3GPP TSG-SA4 Meeting #112e *S4-210187*

1-10 February 2021 revision of S4-210025

**Agenda item** 11.5

**Source:** Intel

**Title:** On SEI Messages for ITT4RT

**Document for** Discussion and Agreement

# Introduction

A draft CR to TS 26.114 introducing video support for ITT4RT in S4-201538 was reviewed and agreed during the SA4#111e meeting.

According to the latest agreed draft CR, the exchange of SEI messages carrying VR metadata for rendering 360-degree video or fisheye video shall be performed using one of the following types of signalling:

*“1- Bitstream-level: SEI messages shall be present in the respective video elementary streams corresponding to the HEVC or AVC bitstreams carrying 360-degree video or fisheye video from the ITT4RT-Tx client to the ITT4RT-Rx client, as per ISO/IEC 23008-2 [119] or ISO/IEC 14496-10 [24]. As expressed more clearly below, the mandatory inclusion of the specific SEI messages in the bitstream by the ITT4RT-Tx client and their decoder and rendering processing by the ITT4RT-Rx client is conditional upon successful SDP-based negotiation of the corresponding 360-degree video or fisheye video capabilities.*

*2- SDP-level: SEI messages shall be signalled in the SDP using the ‘sprop-sei’ parameter based on the procedures specified in IETF RFC 7798 [120] or via other SDP-based means, in the corresponding SDP offer or answer from the ITT4RT-Tx client to the ITTRT-Rx client, during the session setup involving media negotiations for 360-degree video or fisheye video. It should be noted that the signalling based on the ‘sprop-sei’ parameter is only available for the HEVC-based RTP payload formats and is not supported in AVC-based RTP payload formats as defined in IETF RFC 6184 [25].”*

After a careful review of the above alternatives, the proponent is of the view that conducting this signaling in the SDP level is problematic due to the time-varying nature and in many cases frame-dependent behavior of the SEI information - in other words, an SDP-alone solution is not feasible. The SDP approach is only suitable when it is also supported by further signaling at the RTP/RTCP level, e.g. via the use of RTP header extensions, such that the dynamic information contained in the SEI messages can be signalled along with the RTP media. But according to the current ITT4RT timeplan, investigations to perform such work are planned for Phase 2. The proponent believes that until such a solution is defined, it is necessary to consider bitstream-level signaling as the default method for delivering SEI message information.

In this contribution, we propose to include some further text updates into the latest agreed draft CR in S4-201538 in order to clarify the above points. These are presented in Section 2.

# Proposed Text Updates to draft CR to TS 26.114

# X.3 Immersive 360-Degree Video Support

ITT4RT-Rx clients in terminals offering video communication shall support decoding capabilities based on:

- H.264 (AVC) [24] Constrained High Profile, Level 5.1 with the following additional restrictions and requirements on the bitstream:

- the maximum VCL Bit Rate is constrained to be 120 Mbps with cpbBrVclFactor and cpbBrNalFactor being fixed to be 1250 and 1500, respectively.

- the bitstream does not contain more than 10 slices per picture.

- H.265 (HEVC) [119] Main 10 Profile, Main Tier, Level 5.1.

In addition, ITT4RT-Rx clients in terminals may support:

- H.265 (HEVC) [R4] Screen-Extended Main 10 Profile, Main Tier, Level 5.1.

- H.265 (HEVC) [R4] Screen-Extended Main 4:4:4 10 Profile, Main Tier, Level 5.1.

ITT4RT-Tx clients in terminals offering video communication shall support encoding up to the maximum capabilities (e.g., color bit-depth, luma samples per second, luma picture size, frames per second) compatible with decoders compliant with the following on the bitstream:

- H.264 (AVC) [24] Constrained High Profile, Level 5.1 with the following additional restrictions and requirements:

- the maximum VCL Bit Rate is constrained to be 120 Mbps with cpbBrVclFactor and cpbBrNalFactor being fixed to be 1250 and 1500, respectively.

- the bitstream does not contain more than 10 slices per picture.

- H.265 (HEVC) [119] Main 10 Profile, Main Tier, Level 5.1.

In addition, ITT4RT-Tx clients in terminals may support:

- H.265 (HEVC) [R4] Screen-Extended Main 10 Profile, Main Tier, Level 5.1.

