3GPP TSG-SA Plenary Meeting #25
Palm Springs, CA, USA, 13 - 16 September 2004

Tdoc ⌘SP-040657

CR-Form-v7

CHANGE REQUEST

⌘ **26.244** **CR** **004** ⌘ **rev** **1** ⌘ Current version: **6.0.0** ⌘

*For **HELP** on using this form, see bottom of this page or look at the pop-up text over the ⌘ symbols.*

Proposed change affects: UICC apps⌘ ☐ ME ☒ Radio Access Network ☐ Core Network ☐

Title:	⌘	Storage of H.264 (AVC) video in 3GP files
Source:	⌘	Apple Computer, AT&T Wireless Services, Ericsson (editor), France Telecom, Fraunhofer, Nokia, ORANGE, PacketVideo, Panasonic, Philips, RealNetworks, Sharp, STMicroelectronics, Texas Instruments, Toshiba, Vodafone
Work item code:	⌘	PSSrel6-Stage3
Date:	⌘	10/09/2004
Category:	⌘	**B**
Release:	⌘	Rel-6

Use one of the following categories:
F *(correction)*
A *(corresponds to a correction in an earlier release)*
B *(addition of feature),*
C *(functional modification of feature)*
D *(editorial modification)*
Detailed explanations of the above categories can be found in 3GPP TR 21.900.

Use one of the following releases:

2	*(GSM Phase 2)*
R96	*(Release 1996)*
R97	*(Release 1997)*
R98	*(Release 1998)*
R99	*(Release 1999)*
Rel-4	*(Release 4)*
Rel-5	*(Release 5)*
Rel-6	*(Release 6)*

Reason for change:	⌘	Enable 3GP files to carry H.264 (AVC) video.
Summary of change:	⌘	The 3GP file format is extended to inlcude the AVC file format. The storage of video buffer information for H.264 video in 3GP files is defined.
Consequences if not approved:	⌘	An MMS cannot contain H.264 clips. Terminals cannot download video clips with H.264. PSS servers cannot stream H.264 video streams from 3GP files.

Clauses affected:	⌘	2, 3.2, 5.2.2, 5.2.3, 5.2.4, 5.4.3, 5.5, 6.1, 9.1, 9.2, 9.2.1, 9.2.2

		Y	N			
Other specs affected:	⌘	Y		Other core specifications	⌘	TS 26.140, TS 26.234
			X	Test specifications		
			X	O&M Specifications		

Other comments:	⌘	3GP files with H.264 can be used by MMS (TS 26.140) and PSS (TS 26.234).

How to create CRs using this form:
Comprehensive information and tips about how to create CRs can be found at http://www.3gpp.org/specs/CR.htm. Below is a brief summary:

1) Fill out the above form. The symbols above marked ⌘ contain pop-up help information about the field that they are closest to.

2) Obtain the latest version for the release of the specification to which the change is proposed. Use the MS Word "revision marks" feature (also known as "track changes") when making the changes. All 3GPP specifications can be downloaded from the 3GPP server under ftp://ftp.3gpp.org/specs/ For the latest version, look for the directory name with the latest date e.g. 2001-03 contains the specifications resulting from the March 2001 TSG meetings.

3) With "track changes" disabled, paste the entire CR form (use CTRL-A to select it) into the specification just in front of the clause containing the first piece of changed text. Delete those parts of the specification which are not relevant to the change request.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document.*

[1] 3GPP TS 22.233: "Transparent End-to-End Packet-switched Streaming Service; Stage 1".

[2] 3GPP TS 26.233: "Transparent end-to-end packet switched streaming service (PSS); General description".

[3] 3GPP TS 26.234: "Transparent end-to-end packet switched streaming service (PSS); Protocols and codecs".

[4] 3GPP TS 26.245: "Transparent end-to-end packet switched streaming service (PSS); Timed text format".

[5] 3GPP TS 26.246: "Transparent end-to-end packet switched streaming service (PSS); 3GPP SMIL Language Profile".

