**Agenda item:** 10.8

**Source:** InterDigital Communications

**Title: [5G\_RTP] RTP header extension for pose information**

**Document for:** Discussion andAgreement

# Introduction

In the SA4 122nd meeting, S4-230359 [1] and S4-230390 [2] are discussed and agreed in 5G\_RTP and MeCAR WI discussions respectively. The RTP header extension for rendered pose and timestamp are described in [1] while various timestamps capturing events occurred in device and split rendering function are described in [2]. In the SA4 123rd meeting, S4-230711 [3] was discussed and it was agreed that transmitting the timestamps estimated or measured by the device (e.g., estimated-at-time T1), and the time when the rendering started (T3) at the split rendering function should be done using an RTP HE.

In the SA4 123rd meeting, S4-230708 [4] discussed the use of time information recorded at the output of the Split Rendering Server (T5), to measure the server processing delay and the overall application delay excluding the server processing delay. This was agreed in S4-230738 [5] (MeCAR PD).

This contribution proposes a method to transmit the time information recorded at the output of the Split Rendering Server (SRS) to UE, to measure the server processing delay. This contribution also proposes a method to transmit *SceneUpdateTime* recorded in the SRS to UE. *SceneUpdateTime* is used to measure the user interaction delay, age of content and the round-trip interaction delay. The semantics of the timestamps can be referred to clause 8.3 and 8.6 of MeCAR PD v7.0 [5].

# Background

The user interaction QoE metrics are defined in MeCAR PD v7.0 as below.

In shared interactive immersive services use cases, the user interaction is sent from a UE to a server. The server handles the user’s request to the immersive media scene (e.g., changing the context such as translation, rotation, and scaling or adding a new object in the scene). With the edge assisted UE type, the UE offloads the scene rendering to the Split Rendering Server, the server rasterizes the XR viewport and does pre-rendering to generate a XR media which is encoded and delivered to the UE.

In the context of interactive immersive services, one important parameter to estimate the user quality of experience is the *roundtrip interaction delay* which is defined in the TR 26.928section 4.2.2. “Interaction Delays and Age of Content”.

The *roundtrip interaction delay* is defined as the sum of the *Age of Content* and the *User Interaction Delay.*

The *User interaction delay* is defined as the time duration between the moment at which a user action is initiated and the time such an action is taken into account by the content creation engine. It is impacted by the uplink latency of the wireless network.

The*Age of content* is defined as the time duration between the moment the content is created and the time it is presented to the user. It is impacted by the downlink latency of the wireless network.

* Action and timestamp information from the device:
  + Action information: the user action information which are grouped into action sets. Each action has a unique identifier of the action
  + *lastChangeTime*: the time when the user action is made. It corresponds to the lastChangeTime field in the action information defined as the timestamp of the last change to the state of the action.
* Action and timestamp information associated with rendered media frame from the Split Rendering Server:
  + Action identifiers: The identifiers of the actions which are handled by the scene manager and rendered in the associated media frame
  + *sceneUpdateTime*: the time when the Scene manager processes the interaction task according to the actions in the action message from the UE and updates the scene.

The procedures of interactivity pipeline are shown in Figure 1.



Figure 1 - User action call flow

Using all the timestamps from the SRS and the UE, the application can calculate the interaction delays as below:

* *User-interaction-delay = sceneUpdateTime - lastChangeState*
* *Age-of-content = T2.actual – sceneUpdateTime*
* *Roundtrip-interaction-delay = T2.actual – lastChangeState*

TR 26.928 defines the User interaction delay, Age of content and round-trip interaction delay measurements as Quality of experience metrics for XR content. These delay measurements metrics need to be calculated at UE for providing better QoE. Also, the server processing delay measurement helps the UE in adaptation process with the split rendering server for achieving better QoE. It is expected that the time synchronization between the UE and SRS is performed before the start of media delivery. The in-band end-to-end delay measurement by piggybacking a delay measurement message in an RTP packet was discussed in TR 26.806. The similar approach was explored to transmit the *start-to-render-at-time* (T3), *split-rendering-server-output-time* (T5) and the *sceneUpdateTime* (T6).

