**3GPP TSG SA WG4#109-e meeting *S4-200969***

**20th May – 3rd June 2020**

|  |
| --- |
| *CR-Form-v12.0* |
| **CHANGE REQUEST** |
|  |
|  | **26.244** | **CR** | **0065** | **rev** | **-** | **Current version:** | **15.0.0** |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **X** | Radio Access Network |  | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  | Removing MPEG-4 Visual from 3G FF |
|  |  |
| ***Source to WG:*** | Qualcomm Incorporated |
| ***Source to TSG:*** | SA4 |
|  |  |
| ***Work item code:*** | RM\_H263\_MP4V |  | ***Date:*** | 2020-06-11 |
|  |  |  |  |  |
| ***Category:*** | **C** |  | ***Release:*** | Rel-16 |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-8 (Release 8)Rel-9 (Release 9)Rel-10 (Release 10)Rel-11 (Release 11)Rel-12 (Release 12)**Rel-13 (Release 13)Rel-14 (Release 14)Rel-15 (Release 15)Rel-16 (Release 16)* |
|  |  |
| ***Reason for change:*** | H.263 was a state-of-the art codec in the last millennium and made mobile video possible and an actual reality. Many 3GPP specs adopted H.263 and H.263 was the format of choice for the first mobile video deployments. For a brief history, please check https://en.wikipedia.org/wiki/H.263. However, more than 20 years later, this format has finally done its duty and 3GPP should feel good about sending this codec to retirement as part of their Rel-16 specs.In 2012 (Rel-11), 3GPP already addressed to change the status of H.263 and MPEG-4 Video in several specifications, but did not fully remove the technology for all services.The same applies for MPEG-4 Visual. Hence, all unnecessary or confusing information on those two codecs relating to 3GPP Service Specifications is removed.However, the TS26.244 serves as the unique reference for the H.263 in the MP4 File format according to the MP4RA. Hence, maintaining H.263 in TS26.244 is essential. This aspect is clarified in the introduction. |
|  |  |
| ***Summary of change:*** | Remove reference for MPEG-4 VisualClarify the importance of TS26.244 to be the interop spec for H.263. |
|  |  |
| ***Consequences if not approved:*** | Unnecessary costs for testing and implementation.Unclear importance of TS26.244 for H.263. |
|  |  |
| ***Clauses affected:*** | 4, 5.2.1, 5.2.2, 6.1, 6.2, 6.3, 6.6, 7.7, 9.1, 10.2, A.2.2 |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **X** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **X** |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  | **X** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |  |

**===== CHANGE =====**

# 4 Overview

The 3GPP file format (3GP) is defined in this specification as an instance of the ISO base media file format [7]. 3GP is mandated in [8] to be used for continuous media along the entire delivery chain envisaged by the MMS, independent of whether the final delivery is done by streaming or download, thus enhancing interoperability.

In particular, the following stages are considered:

- upload from the originating terminal to the MMS proxy;

- file exchange between MMS servers;

- transfer of the media content to the receiving terminal, either by file download, by streaming or MBMS download delivery [40]. In the first and last case the self-contained file is transferred, whereas in the second case, for RTP streaming, the content is extracted from the file and streamed according to open payload formats. In this case, no trace of the file format remains in the content that goes on the wire/in the air. In segmented streaming over DASH [49], the file is divided into segments for transfer.

For the PSS, the 3GPP file format is mandated in [3] to be used for timed text and it should be supported by PSS servers; 3GP files with streaming-server extensions should be used for storage in streaming servers and the "hint track" mechanism should be used for the preparation for streaming. For Adaptive HTTP Streaming, HTTP streaming extensions are defined.

This specification also defines the necessary structure for integration over several codecs into the ISO base media file format [7]. In particular, this specification defines:

* the necessary structure for integration of 3GPP defined codecs such as AMR, AMR-WB, Enhanced aacPlus and EVS in clauses 6.
* the necessary structure for integration of H.263 video [9] in clause 6.6.

**===== CHANGE =====**

### 5.2.1 Limitations to the ISO base media file format

The following limitation to the ISO base media file format [7] shall apply to a 3GP file:

- compact sample sizes ('stz2') shall not be used for tracks containing H.263, AMR, AMR-WB, AAC or Timed text.

NOTE: The extended presentation format (see clause 11) is defined by using the Meta box of the ISO base media file format [7] that was not present in the first edition. Hence, extended presentations in 3GP files are explicitly signalled via the Extended-presentation profile (see clause 5.4.6).