[6] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

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

[8] 3GPP TS 26.140: "Multimedia Messaging Service (MMS); Media formats and codecs".

[9] ITU-T Recommendation H.263 (1998): "Video coding for low bit rate communication".

[10] ISO/IEC 14496-2:2001: "Information technology ñ Coding of audio-visual objects ñ Part 2: Visual".

[11] 3GPP TS 26.071: "Mandatory Speech CODEC speech processing functions; AMR Speech CODEC; General description".

[12] 3GPP TS 26.171: "AMR Wideband Speech Codec; General Description".

[13] ISO/IEC 14496-3:2001: "Information technology ñ Coding of audio-visual objects ñ Part 3: Audio".

[14] ISO/IEC 14496-14:2003: "Information technology ñ Coding of audio-visual objects ñ Part 14: MP4 file format".

[15] IETF RFC 3267: "Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", Sjoberg J. et al., June 2002.

[16] 3GPP TS 26.101: "Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec frame structure".

[17] 3GPP TS 26.201: "Speech Codec speech processing functions; AMR Wideband Speech Codec; Frame Structure".

[18] ITU-T Recommendation H.263 ñ Annex X (2001): "Annex X: Profiles and levels definition".

[19] IETF RFC 3711: "The Secure Real-time Transport Protocol", Baugher M. et al., Feb 2004.

[20] ISO/IEC 14496-15: "Information technology ñ Coding of audio-visual objects ñ Part 15: Advanced Video Coding (AVC) file format".

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

[22] IETF Internet Draft: "RTP payload Format for H.264 Video", Wenger S. et al, http://www.ietf.org/internet-drafts/draft-ietf-avt-rtp-h264-10.txt, July 2004.

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [6] and the following apply.

3GP	3GPP file format
AAC	Advanced Audio Coding
AVC	Advanced Video Coding
BIFS	Binary Format for Scenes
ITU-T	International Telecommunications Union ñ Telecommunications
MIME	Multipurpose Internet Mail Extensions
MMS	Multimedia Messaging Service
MP4	MPEG-4 file format
PSS	Packet-switched Streaming Service
RTP	Real-time Transport Protocol
RTSP	Real-Time Streaming Protocol
SDP	Session Description Protocol
SRTP	Secure Real-time Transport Protocol

5.2.2 Registration of codecs

Code streams for H.263 video [9], MPEG-4 video [10], H.264 (AVC) video [21], AMR narrow-band speech [11], AMR wide-band speech [12], MPEG-4 AAC audio [13], and timed text [4] can be included in 3GP files as described in clause 6 of the present document.

5.2.3 Extensions

The following extensions to the ISO base media file format [7] can be used in a 3GP file:

- streaming-server extensions (see clause 7);
- asset information (see clause 8);
- video-buffer information (see clause 9);
- encryption (see clause 10);
- AVC file format (see [20]).

If SDP information is included in a 3GP file, it shall be used as defined by the streaming-server extensions.

5.2.4 MPEG-4 systems specific elements

For the storage of MPEG-4 media specific information in 3GP files, this specification refers to MP4 [14] and the AVC file format [20], which are also based on the ISO base media file format. However, tracks relative to MPEG-4 system architectural elements (e.g. BIFS scene description tracks or OD Object descriptors) are optional in 3GP files and shall be ignored. The inclusion of MPEG-4 media does not imply the usage of MPEG-4 systems architecture. Terminals and servers are not required to implement any of the specific MPEG-4 system architectural elements.

5.4.3 Basic profile

The 3GP Basic profile is branded ë3gp6í and is used in MMS and PSS. Conformance to this profile will guarantee the 3GPP file format to be used internally within the MMS service, as well as PSS to interwork with MMS.

The following constraints shall apply to a 3GP file conforming to Basic profile:

- there shall be no references to external media outside the file, i.e. a file shall be self-contained;
- the maximum number of tracks shall be one for video, one for audio and one for text;
- the maximum number of sample entries shall be one per track for video and audio (but unrestricted for text).

