**3GPP TSG SA WG4#116e S4-211368**

**E-meeting, 10th – 19th November 2021**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.1* | | | | | | | | |
| **PSEUDO CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  | 26.998 | **CR** |  | **rev** |  | **Current version:** | 1.0.3 |  |
|  | | | | | | | | |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* | | | | | | | | |
|  | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME |  | Radio Access Network |  | Core Network |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| ***Title:*** | [FS\_5GSTAR] Scenario KPIs and QoS | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** | Qualcomm Incorporated | | | | | | | | | |
| ***Source to TSG:*** |  | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** | FS\_5GSTAR | | | | |  | ***Date:*** | | | 2021-11-02 |
|  |  | | | |  | |  | | |  |
| ***Category:*** | C |  | | | | | ***Release:*** | | |  |
|  | *Use one of the following categories:* ***F*** *(correction)* ***A*** *(mirror corresponding to a change in an earlier release)* ***B*** *(addition of feature),* ***C*** *(functional modification of feature)* ***D*** *(editorial modification)*  Detailed explanations of the above categories can be found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | | | | | | | | *Use one of the following releases: Rel-8 (Release 8) Rel-9 (Release 9) Rel-10 (Release 10) Rel-11 (Release 11) … Rel-15 (Release 15) Rel-16 (Release 16) Rel-17 (Release 17) Rel-18 (Release 18)* | |
|  |  | | | | | | | | | |
| ***Reason for change:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | | 6.2.5, 6.2.6, 6.2.7, 6.3.6, 6.3.7, 6.4.6, 6.4.7 | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **Y** | **N** |  | | | |  | | |
| ***Other specs*** | |  |  | Other core specifications | | | | TS/TR ... CR ... | | |
| ***affected:*** | |  |  | Test specifications | | | | TS/TR ... CR ... | | |
| ***(show related CRs)*** | |  |  | O&M Specifications | | | | TS/TR ... CR ... | | |
|  | |  | | | | | | | | |
| ***Other comments:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***This CR's revision history:*** | |  | | | | | | | | |

**===== CHANGE =====**

### 6.2.5 Content formats and codecs

Based on the use cases, the following formats, codecs and packaging formats are of relevance for Media Streaming of AR:

* General

> Basic scene Graph and Scene Description

> 2D Video Formats and video compression codecs

> Regular Audio Formats and audio compression codecs

- In addition, for STAR-based UE

> richer scene graph data

> 3D Formats such as static and dynamic point clouds or meshes

> several video decoding instances

> decoding tools for such formats

> DASH/CMAF based delivery

- In addition, for EDGAR-based UE

> 2D compression tools for eye buffers as defined in clause 4.5.2.

> decoding tools for such formats

> at least two video decoding instances

> low-latency downlink real-time streaming of the above media

>Uplink streaming of pose information and other relevant information, such as input actions

**===== CHANGE =====**

### 6.2.6 KPIs and QoS

The above scenarios relate to the following cases in TR 26.928 [2], clause 6. In particular:

* For STAR:

- Viewport-independent streaming based on clause 6.2.2 as defined TR 26.928 [2],

- Viewport-dependent streaming based on clause 6.2.3 as defined TR 26.928 [2],

* For EDGAR

- Raster-based split rendering based on clause 6.2.5 as defined TR 26.928 [2].

For STAR-based devices and viewport-independent streaming, processing of updated pose information is only done locally in the XR device. Delivery latency requirements are independent of the motion-to-photon latency. Initial considerations on QoE parameters are provided in TR 26.928 [2], clause 6.2.2.5. The XR media delivery are typically built based on download or adaptive streaming such as DASH such that one can adjust quality to the available bitrate to a large extent. Compared to the viewport independent delivery, for viewport dependent streaming, updated tracking and sensor information impacts the network interactivity. Typically, due to updated pose information, HTTP/TCP level information and responses are exchanged every 100-200 ms. For more details, refer to clause 6.2.3 as defined TR 26.928 [2]. Such approaches can reduce the required bitrate compared to viewport independent streaming by a factor of 2 to 4 at the same rendered quality. It is important to note that viewport-dependent streaming technologies are typically also built based on adaptive streaming allowing to adjust quality to the available bitrate. The knowledge of tracking information in the XR Delivery receiver just adds another adaptation parameter. However, generally such systems may be flexibly designed taking into account a combination/tradeoff of bitrates, latencies, complexity and quality. Suitable 5QI values for adaptive streaming over HTTP are 6, 8, or 9 as defined in TS 23.501 [X], clause 5.7.4 and also indicated in clause 4.3.3 of TR 26.928 [2].