**===== CHANGE =====**

### 5.2.2 Registration of codecs

Code streams for H.263 video [9], H.264 (AVC) video [29], H.265 (HEVC) video [51], AMR narrow-band speech [11], AMR wide-band speech [12], Extended AMR wide-band audio [21], EVS [55], Enhanced aacPlus audio [23, 24, 25], MPEG-4 AAC audio [13], and timed text [4] can be included in 3GP files as described in clause 6 of the present document.

**===== CHANGE =====**

## 6.1 General

The purpose of this clause is to define the necessary structure for integration of the H.263, AMR, AMR-WB, Extended AMR-WB (AMR-WB+), EVS, Enhanced aacPlus 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 clause 6.4 about the MP4AudioSampleEntry box in the MPEG-4 file format [14]. The definitions of the Sample Entry boxes for AMR, AMR-WB, AMR-WB+ and H.263 are given in clauses 6.5 to 6.10. The definition of the Sample Entry box for EVS is given in clause 6.14. The integration of timed text in a 3GP file is specified in [4], the integration of H.264 (AVC) is specified in [20], the integration of H.265 (HEVC) is specified in clause 8 of [20], the integration of Quality metrics timed metadata track is specified in clause 4 of [53] and clause 16 of this specification and the integration of DIMS is specified in [36] and clauses 5.4.3, 5.4.6 and 11 of the present document. Requirements for integrating video codecs in the context of the TV Video Profile are documented in TS 26.116 [56].

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. The 3GPP file format is the native storage format for AMR-WB+. The data stream, stored in samples of a 3GP file, shall be formatted according to clause 8.3 of [21]. Each sample contains one or more AMR-WB+ storage units. The number of storage units per sample may differ from sample to sample.

For EVS each sample of the media is one speech frame block as specified in Annex A.2.6.2 of [55]. A speech frame block consists of N ToC entries and N speech frames, where N is the value of channelcount in the EVSSampleEntry box specified in clause 6.14 of the present document.

**===== CHANGE =====**

## 6.2 Sample Description box

In an ISO file, Sample Description Box gives detailed information about the coding type used, and any initialisation information needed for that coding. The Sample Description Box can be found in the ISO file format Box Structure Hierarchy shown in figure 6.1.



Figure 6.1: ISO File Format Box Structure Hierarchy

The Sample Description Box can have one or more Sample Entries.

Valid Sample Entries already defined for ISO and MP4 include MP4AudioSampleEntry and HintSampleEntry. Other Sample Entries shall be according to the following:

- AMR, AMR-WB AMRSampleEntry

- AMR-WB+ AMRWPSampleEntry

- EVS EVSSampleEntry

- H.263 H263SampleEntry

- H.264(AVC) AVCSampleEntry

- H.265(HEVC) HEVCSampleEntry

- Timed text TextSampleEntry

- DIMS DIMSSampleEntry

- CVO timed metadata CVOSampleEntry

- Location timed metadata LocationSampleEntry

- Quality metrics timed metadata QualityMetricsSampleEntry

- Orientation timed metadata OrientationSampleEntry

The format of SampleEntry and its fields are explained as follows:

**SampleEntry ::= MP4AudioSampleEntry |
 AMRSampleEntry |
 AMRWPSampleEntry |**

 **EVSSampleEntry |
 H263SampleEntry |
 AVCSampleEntry |
 TextSampleEntry |
 DIMSSampleEntry |
 HintSampleEntry |
 CVOSampleEntry |
 HEVCSampleEntry |
 LocationSampleEntry |**

 **QualityMetricsSampleEntry |**

 **OrientationSampleEntry**

Table 6.1: SampleEntry fields

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Details | Value |
|  |  |  |  |
| MP4AudioSampleEntry |  | Entry type for audio samples defined in the MP4 specification. |  |
| AMRSampleEntry  |  | Entry type for AMR and AMR-WB speech samples defined in clause 6.5 of the present document. |  |
| AMRWPSampleEntry |  | Entry type for AMR-WB+ audio samples defined in clause 6.9 of the present document. |  |
| EVSSampleEntry |  | Entry type for EVS samples defined in clause 6.14 of the present document. |  |
| H263SampleEntry  |  | Entry type for H.263 visual samples defined in clause 6.6 of the present document. |  |
| AVCSampleEntry |  | Entry type for H.264 (AVC) visual samples defined in the AVC file format specification in clause 5 of [20]. |  |
| TextSampleEntry |  | Entry type for timed text samples defined in the timed text specification |  |
| DIMSSampleEntry |  | Entry type for DIMS scene description samples defined in the DIMS specification. |  |
| HintSampleEntry |  | Entry type for hint track samples defined in the ISO specification. |  |
| CVOSampleEntry |  | Entry type for CVO timed metadata track as defined in clause 6.11 of the present document |  |
| HEVCSampleEntry |  | Entry type for H.265 (HEVC) visual samples defined in the H.265 (HEVC) file format specification in clause 8 of [20]. |  |
| LocationSampleEntry |  | Entry type for Location timed metadata track as defined in clause 6.12 of the present document |  |
| QualityMetricsSampleEntry |  | Entry type for Quality metrics timed metadata track as defined in clause 4 of [53] |  |
| OrientationSampleEntry |  | Entry type for Orientation timed metadata track as defined in clause 6.13 of the present document |  |