NOTE: The Basic profile of 3GP in Release 6 corresponds to 3GP files of earlier releases, which did not define profiles. Files with brands ë3gp4í and ë3gp5í in Release 4 and 5, respectively, correspond to files with brand ë3gp6í in Release 6.

NOTE: For H.264 (AVC) video in a Basic profile 3GP file, the restriction on the number of video tracks implies in particular that there shall be no alternative tracks (including switching tracks) and no separate tracks for parameter sets.

5.5 File-branding guidelines

The file-type brands defined in this specification are used to label 3GP files belonging to Release 6 and conforming to one or more profiles. 3GP files may also conform to earlier Releases or even to other file formats, such as MP4, which is also derived from the ISO base media file format [7].

Table 5.1 contains a non-exhaustive list of examples with 3GP files for various purposes. Note, however, that it only gives typical or suggested uses. Both writers and readers of files should exercise care when using brand identifiers. It is worth repeating the general guidelines here, remembering that a brand identifies a specification or a conformance point in a specification; its presence in a file indicates both:

- that the file conforms to the specification; it includes everything required by, and nothing contrary to the specification (though there may be other material);
- that a reader implementing that specification (possibly only that specification) is given permission to read and interpret the file.

All 3GP files of Release 5 or later shall contain the compatible brand ëisomí indicating that they conform to the ISO base media file format, unless the reader is required to interpret extensions specific to the AVC file format [20], for which case the~~.~~ compatible brand ëavc1í shall be used instead (see note 2). The major brand shall be included in the compatible brands list as well. If a file contains more than one (3GPP) brand in the compatible brands list, the major brand indicates the ìbest useî of the file. For example, a Release-5 file with audio combined with Timed text is best played by a Release-5 player, but may also be played by a Release-4 player that does not support timed text.

NOTE 2: Consider the brands ëisomí and ëavc1í. The first indicates conformance to the base structure of the ISO base media file format (first version) [7]. The second, conformance to the AVC-specific extensions (structures such as sample groups, for example) [20]. A file labelled as ëisomí and ëavc1í conformant is indicating that either these extensions are not present, or if present, they can be ignored (as an ëisomí reader will not understand them). If the writer desires that only readers supporting the extensions read a file, then the ëisomí brand would be omitted. These extensions are all optional (i.e. none are required to be in a file, though if they are, an ëavc1í-conformant reader must interpret them), and therefore a file not using them is still ëavc1í conformant.

Table 5.1: Examples of brand usage in 3GP files

Conformance	Suffix	Brand	Compatible brands	Example content
MMS and download: Files shall contain one or more of the brands 3gp4, 3gp5 and 3gp6. It is good practice to include compatible brands of earlier releases to enable legacy players to play the files.				
Release 4	.3gp	3gp4	3gp4	H.263 and AMR
Release 5, 4	.3gp	3gp5	3gp5, 3gp4, isom	H.263 and AMR
Release 6, 5, 4	.3gp	3gp6	3gp6, 3gp5, 3gp4, isom	H.263 and AMR
Release 6, 5, 4	.3gp	3gp6	3gp6, 3gp5, 3gp4, isom	H.263, AMR and Timed text
Release 6, 5	.3gp	3gp6	3gp6, 3gp5, isom	Timed text
Release 6	.3gp	3gp6	3gp6, isom	H.264 (AVC) and AMR~~Some Release-6 specific codec TBD~~
Progressive download and MMS				
Release 6, 5, 4	.3gp	3gr6	3gr6, 3gp6, 3gp5, 3gp4, isom	H.263
Release 6, 5, 4	.3gp	3gr6	3gr6, 3gp6, 3gp5, 3gp4, isom	interleaved H.263 and AMR
Release 6	.3gp	3gr6	3gr6, 3gp6, avc1	interleaved H.264 (AVC) and AMR
Streaming servers: Some files may in principle also be used for MMS or download.				
Release 6	.3gp	3gs6	3gs6, isom	AMR and hint track
Release 6	.3gp	3gs6	3gs6, isom	2 tracks H.263 and 2 hint tracks
Release 6, 5, 4	.3gp	3gs6	3gs6, 3gp6, 3gp5, 3gp4, isom	H.263, AMR and hint tracks
General purpose: Files that are not yet suitable for MMS, download or PSS streaming servers.				
Release 6	.3gp	3gg6	3gg6, isom	4 tracks H.263 (and no hint tracks)
Release 6	.3gp	3gg6	3gg6, isom	2 tracks H.263, 3 tracks AMR
3GP file, also conforming to MP4				
Release 4, 5 and MP4	.3gp	3gp5	3gp5, 3gp4, mp42, isom	MPEG-4 video
MP4 file, also conforming to 3GP				
Release 5 and MP4	.mp4	mp42	mp42, 3gp5, isom	MPEG-4 video and AAC