- H.265 (HEVC) [R4] Screen-Extended Main 4:4:4 10 Profile, Main Tier, Level 5.1.

Hence, for a Bitstream conforming to the H.264 (AVC) [24] Constrained High Profile, Level 5.1 delivered from an ITT4RT-Tx client to the ITT4RT-Rx client, the following restrictions apply:

- The profile\_idc shall be set to 100 indicating the High profile.

- The constraint\_set0\_flag, constraint\_set1\_flag, constraint\_set2\_flag and constraint\_set3\_flag shall all be set to 0, and constraint\_set4\_flag and constraint\_set5\_flag shall be set to 1.

- The value of level\_idc shall not be greater than 51 (corresponding to the level 5.1) and should indicate the lowest level to which the Bitstream conforms.

Furthermore, for a Bitstream conforming to the H.265 (HEVC) [119] Main 10 Profile, Main Tier, Level 5.1 delivered from an ITT4RT-Tx client to the ITT4RT-Rx client, the following restrictions apply:

- The general\_profile\_idc shall be set to 2 indicating the Main10 profile.

- The general\_tier\_flag shall be set to 0 indicating the Main tier.

- The value of level\_idc shall not be greater than 153 (corresponding to the Level 5.1) and should indicate the lowest level to which the Bitstream conforms.

For 360-degree video delivery across ITT4RT clients, the following components are applicable:

- The RTP stream shall contain an HEVC or an AVC bitstream with possible presence of omnidirectional video specific SEI messages. In particular, the omnidirectional video specific SEI messages as defined in clause D.2.41 of ISO/IEC 23008-2 [119] or ISO/IEC 14496-10 [24] may be present for the respective HEVC or AVC bitstreams.

- The video elementary stream(s) shall be encoded following the requirements in the Omnidirectional Media Format (OMAF) specification ISO/IEC 23090-2 [R1], clauses 10.1.2.2 (viewport-independent case) or 10.1.3.2 (viewport-dependent case) for HEVC bitstreams and clause 10.1.4.2 for AVC bitstreams. Furthermore, the general video codec requirements for AVC and HEVC in clause 5.2.2 of TS 26.114 also apply.

ITT4RT-Rx clients are expected to be able to process the VR metadata carried in SEI messages for rendering 360-degree video according to the relevant processes. Relevant SEI messages contained in the elementary stream(s) with decoder rendering metadata may include the following information for the relevant processes as per clause D.3.41 of ISO/IEC 23008-2 [119] and ISO/IEC 14496-10 [24]:

- Projection mapping information (indicating the projection format in use, e.g., Equirectangular projection (ERP) or Cubemap projection (CMP)), for the projection sample location remapping process as specified in clauses 7.5.1.3 and 5.2 of ISO/IEC 23090-2 [R1]

- Region-wise packing information (carrying region-wise packing format indication, any coverage restrictions or padding/guard region information in ithe packed picture), for the inverse processes of the region-wise packing as specified in clauses 7.5.1.2 and 5.4 of ISO/IEC 23090-2 [R1]

- Sphere rotation information (indicating the amount of sphere rotation, if any, applied to the sphere signal before projection and region-wise packing at the encoder side), for the coordinate axes conversion process as specified in clause 5.3 of ISO/IEC 23090-2 [R1]

- Frame packing arrangement (indicating the frame packing format for stereoscopic content), for the processes as specified in D.3.16 of ISO/IEC 23008-2 [119]

- Fisheye video information (indicating that the picture is a fisheye video picture containing a number of active areas captured by fisheye camera lens), for the fisheye sample location remapping process as specified in clause D.3.41.7.5 of ISO/IEC 23008-2 [119]

 The exchange of SEI messages carrying VR metadata for rendering 360-degree video or fisheye video shall be performed using one of the following types of signalling:

1- Bitstream-level: SEI messages shall be present in the respective video elementary streams corresponding to the HEVC or AVC bitstreams carrying 360-degree video or fisheye video from the ITT4RT-Tx client to the ITT4RT-Rx client, as per ISO/IEC 23008-2 [119] or ISO/IEC 14496-10 [24]. As expressed more clearly below, the mandatory inclusion of the specific SEI messages in the bitstream by the ITT4RT-Tx client and their decoder and rendering processing by the ITT4RT-Rx client is conditional upon successful SDP-based negotiation of the corresponding 360-degree video or fisheye video capabilities.