From the Sample Entries in Table 6.1, only the MP4AudioSampleEntry, H263SampleEntry, AMRSampleEntry, AMRWPSampleEntry, EVSSampleEntry, CVOSampleEntry, LocationSampleEntry and OrientationSampleEntry are taken into consideration here. TextSampleEntry is defined in [4], HintSampleEntry in [7], AVCSampleEntry in clause 5 of [20], HEVCSampleEntry in clause 8 of [20], QualityMetricsSampleEntry in clause 4 of [53] and DIMSSampleEntry in [36].

**===== CHANGE =====**

## 6.3 (void)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**===== CHANGE =====**

## 7.7 Aggregated RTP payloads

An application data unit (ADU), normally being the smallest independently usable data unit, is specified as follows for coding formats and RTP payload formats allowed in 3GP files:

- For audio and speech, an ADU is specified as a coded frame intended for transport.

- For H.263 an ADU consists of an entire RTP payload.

- For H.264 (AVC) or H.265 (HEVC), an ADU is a Network Adaptation Layer Unit (NALU).

- For timed text, an ADU consists of any of the type 1-5 RTP payload units [28].

For encrypted RTP payloads, the actual ADUs are hidden within the encrypted payload. Some RTP payload formats allow aggregation of multiple ADUs into a single RTP payload. When any hint sample in an RTP hint track defines a payload including multiple ADUs, each hint sample in the hint track shall comply with the following requirements:

- The extra-flag in the RTPPacket class of the hint sample shall be set to 1. This indicates that there is extra information before the RTP constructors in the form of type-length-value sets.

- The extra information in the hint sample shall include a '3gau' structure as specified below.

class 3gppApplicationDataUnitInfoTLV extends Box('3gau') {
 unsigned int(16) entrycount;
 for(i=1; i<=entrycount; i++){
 unsigned int(32) numbytes;
 unsigned int(64) decorder;
 unsigned int(32) timestampoffset
 }
}

**entrycount** indicates the number of ADUs in the RTP payload.

**numbytes** indicates the number of bytes of the i'th ADU in the RTP payload.

**decorder** indicates the decoding order of ADUs within the RTP hint track. The smaller value of decorder, the earlier the ADU is in decoding order. All ADUs shall have a unique value of decorder, and the assignment shall be done using consecutive numbers. If two or more ADUs can be decoded virtually simultaneously, i.e. their relative decoding order is undefined, they shall still be assigned consecutive numbers.

**timestampoffset** indicates the RTP timestamp offset of the i'th ADU relative to the timestamp of RTP header of the packet it will be transmitted in. Where the ADU's timestamp value is equal to what it would have had if it were transmitted in an RTP packet containing only the ADU.

**===== CHANGE =====**

## 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.264 (AVC) or H.265 (HEVC), there are two types of buffers:

- Hypothetical Reference Decoder (HRD) model;

- For H.264 (AVC) the de-interleaving buffer of the interleaved RTP packetization mode as specified in [30], or for H.265 (HEVC), the de-packetization buffer as specified in [52].

Buffer parameters for several operation points and synchronization points of the HRD model may be specified by a video HRD sample grouping as defined in clause 9.2.2.

Only one set of de-interleaving parameters for H.264 (AVC) and only one set of de-packetization parameters for H.265 (HEVC) can be associated to a stream and therefore the de-interleaving or de-packetization parameters are included in the corresponding media-level SDP fragment according to the H.264 (AVC) MIME/SDP specification in [30] or the H.265 (HEVC) MIME/SDP specification in [52].

NOTE: Any HRD parameters in parameter sets and SEI message in the bitstream or included in the MIME/SDP parameters of a media-level SDP fragment must not contradict each other or the information in the video HRD sample grouping, if any.