6 Codec registration

6.1 General

The purpose of this clause is to define the necessary structure for integration of the H.263, MPEG-4 video, AMR, AMR-WB, and AAC media specific information in a 3GP file. Clause 6.2 gives some background information about the Sample Description box in the ISO base media file format [7] and clauses 6.3 and 6.4 about the MP4VisualSampleEntry box and the MP4AudioSampleEntry box in the MPEG-4 file format [14]. The definitions of the Sample Entry boxes for AMR, AMR-WB and H.263 are given in clauses 6.5 to 6.8. The integration of timed text in a 3GP file is specified in [4] and the integration of H.264 (AVC) is specified in [20].

AMR and AMR-WB data is stored in the stream according to the AMR and AMR-WB storage format for single channel header of Annex E [15], without the AMR magic numbers.

9 Video buffer information

9.1 General

A 3GP file can include video-buffer parameters associated with video streams. For the case when only one set of parameters is associated to an entire video stream, these can be included in the corresponding media-level SDP fragment. However, in order to provide buffer parameters for different operation points, as defined below, and for different synchronization points, a track can contain a video buffer sample grouping. The type of sample grouping depends on which video-buffer model that is used for a particular video codec.

For H.263 and MPEG-4 visual, ~~of~~ the PSS buffering model, defined in Annex G of TS 26.234 [3] (PSS Annex G), is used. Buffer parameters for several operation points and synchronization points ~~that an associated video stream conforms to. For the case when only one set of parameters is associated to an entire video stream, these can be included in the corresponding media-level SDP fragment. However, in order to provide buffer parameters for different operation points, as defined below, and for different synchronization points, a track can contain~~ may be specified by a 3GPP PSS Annex G sample grouping as defined in ~~this~~ clause 9.2.1.

For H.264 (AVC), there are two types of buffers:

- H.264 (AVC) Hypothetical Reference Decoder (HRD) model;
- de-interleaving buffer of the interleaved RTP packetization mode of H.264 (AVC).

Buffer parameters for several operation points and synchronization points of the HRD model may be specified by an AVC HRD sample grouping as defined in clause 9.2.2. Only one set of de-interleaving parameters can be associated to a stream and therefore the de-interleaving parameters are included in the corresponding media-level SDP fragment according to the H.264 (AVC) MIME/SDP specification in [22].

NOTE: Any VUI HRD parameters, buffering period SEI message, and picture timing SEI message in H.264 (AVC) streams or included in the sprop-parameter-sets MIME/SDP parameter of a media-level SDP fragment must not contradict each other or the information in the AVC HRD sample grouping, if any.

9.2 Sample groupings for video-buffer parameters

A sample grouping is an assignment of each sample in a track to be a member of one sample group, based on a grouping criterion. The assignment of buffer parameters to synchronization points (sync samples) provides one sample grouping of all samples in a track. The usage of sample groups in 3GP files shall follow the syntax defined in [20].

Each sample group is associated to zero or one sample group entries in the sample group description box ('sgpd'). Sample group entries for sample groups defined by the grouping type '3gag' are given by the 3GPP PSS Annex G Sample group entry, defined in Table 9.1, and sample group entries for sample groups defined by the grouping type 'avcb' are given by the AVC HRD Sample group entry, defined in Table 9.2.

