3GPP TSG SA WG4 128 S4-241302

Jeju, Korea, 20 - 24 May, 2024

**Title: Draft Reply LS on MV-HEVC Integration and Media Messaging Application Format**

**Response to: -**

**Release: Rel-19**

**Work Item: VOPS**

**Source:** **3GPP SA4**

**To:** **ISO/IEC JTC 1/SC 29/WG 03 (MPEG Systems)**

**Cc:**

**Contact person: Waqar Zia**

 **waqar\_zia (at) apple.com**

**Send any reply LS to: 3GPP Liaisons Coordinator,** **mailto:3GPPLiaison@etsi.org**

**Attachments: None**

# 1 Overall description

3GPP TSG SA WG4 (SA4) would like to thank MPEG Systems for Reply Liaison statement to 3GPP SA4 on MV-HEVC Integration and Messaging Media Application Format. On the topic of ongoing 3GPP SA4 work of MV-HEVC integration work VOPS (Video Operating Points - Harmonization and Stereo MV-HEVC), SA4 would like to exchange information on two topics in the following.

## 1.A. Codec Strings

The first topic is related to codec strings already included in MPEG Systems' reply LS. SA4 would like to provide further information on this topic to further clarify the issue.

One of the possibilities to carry layered HEVC (L-HEVC) video in mp4 is by using the 'hvc1' or 'hev1' sample entry type as specified in clause 9 of ISO/IEC 14496-15 in a backwards compatible manner. Based on discussions in SA4, it has become evident that existing products in the market are using this concept to carry stereoscopic content and alpha using the Multiview extensions of the HEVC standard as L-HEVC in mp4. Such profiles with existing support in the mobile ecosystem are targeted for the ongoing SA4 VOPS work item. However, when constructing the MIME types 'codecs' parameter, according to Annex E of ISO/IEC 14496-15, it does not provide necessary signalling for all layers. Even if additional information can be specified with other MIME type parameters, these are often not processed by the APIs. For example, W3C API accepts a MIME type with no extra MIME parameters except for codecs.

Furthermore, the signalling inside the codecs string does not expose other important information such as the types of auxiliary information that would allow us to obtain necessary information about rendering aspects of a stream. Given these challenges, it is important that a standardized solution be developed to enhance interoperability, accuracy, and efficiency of multi-layer video stream handling. SA4 has identified some high-level requirements for such a solution:

* **Comprehensive Layer Signalling**: Enable the signalling of multiple video layers to be used for 3GPP-based services, including DASH MPDs (see TS 26.511), capability checks in the context of 5G Media Streaming (see TS 26.511), as well as for Media Messaging Services (see TS 26.143) . This should include but not be limited to the number of layers, types of each layer, their inter-dependencies, etc.
* **Backward Compatibility**: Ensure that the solution maintains backward compatibility, allowing existing players and systems to continue functioning without modifications while enabling enhanced capabilities for updated systems.
* **Adaptability and Extensibility**: Design the solution to be adaptable for future extensions and new types of layers or enhancements without requiring significant overhauls. Consider making the signalling codec agnostic.

## 1.B. Carriage of metadata in CMAF

The current version of ISO/IEC 23000-19:2024 (CMAF) allows for only a single track carrying media data to be present in the MovieBox, as noted in section 7.3.2.1 of CMAF specification. This section mentions that timed metadata tracks can be provided as separate CMAF tracks in a distinct selection set. This restriction precludes the packaging of metadata tracks alongside media data within a single CMAF track.

Using separate switching sets in CMAF to provide necessary metadata for such use cases is impractical, particularly given the volume of data involved. There are two primary purposes of carrying the metadata in a CMAF track: 1) the metadata annotating the media samples in the track, i.e. self-contained metadata for the playback of the track, and 2) the metadata providing the relevant information for dynamic switching to/from other tracks in the CMAF switching sets during the playback. In neither of the cases, is the metadata subject to late binding and maintains a clear one-to-one dependency with the video content. Thus, such metadata is not selected or need to be “late-binded” during the playback. Carrying such information in a separate track requires the player to fetch the corresponding metadata track, which can be resource-intensive and complicate player logic.

We believe this limitation is unnecessary for use cases where certain media data is closely associated with timed metadata. Adding frame accurate metadata to CMAF tracks is a **generic problem** **and** **not only a problem that occurs in layered video use-cases**. For the layered video use-cases, there are already existing products on the market that already combine media and metadata tracks in the same segments. For instance, based on discussions in SA4 it has become evident that existing products employ the 'mebx' metadata track together with the stereo video track to mitigate user discomfort of caption parallax when captions intersect with stereoscopic elements of the video and these payloads also appear together in fragmented ISOBMFF files.

SA4 is also aware that other methods exist to embed metadata within media data samples (e.g., SEI, T.35), event message box and sample groups. However, metadata tracks offer significant advantages for dynamic metadata that is synchronized and closely associated with media data. Additionally, ISO/IEC 14496-12:2022 (ISOBMFF) define multiplexed metadata track format ('mebx') capable of carrying multiple metadata items over a time range that allows multiplexing different types of metadata and reduce the number of metadata tracks to a single metadata track.

SA4 has identified the following high-level requirements for consideration by MEPG to suitably extend CMAF specification to support CMAF tracks that:

* can multiplex one media type track with its associated small size timed aligned metadata, with the possibility of multiplexing at different granularity level.
* can be identified at the track level that is a multiplexed track, and
* under what constraints can they be included together in a CMAF switching set.

# 2 Actions

**To ISO/IEC JTC 1/SC 29/WG 03 (MPEG Systems)**

**ACTION:** MPEG Systems is requested to take the above information into account and kindly develop solutions to fulfil the proposed requirements on the two topics identified above.

# 3 Dates of next TSG SA WG 4 meetings

SA4#129 19-23 August 2024 Online

SA4#130 18-22 November 2024 Orlando, US