**Source**: Interdigital Finland Oy

**Title:** [MeCAR] Pose information QoE

**Agenda item:** 9.5

**Document for:** Discussion and Agreement

#### 1 Discussion

A contribution S4-230390 introduce timing estimation between the pose information and the rendered content for split rendering service use-case. The timing information may be used to measure the QoE metric associated with the motion to photon.

This contribution details the different steps of the split rendering loop to collect and use that timing information in the device and the server. We propose to introduce a timestamp (T5), the time information at the output of the Split Rendering Server, used to measure the server processing delay.

#### 2 Proposed changes

--------------------------------------------- Begin First change ----------------------------------------------------------------------

1. 8.3 Pose information
2. 8.3.1 QoE timing information

The MeCAR device sends a group of pose information to the server's split render function to generate rendered media frames based on the poses. Each pose is associated with time metadata, such as the time when the pose estimation was made (T1), the estimated target display time of the content (T2.estimated), and the time the group of poses was sent (T1').

The gap between the actual display time (T2.*actual*) and the pose estimate time (T1) is the pose-to-render-to-photon delay, which allows the MeCAR device to know the amount of processing time as well as the connection delay required for a loop of split rendering. The next round of pose estimation should refer to the pose-to-render-to-photon delay for the estimation of a new T2.estimated. The split render function in the server may refer to T1', which is the time when the group of poses is sent from device, if multiple pairs of pose and metadata for the same target display time are received from the device. The T1’ information may be used to manage poses by the server, such as allowing the MeCAR device to update former estimations by resubmitting a new pose with the same estimated-target-display-time.

The split render function in the server sends rendered media frames and associated metadata. The metadata shall include the pose used for the rendered frame, as well as corresponding time information, such as T1, T2.estimated, and may include the time when the rendering started (T3), all sent to the MeCAR device in order to measure the render-to-photon delay.

The (T5) is the time information at the output of the Split Rendering Server (SRS).  
The (T5) timestamp can be used to:

* measure the server processing delay, (T3 – T5)
* the overall application delay excluding the server processing delay,

Note: the downlink delay can be measured with T5 and inferred timing from the MAF, and the uplink delay can be measured with (T1’ - T3).

In the case where there are poses stacked in the server’s pose buffer, for example with a granularity finer than the device's supported frame rate, the split render function should select the pose closest to the display time, according to the previous-render-to-photon delay. The previous-render-to-photon delay from the most recent frame information may help the server to make this selection.



**Procedure**

1. The raw pose is acquired by the XR Source Management from the XR Runtime.
2. The XR Source Management estimates the target display time (T2.estimated). To estimate that time, it uses the previous timestamps and delays measurements of the pose-to-render-to-photon which is the gap between the (T2.actual) and the pose estimate time (T1)for the most recent frame.
3. The XR Source Management predicts the pose at the estimated target (T2.estimated)of the content. The pose prediction is made at (T1).
4. The UE MAF gets the predicted poses or group of poses with the associated times metadata (T2.estimated, T1, previous-render-to-photon).
5. The UE MAF sends the predicted pose or group of poses and the associated time metadata. The MAF appends to the time metadata the (T1’) which is the time when the group of poses is sent from device.
6. The Scene Manager in the Split Rendering Server renders the scene based on the received predicted pose. It records the time when the rendering started (T3) and appends it to the time metadata.  
   The predicted pose used for rendering with its associated time metadata is stored with the output rendered media frame.  
   In the case where there are poses stacked in the server’s pose buffer, the Scene Manager in the Split Rendering Server should select one of the (predicted pose, T2.estimated )pairfollowing a local estimation of the display time by using the previous-render-to-photon received in the time metadata.
7. The rendered media frame is sent to the video encoder with the associated time metadata.
8. The Split Rendering Server encodes the rendered media frame together with the associated time metadata. The (T5) timestamp is recorded at the output of the Split Rendering Server.The encoded media frame is sent from the Split Rendering Server to the UE MAF with the associated time metadata.
9. The UE MAF decodes the media data. The rendered frame is then shared to the Presentation Engine and the XR runtime.
10. The XR runtime performs further processing such pose correction using the latest pose. The UE records the time (T4)of the latest pose.
11. The frame is displayed at the actual display time (T2.actual).
12. 8.3.2 Pose information delays and QoE

The timing metadata are used to measure the following delays:

* *pose-to-render-to-photon delay = T2.actual – T1*
* *render-to-photon delay = T2.actual – T3*
* *pose-to-render delay = T3 – T1*
* *motion-to-photon = T2.actual – T4*
* *server-processing-delay = T5 – T3*

By using all the history of delay measurements, the Application can estimate the delays for the next poses and rendered frames.

* Pose and timestamp information from the device:
  + Predicted pose: it includes location and direction information
  + (T1): the time when the pose estimation was made
  + (T2.estimated): the estimated target display time for the media frame which will be rendered, using to this pose.
  + (T1’): the actual time when a pose or a group of poses is sent from the device to the Split Rendering Server
  + previous-render-to-photon: the render-to-photon delay for the most recent frame (*T2.actual – T3*).
* Pose and timestamp information associated with rendered media frame from the Split Rendering Server:
  + Pose used for rendering
  + (T1): the time when the pose estimation was made
  + (T2.estimated): the estimated target display time for the media frame which is rendered.
  + (T3): is the actual time when the renderer in the Split Rendering Server starts to render the associated media frame
  + T5: is the time when the rendered media frame is outputed from the Split Rendering Server

--------------------------------------------- End First change -------------------------------------------------------------------------



#### 3 Proposal

We propose to update the section 7.2 and the section 8.3 of MeCAR PD v0.5.1 with the proposed changes in clause 2.