Sample entries provide buffer parameters relevant to all samples in the corresponding sample group(s). A sync sample and all following non-sync samples before the next sync sample shall be members of the same sample group with respect to the video-buffer grouping type. The indicated buffer parameters for a sync sample are applicable for the stream from that sync sample onwards.

NOTE: A file, in which some but not all samples are associated with sample groups with respect to the grouping type '3gag' or 'avcb', may have been edited and may therefore no longer conform to corresponding buffer model.

9.2.1 3GPP PSS Annex G sample grouping

~~A sample grouping is an assignment of each sample in a track to be a member of one sample group, based on a grouping criterion. The assignment of buffer parameters to synchronization points provides one sample grouping of all samples in a track. The usage of sample groups in 3GP files shall follow the syntax defined in [20].~~

The grouping type '3gag' defines the grouping criterion for 3GPP PSS Annex G buffer parameters. Zero or one sample-to-group box ('sbgp') for the grouping type '3gag' can be contained in the sample table box ('stbl') of a track. It shall reside in a hint track, if a hint track is used, otherwise in the video track. The presence of this box and grouping type indicates that the associated video stream complies with PSS Annex G. Note that the nature of the track defines the media transport for which the buffer parameters are calculated, e.g. for an RTP hint track, the media transport is RTP.

~~Each sample group is associated to zero or one sample group entries in the sample group description box ('sgpd'). Sample group entries for sample groups defined by the grouping type '3gag' are given by the 3GPP PSS Annex G Sample group entry defined in Table 9.1. Such a sample entry provides buffer parameters relevant to all samples in the corresponding sample group(s). Note that samples that are not synchronization points shall not be associated with a sample group.~~

Table 9.1: 3GPP PSS Annex G sample group entry

Field	Type	Details	Value
BufferParameters	AnnexGstruc	Structure which holds the buffer parameters of PSS Annex G	

BufferParameters: the structure where the PSS Annex G buffer parameters reside.

AnnexGstruc is defined as follows:

```
struct AnnexGstruc{
                    Unsigned int(16)    operation_point_count
                    for (i = 0; i < operation_point_count; i++){
                        Unsigned int (32)    tx_byte_rate
                        Unsigned int (32)    dec_byte_rate
                        Unsigned int (32)    pre_dec_buf_size
                        Unsigned int (32)    init_pre_dec_buf_period
                        Unsigned int (32)    init_post_dec_buf_period
                    }
}
```

The definitions of the AnnexGstruc members are as follows:

operation_point_count: specifies the number of operation points, each characterized by a pair of transmission byte rate and decoding byte rate. Values of buffering parameters are specified separately for each operation point. The value of operation_point_count shall be greater than 0.

tx_byte_rate: indicates the transmission byte rate (in bytes per second) that is used to calculate the transmission timestamps of media-transport packets for the PSS Annex G buffering verifier as follows. Let t1 be the transmission time of the previous media-transport packet and size1 be the number of bytes in the payload of the previous media-transport packet in transmission order, excluding the media-transport payload header and any lower-layer headers. For the first media-transport packet of the stream, t1 and size1 are equal to 0. The media track shall comply with PSS Annex G when each sample is packetized in one media-transport packet, the transmission order of media-transport packets is the same as their decoding order, and the transmission time of an media-transport packet is equal to t1 + size1 / tx_byte_rate. The value of tx_byte_rate shall be greater than 0.

dec_byte_rate: indicates the peak decoding byte rate that was used in this operation point to verify the compatibility of the stream with PSS Annex G. Values are given in bytes per second. The value of dec_byte_rate shall be greater than 0.

pre_dec_buf_size: indicates the size of the PSS Annex G hypothetical pre-decoder buffer in bytes that guarantees pauseless playback of the entire stream under the assumptions of PSS Annex G.

init_pre_dec_buf_period: indicates the required initial pre-decoder buffering period that guarantees pauseless playback of the entire stream under the assumptions of PSS Annex G. Values are interpreted as clock ticks of a 90-kHz block. That is, the value is incremented by one for each 1/90 000 seconds. For example, value 180 000 corresponds to a two second initial pre-decoder buffering.

init_post_dec_buf_period: indicates the required initial post-decoder buffering period that guarantees pauseless playback of the entire stream under the assumptions of PSS Annex G. Values are interpreted as clock ticks of a 90-kHz clock.