# Proposed changes

[Begin Change 1]

4.3 RTP Header Extension for Rendered Pose

The split rendering server streams the rendered frame using one or more video streams, depending on the view and projection configuration that is selected by the UE. The server uses the proposed RTP header extension to associate the selected pose with the rendered frame. An RTP header extension is the most appropriate option to associate the rendered frame with its pose as it is carried as part of the RTP packets that carry the rendered images of a frame. The RTP header extension may also be used with audio streams of a split rendering process. The frequency of RTP header extension for pose is once in a frame/PDU set. It may be sent multiple times but not necessarily in every RTP packet.

Header extensions are declared in the SDP using the “a=extmap” attribute as defined in RFC8285. The header extension is identified through an association between the URI of the header extension and an ID value that is contained as part of the extension. The rendered pose header extension should use the following URN: “urn:3gpp:xr-rendered-pose”.

The two-byte header format of the header extension is used for signaling the rendered pose. The 2-byte (RFC8285) RTP header extension format of the rendered pose header extension is as follows:

0 1 2 3 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+| 0x100 |appbits| ID | length=40+2n |+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+| x |+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+| y |+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+| z |+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+| rx |+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+| ry |+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+| rz |+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+| rw |+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+| time\_info | estimated-at-time (T1) +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+ |start-to-render-at-time (T3) +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+ |split-rendering-server-output-time (T5) +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+ | scene update time (T6) +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+ | action\_id #1 | ... |+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

The (x,y,z) provides the position of the rendered pose and the (rx,ry,rz,rw) provides the orientation of the rendered pose.

The estimated-at-time (T1) and start-to-render-at-time (T3) provide the times when the pose was estimated and when the split rendering server (SRS) started to render the rendered frame, respectively. The split-renderer-output-time (T5) provides the time when the output of the SRS for a rendered frame is available. This T5 information can be used to measure the server processing delay and the overall application delay excluding the server processing delay. The SRS processes the interaction according to the actions in the action message from the UE and updates the scene. The Scene Manager records the *sceneUpdateTime* (T6) timestamp when it starts to process the actions. The s*ceneUpdateTime* is used to measure the user interaction delay, age of content and the roundtrip interaction delay. The details of s*ceneUpdateTime,* measurement of *User-interaction-delay, Age-of-content and Roundtrip-interaction-delay* QoE interaction metrics are defined in clause 8.3 and 8.6 of MeCAR PD.

The *time\_info* (8 bits) field bits represent the time stamps that are present in the RTP HE data. When T1 is present in the RTP HE, 1st bit (least significant bit) is set to 1. When T3, T5 and T6 are present in the RTP HE data, bits 3, 5 and 6 are set to 1 respectively. When T1, T3, T5 and T6 are present in an RTP HE, the *time\_info* field value will be b00110101. The transmission frequency of T3, T5 and T6 time information in RTP HE can be negotiated during the configuration phase.

These timestamps use the XR system clock. There is no requirement to synchronize the RTP stream timestamps to the XR system clock. The timestamps are passed to the XR runtime together with the rendered swapchain images (e.g., as part of the xrEndFrame call in OpenXR).

Alternatively to this format, the application and the rendering server may use unique identifiers for the transmitted pose information to reduce the required extension header size. The UE saves a timestamp, for example the estimated-at-time (T1) and associates it with a unique identifier which will be sent to split rendering server. The server transmits this unique identifier with the rendered frame that is sent to the UE. The reception of the identifier from the server triggers the UE to calculate the round-trip delay measurement which is the elapsed time between the time when the UE received the identifier from the server and the time at which the T1 timestamp is saved in the UE.

The header also provides the identifiers of all actions that were processed for the rendering of the frame. A maximum limit of 10 actions per RTP header extension for rendered pose is allowed. Hence, the total size of the header extension is set to 4 (7 + m) + 2 n, where m is the number of timestamps and n is the number of action identifiers in the header extension.

# [End of Change 1]Proposal

We propose to agree the proposed change in clause 3 into clause 4.3 of 5G\_RTP permanent document.

# References

1. S4-230359, “Signalling the render pose and other related information”.
2. S4-230390, “[MeCAR] On pose information”
3. S4-230711, “RTP header extension for pose information.”
4. S4-230708, “Pose information QoE”
5. S4-230738, “MeCAR Permanent Document v7.0.”
6. 3GPP TR 26.928, “Extended Reality (XR) in 5G”