**===== CHANGE =====**

## 10.2 Sample entries for encrypted media tracks

The sample entries stored in the sample description box of a media track in a 3GP file identify the format of the encoded media, i.e. codec and other coding parameters. All valid sample entries for unencrypted media in a 3GP file are described in Clause 6. The principle behind storing encrypted media in a track is to "disguise" the original sample entry with a generic sample entry for encrypted media. Table 10.1 gives an overview of the formats (identifying sample entries) that can be used in 3GP files for signalling encrypted video, audio and text.

Table 10.1: Formats for encrypted media tracks

|  |  |  |
| --- | --- | --- |
| Format | Original format | Media content |
| 'encv' | 's263', 'avc1', … | encrypted video: H.263, H.264(AVC), … |
| 'enca' | 'samr', 'sawb', 'sawp', 'mp4a', … | encrypted audio: AMR, AMR-WB, AMR-WB+, Enhanced aacPlus, AAC, … |
| 'enct' | 'tx3g', … | encrypted text: timed text, … |

The generic sample entries for encrypted media replicate the original sample entries and include a Protection scheme information box with details on the original format, as well as all requirements for decrypting the encoded media. The EncryptedVideoSampleEntry and the EncryptedAudioSampleEntry are defined in Tables 10.2 and 10.3, where the ProtectionSchemeInfoBox (defined in clause 10.2) is simply added to the list of boxes contained in a sample entry.

Table 10.2: EncryptedVideoSampleEntry

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Details | Value |
| **BoxHeader**.Size | Unsigned int(32) |  |  |
| **BoxHeader**.Type | Unsigned int(32) |  | 'encv' |
| All fields and boxes of a visual sample entry H263SampleEntry. |
| **ProtectionSchemeInfoBox** |  | Box with information on the original format and encryption |  |

Table 10.3: EncryptedAudioSampleEntry

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Details | Value |
| **BoxHeader**.Size | Unsigned int(32) |  |  |
| **BoxHeader**.Type | Unsigned int(32) |  | 'enca' |
| All fields and boxes in an audio sample entry, e.g. MP4AudioSampleEntry or AMRSampleEntry. |
| **ProtectionSchemeInfoBox** |  | Box with information on the original format and encryption |  |

The EncryptedVideoSampleEntry and the EncryptedAudioSampleEntry can also be used with any additional codecs added to the 3GP file format, as long as their sample entries are based on the SampleEntry of the ISO base media file format [7].

The EncryptedTextSampleEntry is defined in Table 10.4. Text tracks are specific to 3GP files and defined by the Timed text format [4]. In analogy with the cases for audio and video, a ProtectionSchemeInfoBox is added to the list of contained boxes.

Table 10.4: EncryptedTextSampleEntry

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Details | Value |
| **BoxHeader**.Size | Unsigned int(32) |  |  |
| **BoxHeader**.Type | Unsigned int(32) |  | 'enct' |
| All fields and boxes of TextSampleEntry. |
| **ProtectionSchemeInfoBox** |  | Box with information on the original format and encryption |  |

NOTE: The boxes within the sample entries defined in Tables 10.2-10.4 may not precede any of the fields. The order of the boxes (including the ProtectionSchemeInfoBox) is not important though.

**===== CHANGE =====**

## A.2.2 Codecs parameter

The codecs parameter is defined in RFC 6381. The ISO file format name space and ISO syntax in clauses 3.3 and 3.4 of RFC 6381 [32] shall be used together with extensions to the ISO syntax specified here.

The syntax in clause 3.4 of RFC 6381 defines the usage of the codecs parameter for files based on the ISO base media file format and specifies that the first element of a parameter value is a sample description entry four-character code. It also includes specific definitions for MPEG audio ('mp4a') where each value in addition to the four-character code includes two elements signalling Object Type Indications and Profile Level Indications (video only). It also includes specific definitions for Advanced Video Coding ('avc1') where each value in addition to the four-character code includes a second element (referred to as 'avcoti' in the formal syntax), which is the hexadecimal representation of the following three bytes in the (subset) sequence parameter set Network Abstraction Layer (NAL) unit specified in [29]: (1) profile\_idc,(2) the byte containing the constraint\_set flags (currently constraint\_set0\_flag through constraint\_set5\_flag, and the reserved\_zero\_2bits), and (3) level\_idc. Note also that reserved\_zero\_2bits is required to be equal to 0 in [29], but other values for it may be specified in the future by ITU-T or ISO/IEC. These definitions apply to the MPEG codecs used by the 3GP file format, such as H.264 (AVC) [29], MPEG-4 AAC [13] and Enhanced aacPlus [23, 24, 25]. Values for other codecs used by the 3GP file format are specified below.

