5G Real-time Media Transport Protocol Configurations for XR services

By Razvan-Andrei Stoica (Lenovo), Igor D.D. Curcio (Nokia)

In various 3GPP groups, including SA2, RAN1 and RAN2, multiple Extended Reality (XR) enhancements were added across the communication layers for high data rate and low latency XR traffic.

These enablers empowered XR content awareness in QoS handling of media delivery from application down to radio layer for an improved QoE of XR immersive content. Working Group SA4 defined transport protocol extensions and configurations to help 5G XR developers to leverage the potential of these new cross-layer optimizations.

The WG SA4 XR investigations revealed the relevance of the concept of Application Data Units (ADUs) granularity, e.g., video frame/slice, metadata, etc., in XR traffic transport optimizations. Working Groups RAN2 and SA2 integrated this core concept into their radio resource allocation and QoS flows mechanisms under the “PDU Set” guise, i.e., one or more PDUs carried by an ADU.

The 5GS PDU Set awareness is based on acquiring different PDU Set information: PDU Set Sequence Number (PSSN), End PDU of the PDU Set (E), PDU Sequence Number within a PDU Set (PSN), PDU Set Size (PSSize) and PDU Set Importance (PSI).

****** WGSA4 enabled XR services and applications to expose such information to the 5GS through the media transport by an RTP header extension (HE) for PDU Set marking.

The RTP HE is using the general RTP HE mechanism in IETF RFC 8285, applicable to (S)RTP and WebRTC, to mark the boundaries of a PDU Set, its size and its importance, whereas the marking is applied to all PDUs of the PDU Set, as in Figure 1.

The PSI indicates the importance of a PDU Set relative to other PDU Sets within the same multimedia session. This information may support RAN-based informed discard of PDUs and PDU Sets when needed, e.g., during air-interface congestion. In TS 26.522, WG SA4 provides general guidelines for XR services and application developers in setting the PSI field for the H.264 and H.265 video codecs. In general, the concept assigns higher importance to Intra-coded frames/slices relative to P/B-frames or P/B-slices.

The RTP HE for PDU Set marking also includes an indication of the end of a data burst (D), i.e., one or more PDU Sets. This information may be utilized by the RAN to optimize power savings and configure XR devices in a deeper sleep state.

XR services often incorporate split rendering to maintain low energy consumption on XR devices while still delivering immersive experiences. In a nutshell, a dedicated XR split rendering server receives, e.g., at a network edge, from an XR client user pose information capturing user motion, pre-renders content based on the available pose information, encodes the pre-rendered content, and delivers it back to the XR client. The latter decodes the content and performs a power-efficient rendering to the user.

To preserve low (50-70 ms) motion-to-render-to-photon (M2R2P) latency for XR services and to enable XR split rendering, WG SA4 has specified a RTP HE for exchanging XR pose information, eg. to indicate rendered pose for split rendered media. This HE piggybacks on available media streams using the (S)RTP HE general format of RFC 8285. The RTP HE for XR pose provides thus a real-time XR pose transport mechanism including 3DoF or 6DoF coordinates, an XR timestamp associated with the XR pose in the context of the XR service, and a list of action identifiers determining the XR action spaces correspondent to the XR pose coordinates. The XR client may so use the rendered pose to optimize for QoE the late stage reprojection and on-device rendering routines at runtime.

Two RTP HEs for E2E delay measurements are also specified in order to support XR immersive experiences. One HE is meant to measure the one-way delay between time-synchronized XR endpoints communicating via (S)RTP, based on single timestamp, T1, i.e., the “Originate Timestamp”. The second HE is a response to the first HE indexed by the timestamp T1 and includes two more timestamps, T2, i.e., the “Receive Timestamp”, and T3, i.e., the “Transmit Timestamp”.

By recording a fourth timestamp T4, i.e., the “Destination Timestamp”, one of the XR endpoints can thus calculate the loaded data flow traffic in terms of one-way delay (uplink or downlink), round-trip delay and processing delay metrics. Figure 2 displays the RTP HE pairs for XR pose and E2E delay measurement in a split rendering scenario.



Additional timing reporting for split rendering is enabled by RTCP extended reports for QoE measurements. These include XR server timing information allowing XR devices to determine QoE metrics such as round-trip delay, server processing delay, user interaction delay, age of content and round-trip interaction delay.

The 3GPP XR RTP protocol extensions are a toolset for service providers and application developers to unlock immersive XR media delivery over the 5GS and will be further complemented and enhanced with additional features in Rel-19.

For more on WG SA4: www.3gpp.org/3gpp-groups