2- SDP-level: SEI messages shall be signalled in the SDP using the ‘sprop-sei’ parameter based on the procedures specified in IETF RFC 7798 [120] or via other SDP-based means, in the corresponding SDP offer or answer from the ITT4RT-Tx client to the ITTRT-Rx client, during the session setup involving media negotiations for 360-degree video or fisheye video. It should be noted that the signalling based on the ‘sprop-sei’ parameter is only available for the HEVC-based RTP payload formats and is not supported in AVC-based RTP payload formats as defined in IETF RFC 6184 [25].

Editor’s Note: The details of which one of the above signalling types shall be used for specific SEI messages is FFS. The frequency of the signalling of each SEI message is FFS.

Editor’s Note: The feasibility of the SDP solution is FFS. More broadly, the feasibility of an out-of-band solution for signaling SEI message information is FFS. Such a solution may consider SDP-based approaches as well as those based on RTP header extensions. Until such signaling is specified, the bitstream-level signaling of SEI messages is the default signalling method for ITT4RT.

In particular, using one of the above two signalling types (bitstream-level or RTP-level), the ITT4RT-Tx client supporting 360-degree video for viewport-independent processing shall signal the equirectangular projection SEI message (payloadType equal to 150) to the ITT4RT-Rx client, with the erp\_guard\_band\_flag set to 0.

If viewport-dependent processing (VDP) capability is successfully negotiated by the ITT4RT-Tx client and ITT4RT-Rx client for the exchange of 360-degree video, then, using one of the above two signalling types (bitstream-level or RTP-level), the ITT4RT-Tx client shall signal to the ITT4RT-Rx client either:

- the equirectangular projection SEI message (payloadType equal to 150) with the erp\_guard\_band\_flag set to 0, or

- the cubemap projection SEI message (payloadType equal to 151).

In order to optimize the spatial resolution of specific viewports, the ITT4RT-Tx client and ITT4RT-Rx client may negotiate the use of region-wise packing as part of the exchange of 360-degree video. If this is the case, the region-wise packing SEI message (payloadType equal to 155) shall also be signalled by the ITT4RT-Tx client to the ITT4RT-Rx client using one of the above two signalling types (bitstream-level or RTP-level).

If stereoscopic video capability is successfully negotiated by the ITT4RT-Tx client and ITT4RT-Rx client as part of the exchange of 360-degree video, then the frame packing arrangement SEI message (payloadType equal to 45) shall also be signalled by the ITT4RT-Tx client to the ITT4RT-Rx client using one of the above two signalling types (bitstream-level or RTP-level), with the following restrictions:

- The value of frame\_packing\_arrangement\_cancel\_flag is equal to 0.

- The value of frame\_packing\_arrangement\_type is equal to 4.

- The value of quincunx\_sampling\_flag is equal to 0.

- The value of spatial\_flipping\_flag is equal to 0.

- The value of field\_views\_flag is equal to 0.

- The value of frame0\_grid\_position\_x is equal to 0.

- The value of frame0\_grid\_position\_y is equal to 0.

- The value of frame1\_grid\_position\_x is equal to 0.

- The value of frame1\_grid\_position\_y is equal to 0.

Furthermore, ITT4RT-Tx clients supporting 360-degree fisheye video shall signal the fisheye video information SEI message (payloadType equal to 152) to the ITT4RT-Rx clients, using one of the above two signalling types (bitstream-level or RTP-level).

The bitstream delivered from an ITT4RT-Tx client to the ITT4RT-Rx client shall contain the corresponding SEI message and ITT4RT-Rx client shall process the VR metadata carried in the signalled SEI message(s) for rendering 360-degree video (provided the successful SDP-based negotiation of the corresponding 360-degree video or fisheye video capabilities associated with the SEI messages).

Editor’s Note: The feasibility of an out-of-band solution for signaling of SEI information is FFS. Such a solution may consider SDP-based approaches as well as those based on RTP header extensions. Until such signaling is specified, the bitstream-level signaling of SEI messages is the default signalling method for ITT4RT, as reflected in the paragraph above.