

**INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2016/m39263
October 2016, Chengdu (CH)**

Source **Telecom ParisTech**
Status **For consideration at the 116th MPEG meeting**
Title **Comments on MMT Specification**
Author Jean-Claude Dufourd, Nicolas Derouineau, Cyril Concolato, Jean Le Feuvre

1 Introduction

As part of the GPAC project (<http://gpac.io>), we have started implementing basic MMT support. This document contains comments and suggestions following that implementation.

2 Comments and suggestions

2.1 Mandatory Moov

The following sentence: "An MPU file should contain an `'ftyp'`, an `'mmpu'`, and a `'moov'`, uses "should", which means that some MPU may not have a 'moov' at all.

But the standard also says: "An MPU shall be a conformant ISOBMFF file" which requires a "moov".

These sentences are contradictory. We suggest clarifying this aspect.

2.2 Optionality of mmpu box

The mmpu box is optional in an MPU. It is unclear how the asset identifier of an MPU without mmpu box can be found, in particular when used in PI.

We suggest clarifying this point.

2.3 Number of tracks

There are several inconsistencies regarding the number of media tracks in an MPU.

Section 6.2 declares "The `'moov'` box shall contain at most one media track" and a sentence in 6.1 says "Timed media data are stored as track of the ISOBMFF (a single media track is allowed)."

Additionally, it says:

"The maximum number of independent (e.g. empty `'tref'` box) media tracks in this file shall be one. Additionally, tracks with non-empty `'tref'` box (e.g. hint track) may be available. "

This seems to imply that a file may have 2 media tracks, one dependent on the other. This is contradictory with 6.2.

Then, independent tracks do not necessarily have an empty 'tref' box (e.g. a metadata track (independently decodable) may have a reference to another independent track). A track with a non-empty 'tref' box is not necessarily dependent on another track for decoding.

We suggest clarifying this text, in particular defining dependency in terms of decoding and not tref.

2.4 On SAP in MPU

Section 5.4 states that “the first access unit of an MPU shall be a SAP for processing by an MMT entity”.

This definition does not indicate which type of SAP is allowed, which means that all types would be allowed, including SAP of type 1 or 2 (Closed GoP Access Point), or SAP of type 3 (Open GoP dependency) and others.

Additionally, the sentence "For timed media, this implies that the decoding order of the first AU in the MPU payload is always '0'." is unclear:

- the "first AU in the MPU": Is it the "first" in bitstream order, in which case it will be necessarily the first to be decoded?
- Should it say "the first AU in presentation order is also the first in decoding order", which would require only SAP of type 1?

The following is suggested:

- A clarification of the actual type of SAP supported in MPU should be made. If Open GoP is supported, the authors suggest the possibility of profiles creation (Broadcast for Open-GoP, and Adaptive Streaming for Closed GoP) and associated signaling (e.g. brands).
- The notion of decoding number (as used above: '0') is not defined in ISO/IEC 14496-12. Only the decoding order is (according to the DTS). The sentence indicated above should be fixed.

2.5 Assets dependency

In section 3.3.1, a dependent asset is defined as "an asset for which one or more assets are necessary for decoding the contained media content".

But an asset is made of MPUs, which are defined as independent:

"3.1.14

Media Processing Unit

MPU

generic container for independently decodable timed or non-timed data that is media codec agnostic"

We suggest clarifying the notion of dependency at MPU level and asset level, in particular if it is meant only for only for scalable representations? or also for Open GoP dependencies?

2.6 MPU Metadata

Section 6.2 says: “An MPU file should contain an 'ftyp', an 'mmpu', and a 'moov', and optionally an 'sidc' box, all of which are part of the MPU metadata.”

In section 7.1, it says: "MPU metadata may include the 'ftyp', 'sidc', 'mmpu', and 'moov' boxes."

In section 8.3.2.3, Table 4, "MPU metadata" is defined as: "contains the ftyp, mmpu, moov, and meta boxes as well as any other boxes that appear in between."

The above definitions are almost the same but not quite the same ('sidx' box, 'meta' box, optional boxes). Additionally, it does not indicate that moov and mdat are not part of this metadata although they could appear before between moov and meta.

We suggest formally defining it in section 3 and that all redefinitions are removed.

2.7 MPU Concatenation

In section 5.4, the following statement is done: "In the absence of an 'sidx' box, the concatenation of the j -th MPU of Asset with `asset_id Y` with the j -th MPU of the Asset with `asset_id X` without its MPU metadata results in a valid MPU"

First, concatenation of A with B may mean A+B or B+A. It would be good to clarify that part, using terms such as "The concatenation of A followed by B".

Then, the concatenation of MPUs without MPU metadata may result in invalid bitstreams, unless some constraints are met. For instance, in the case of layered representations, one can store the base layer in one track in the MPUs of an asset Y, and the other layer in the MPUs of a dependent asset X. For the concatenation to be valid, the MPU metadata of Y has to declare the 2 tracks (see discussion above on the number of media tracks).