When the first element of a value is 's263', indicating H.263 video [9], the second element is the decimal representation of the profile, e.g., 0 or 3, and the third element is the decimal representation of the level, e.g. 10 or 45.

When the first element of a value is one of the following elements, no other elements are defined for that value:

- 'samr', indicating AMR narrow-band speech [11];

- 'sawb', indicating AMR wide-band speech [12];

- 'sawp', indicating Extended AMR wide-band audio [21];

- 'sevs', indicating EVS speech [55];

- 'tx3g', indicating timed text [4];

- '3gvo', indicating CVO as defined in clauses 6.11 and 14.

The following syntax defines all values above in ABNF (RFC 4234 [31]) by extending the definition in clause 3.4 of RFC 6381:

id-iso = iso-gen / iso-mpega / iso-mpegv / iso-amr / iso-amr-wb / iso-amr-wbp / iso-evs / iso-tt / iso-h263; = iso-gen, iso-mepga, iso-mpegv, iso-avc as defined in RFC 6381

iso-amr = %x73.61.6d.72 ; 'samr'

iso-amr-wb = %x73.61.77.62 ; 'sawb'

iso-amr-wbp = %x73.61.6d.70 ; 'sawp'

iso-evs = %x73.65.76.73 ; 'sevs'

iso-tt = %x74.78.33.67 ; 'tx3g'

iso-cvo = %x33.67.76.6f ; '3gvo'

iso-h263 = s263 "." h263-profile "." h263-level

s263 = %x73.32.36.33 ; 's263'

h263-profile = 1\*DIGIT

h263-level = 1\*DIGIT

The elements of the codecs parameter for H.265 (HEVC) are specified as below.

NOTE: The following specification replaces that in clause E.3 of [20].

When the first element of a value is a code indicating a codec from the High Efficiency Video Coding specification (ISO/IEC 23008-2), as documented in clause 8 of [20] (such as 'hev1' or 'hvc1'), the elements following are a series of values from the HEVC decoder configuration record, separated by period characters ("."). In all numeric encodings, leading zeroes may be omitted,

- the general\_profile\_space, encoded as no character (general\_profile\_space == 0), or 'A', 'B', 'C' for general\_profile\_space 1, 2, 3, followed by the general\_profile\_idc encoded as a decimal number;

- the 32 bits of the general\_profile\_compatibility\_flags, but in reverse bit order, i.e. with general\_profile\_compatibility\_flag[ 31 ] as the most significant bit, followed by , general\_profile\_compatibility\_flag[ 30 ], and down to general\_profile\_compatibility\_flag[ 0 ] as the least significant bit, where general\_profile\_compatibility\_flag[ i ] for i in the range of 0 to 31, inclusive, are specified in ISO/IEC 23008-2, encoded in hexadecimal (leading zeroes may be omitted);

- the general\_tier\_flag, encoded as 'L' (general\_tier\_flag==0) or 'H' (general\_tier\_flag==1), followed by the general\_level\_idc, encoded as a decimal number;

- each of the 6 bytes of the constraint flags, starting from the byte containing the general\_progressive\_source\_flag, each encoded as a hexadecimal number, and the encoding of each byte separated by a period; trailing bytes that are zero may be omitted.

Examples:

codecs=hev1.1.6.L93.B0

a progressive, non-packed stream, Main Profile, Main Tier, Level 3.1. (Only one byte of the constraint flags is given here; The value after the second period is 6 instead of 2 because according to ISO/IEC 23008-2 a Main Profile bitstream should also be marked as compatible to the Main 10 Profile).

codecs=hev1.A4.41.H120.B0.23

a (mythical) progressive, non-packed stream in profile space 1, with general\_profile\_idc 4, some compatibility flags set, and in High tier at Level 4 and two bytes of constraint flags supplied.

For Quality metrics metadata track, the usage of the codecs parameter is specified in [53].

Table 10.6: OriginalFormatBox

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Details | Value |
| **BoxHeader**.Size | Unsigned int(32) |  |  |
| **BoxHeader**.Type | Unsigned int(32) |  | 'frma' |
| DataFormat | Unsigned int(32) | original format |  |

**DataFormat** identifies the format (sample entry) of the decrypted, encoded data. The currently defined formats in 3GP files include 'h263', 'avc1', 'hvc1', 'hev1', 'mp4a', 'samr', 'sawb', 'sawp' and 'tx3g'.