9.2.2 AVC HRD sample grouping

The grouping type 'avcb' defines the grouping criterion for AVC HRD parameters. Zero or one sample-to-group box ('sbgp') for the grouping type 'avcb' can be contained in the sample table box ('stbl') of a track. It shall reside either in a hint track or a video track. The presence of this box and grouping type indicates that the associated video stream complies with AVC HRD with the indicated parameters.

Table 9.2: AVC HRD sample group entry

Field	Type	Details	Value
AVCHRDParameters	AVCHRDstruc	Structure which holds the AVC HRD parameters	

AVCHRDParameters: the structure where the AVC HRD parameters reside.

AVCHRDstruc is defined as follows:

```
struct AVCHRDstruc{
        Unsigned int(16)    operation_point_count
        for (i = 0; i < operation_point_count; i++){
                Unsigned int (32)    tx_byte_rate
                Unsigned int (32)    pre_dec_buf_size
                Unsigned int (32)    post_dec_buf_size
                Unsigned int (32)    init_pre_dec_buf_period
                Unsigned int (32)    init_post_dec_buf_period
        }
}
```

The definitions of the AVCHRDstruc members are as follows:

operation_point_count: specifies the number of operation points. Values of AVC HRD parameters are specified separately for each operation point. The value of operation_point_count shall be greater than 0.

tx_byte_rate: indicates the input byte rate (in bytes per second) to the coded picture buffer (CPB) of AVC HRD. The bitstream is constrained by the value of BitRate equal to 8 * the value of tx_byte_rate for NAL HRD parameters as specified in [21]. For VCL HRD parameters, the value of BitRate is equal to tx_byte_rate * 40 / 6. The value of tx_byte_rate shall be greater than 0.

pre_dec_buf_size: gives the required size of the pre-decoder buffer or coded picture buffer in bytes. The bitstream is constrained by the value of CpbSize equal to pre_dec_buf_size * 8 for NAL HRD parameters as specified in [21]. For VCL HRD parameters, the value of CpbSize is equal to pre_dec_buf_size * 40 / 6.

At least one pair of values of tx_byte_rate and pre_dec_buf_size of the same operation point shall conform to the maximum bitrate and CPB size allowed by profile and level of the stream.

post_dec_buf_size: gives the required size of the post-decoder buffer, or the decoded picture buffer, in unit of bytes. The bitstream is constrained by the value of max_dec_frame_buffering equal to Min(16, Floor(post_dec_buf_size) / (PicWidthMbs * FrameHeightInMbs * 256 * ChromaFormatFactor))) as specified in [21]. If the SDP attribute 3gpp-videopostdecbufsize is not present for an H.264 (AVC) stream, the value of max_dec_frame_buffering is inferred as specified in [21].

init_pre_dec_buf_period: gives the required delay between the time of arrival in the pre-decoder buffer of the first bit of the first access unit and the time of removal from the pre-decoder buffer of the first access unit. It is in units of a 90 kHz clock. The bitstream is constrained by the value of the nominal removal time of the first access unit from the coded picture buffer (CPB), $t_{r,n}(0)$, equal to init_pre_dec_buf_period as specified in [21].

init_post_dec_buf_period: gives the required delay between the time of arrival in the post-decoder buffer of the first decoded picture and the time of output from the post-decoder buffer of the first decoded picture. It is in units of a 90 kHz clock. The bitstream is constrained by the value of dpb_output_delay for the first decoded picture in output order equal to init_post_dec_buf_period as specified in [21] assuming that the clock tick variable, t_c, is equal to 1 / 90 000.