We suggest clarifying this aspect.

2.8 Media interval and media interval concatenation

The specification uses the notion of "media intervals" in unclear ways. It should probably use the term "sub-segment" as defined in ISO/IEC 14496-12.

Then, the following sentence should talk about Assets X and Y:

"Additionally, if the 'sidx' box is present the media intervals defined by the 'sidx' box shall be non-overlapping, i.e. no media sample in the k -th media interval (defined by the 'sidx' box) in an MPU is inside the time interval defined by the sample boundaries of the j -th media time interval (defined by the 'sidx' box) for j different from k ."

Within a single asset, it is obvious that the 'sidx' does not define overlapping media intervals. Should the text above mention Asset X and Y?

Also, the following sentence is also unclear:

"When an 'sidx' box is present the concatenation of the k -th media interval (defined by the 'sidx' box) of the j -th MPU of Asset with `asset_id Y` with the k -th media interval (defined by the 'sidx' box) of the j -th MPU of the Asset with `asset_id X` following the metadata of the MPU with `asset_id Y` results in a valid MPU."

What is the concatenation of a "media interval"? Is it the concatenation of the associated movie fragments?

2.9 MPU Box

2.9.1 Syntax problem

In section 6.3.2, The following syntax is described for the MPU Box:

```
aligned(8) class MPUBox
    extends FullBox('mmpu', version, 0){

    unsigned int(1) is_complete;
    unsigned int(1) is_adc_present;
    unsigned int(6) reserved;

    unsigned int(32) mpu_sequence_number;

    AssetIdentifierBox();
}
```

```
aligned(8) class AssetIdentifierBox {
    unsigned int(32) asset_id_scheme;
    unsigned int(32) asset_id_length;
    unsigned int(8)  asset_id_value[asset_id_length];
}
```

There is a lack of consistency in this definition. The AssetIdentifierBox doesn't extend from an IsoBMFF Box Type.

The following is suggested:

- Case 1

```
aligned(8) class MPUBox
    extends FullBox('mmpu', version, 0){

    unsigned int(1) is_complete;
    unsigned int(1) is_adc_present;
    unsigned int(6) reserved;

    unsigned int(32) mpu_sequence_number;

    AssetIdentifierBox();
}
```

could be replaced by

```
aligned(8) class MPUBox
    extends FullBox('mmpu', version, 0){

    unsigned int(1) is_complete;
    unsigned int(1) is_adc_present;
    unsigned int(6) reserved;

    unsigned int(32) mpu_sequence_number;

    unsigned int(32) asset_id_scheme;
    unsigned int(32) asset_id_length;
    unsigned int(8)  asset_id_value[asset_id_length];
}
```

- Case 2

```
aligned(8) class AssetIdentifierBox {
    unsigned int(32) asset_id_scheme;
    unsigned int(32) asset_id_length;
    unsigned int(8)  asset_id_value[asset_id_length];
}
```

could be replaced by

```
aligned(8) class AssetIdentifierRecord {
    unsigned int(32) asset_id_scheme;
    unsigned int(32) asset_id_length;
    unsigned int(8)  asset_id_value[asset_id_length];
}
```

2.9.2 Semantics

2.9.2.1 is_complete

When an MPU is generated from an MP4, it is unclear what the value of 'is_complete' should be, since no MFU is generated. Additionally, 'is_complete' is not used elsewhere in the specification. We suggest clarifying this point.

2.9.2.2 asset_id_scheme

asset_id_scheme is a 32 bits value. "URI" is made of 3 chars. It is unclear what the value of the 4 bytes should be (can we use 0 or should we use printable characters such as space) and where each byte should be such as: (0, 'U', 'R', 'I') or ('U', 'R', 'I', 0) or (20, 'U', 'R', 'I') or ('U', 'R', 'I', 20)

Additionally, the meaning of the following sentence is unclear:

"It is recommended to use scheme-length-value, and not to define a new identification scheme."

" scheme-length-value " is used only here. We suggest removing it.

2.10 Track Fragment Decode Time

It is unclear if the value of the tfdt of the first movie fragment of an MPU is allowed to have a value different from 0, for example in the 2nd MPU of an Asset.

A similar problem is the alignment of an audio MPU with a video MPU. In this case, it is unlikely that both MPU start with a tfdt of 0.

The use of tfdt should be clarified.

2.11 Movie Fragment Sequence Number

The text does not indicate how movie fragment sequence numbers are handled. Should they be monotonically increasing within an MPU and across MPUs?

2.12 Typos and editorial problems

"support the other brand" should be "support for other brands"

"move fragment" should be " movie fragment"

The following sentence is duplicated (5.4 and 6.2):

"If any 'elst' box is available, the indicated offset shall be applied to the composition time of the first sample in presentation order of the MPU in addition to the presentation time provided by any presentation information."

3 Recommendation

We recommend MPEG to issue a corrigendum on MMT with the above clarifications.

4 References

[1] m37148, Draft for ISO/IEC 23008-1 2nd Edition