp42'.

The presence of the brand identifier 'ism2' in the compatible brands list indicates that the file conforms to this specification and is in accordance with the restrictions imposed above. Furthermore, if the ISMA 2.0 file also complies with the above progressive download guidelines, the brand identifier 'ismp' SHOULD be included in the compatible brands list. However, if the brand identifier 'ismp' is included in the compatible brands list then the above guidelines for progressive download SHALL be followed.

Annex A: Transmission Offsets and Seeking in Hint Tracks

A.1 Problem Area

There is a standard format for RTP hint tracks in the ISO (MP4) file format, and that format includes a field called relative-time, a signed integer, which allows 'warping' the transmission time away from the nominal time corresponding to the RTP timestamp.

This annex provides more text on the use and interpretation of this field.

A.2 Background

An RTP hint track consists of a sequence of RTP Hint samples. Each sample consists of a set of packets. Each packet has a field, relative-time, which is the field in question.

Each sample has a timestamp which is coded by the time-to-sample table (by difference values; the time of each sample is computed by summing the prior durations, with the first sample having a timestamp of zero). This timestamp is used to form the RTP timestamp; therefore all packets within a sample share the same RTP timestamp. Normally RTP packets are sent at times corresponding to their timestamp, but it is useful both to be able to 'spread out' the burst of packets that come from a large sample, and to smooth traffic. The documentation of the relative-time field (a signed 32-bit number) says:

> The relative-time field 'warps' the actual transmission time away from the sample time. This allows traffic smoothing.

A.3 Writer constraints

The relative-time field MUST NOT be used to cause transmission re-ordering. Packets are sent:

a) within a sample, in order;

b) the first packet of a sample is sent after the last packet of the previous sample.

This means that within a sample, the packets MUST be ordered so that the relative-time fields are non-decreasing. The calculated transmission times for the packets in a sample N MUST be greater than the packets in samples K<N and MUST be less than the transmission time of packets in samples L>N.

Note: This means that the server can seek and that seek points perfectly divide the timeline into packets preceding the seek time, and succeeding it. (The algorithms to 'back up' to a sync point are still needed, of course).

If the use of the relative-time field requires a client buffer larger than the default, then the writer MUST indicate that in the SDP information using the X-initpredecbufperiod parameter.

A.4 Server Behavior

The name 'hint track' suggests correctly that the server behavior is not normative, but informative. Servers are free to apply other data-rate smoothing than that suggested by the relative-time field, and other transmission principles. However, they should take care to operate within the bounds of the client buffer. For improved interoperability it is RECOMMENDED to use the algorithm provided in this sub-section.

When a server is requested to start playing, it first identifies the first sample, N, to send. The transmission time of the first packet in that sample, T[N,1] is then calculated, as:

$$T[N,1] = \sum_{k=1}^{N-1} sampleduration[k] + \text{relative-time}[N,1].$$

That packet is sent at local time L[N,1]. The second packet in the sample is then sent at the correct interval after the first packet:

L[N,2] = L[N,1] + (T[N,2] - T[N,1])

and so on; each packet is sent correctly spaced from the previous packet.

If there is more than one track, this algorithm can be applied to the tracks independently; for each track, one packet is sent 'immediately' and following packets are sent correctly timed after that. This avoids the problem of one track being 'starved' if its initial relative-time value is significantly more.

References

[1] ITU-T Recommendation H.264 (2003): “Advanced video coding for generic audiovisual services” | ISO/IEC 14496-10:2003: “Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding”.

[2] ISO/IEC 14496-3:2001, “Information Technology – Coding of audio-visual objects – Part 3: Audio”.

[3] ISO/IEC 14496-3:2001/Amd 1:2003, “Bandwidth Extension”.

[4] ISO/IEC 14496-12:2003 | 15444-12:2003: “Information technology – Coding of audio-visual objects – Part 12: ISO base media file format” | “Information technology – JPEG 2000 image coding system – Part 12: ISO base media file format”.

[5] ISO/IEC 14496-14:2003: “Information technology – Coding of audio-visual objects – Part 14: MP4 file format”.

[6] ISO/IEC 14496-15: 2004: “Information technology – Coding of audio-visual objects – Part 15: Advanced Video Coding (AVC) file format”.

[7] IETF RFC 3984: “RTP payload Format for H.264 Video”, Wenger S. et al., February 2005.

[8] IETF RFC 2119: “Key words for use in RFCs to Indicate Requirement Levels”, S. Bradner, March 1997.

[9] IETF STD 0006: “User Datagram Protocol”, Postel J., August 1980.

[10] IETF STD 0007: “Transmission Control Protocol”, Postel J., September 1981.

[11] IETF RFC 3550: “RTP: A Transport Protocol for Real-Time Applications”, Schulzrinne H. et al., July 2003.

[12] IETF RFC 3551: “RTP Profile for Audio and Video Conferences with Minimal Control”, Schulzrinne H. and Casner S., July 2003.

[13] IETF RFC 2326: “Real Time Streaming Protocol (RTSP)”, Schulzrinne H., Rao A. and Lanphier R., April 1998.

[14] IETF RFC 3640: “RTP Payload Payload Format for MPEG-4 Streams”, J. Van Der Meer et al., Nov 2003.

[15] IETF RFC 2327: “SDP: Session Description Protocol”, Handley M. and Jacobson V., April 1998.

[16] ISMA 1.0.1: “Internet Streaming Media Alliance Implementation Specification, Version 1.0.1”, June 3, 2004.

[17] 3GPP TS 26.234 V6.1.0, “Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs (Release 6)”