For EDGAR-based devices, raster-based split rendering based on clause 6.2.5 as defined TR 26.928 [2] applies. With the use of pose corrections, the key latency for the network is the motion-to-render-to-photon delay as introduced in clause 4.5.2 and 4.5.3, i.e. the end-to-end latency between the user motion and the rendering is 50-60ms. The formats are defined in clause 4.5.2 as follows

- for 30 x 20 degrees, 1.5K by 1K per eye is required and 1.8K by 1.2K per eye is desired

- for 40 x 40 degrees, 2K by 2K required and 2.5 K by 2.5 K desired

Colours are typically RGB but may be converted to YUV. Framerates are typically 60fps to 90fps. The above formats results in typically in maximum 4K content at 60 fps. Modern compression tools can compress this to 30 to 50 Mbit/s. Regular stereo audio signals are considered, requiring bitrates that are negligible compared to the video signals. In order to support warping and late stage reprojection, some depth information may be added. For communication a real-time capable content delivery protocol is needed, and the network needs provide reliable delivery mechanisms. 5QI values exist that may address the use case, such 5QI value number 80 with 10ms, however this is part of the non-GBR bearers (see clause). In addition, it is unclear whether the 10ms with such high bitrates and low required error rates may be too stringent and resource consuming.

Hence, for simple split rendering in the context of the requirements in this clause, suitable 5QIs 89 and 90 have been defined in Rel-17 in TS 23.501 in Rel-17 addressing the latency requirements in the range of 10-20ms and bitrate guarantees to be able to stream up to 50 Mbps consistently. Significant opportunities exist to support split rendering with advanced radio tools, see for example TR 26.926 [48] for performance evaluation.

The uplink is predominantly the pose information. Data rates are several 100 kbit/s and the latency should be small in order to not add to the overall target latency. Suitable 5QIs 87 and 88 have been defined in Rel-17 in TS 23.501 to stream uplink pose information.

**===== CHANGE =====**

### 6.2.7 Standardization areas

The list of potential standardization area that has been collected is provided in the following:

- HTTP-Streaming of immersive scenes with 2D and 3D media formats and objects to STAR-based devices including

- Immersive media format and profile with integration into 5GMS for STAR-based devices

- Scene description format, functionality, and profile as an entry point of immersive media

- Relevant subset of media codecs for different media types and formats

- CMAF encapsulation of immersive media for 5G media streaming

- Viewport independent and viewport dependent streaming

- Split rendering delivery of immersive scenes to EDGAR-based devices

- simple 2D media formats that match AR glass display capabilities

- Media payload format to be mapped into RTP streams

- Capability exchange mechanism and relevant signalling

- Protocol stack and content delivery protocol

- Cross-layer design, radio and 5G system optimizations for QoS support

- uplink streaming of predicted pose information and input actions

- Required QoE metrics

**===== CHANGE =====**

### 6.3.6 KPIs and QoS

The above scenarios relate to the following cases in TR 26.928 [2], clause 6. In particular:

* For STAR:

- Viewport-dependent streaming based on clause 6.2.3 as defined TR 26.928 [2],

- Raster-based split rendering based on clause 6.2.5 as defined TR 26.928 [2],

- Generalized XR split rendering based on clause 6.2.6 as defined TR 26.928 [2].

* For EDGAR

- Raster-based split rendering based on clause 6.2.5 as defined TR 26.928 [2].

For STAR-based delivery, a basic architecture as shown in Figure 6.3.6-1 applies. Two important latency considerations are important:

* User interaction latency, i.e. 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 stage performer or content creation engine. In the context of gaming, this is the time between the moment the user interacts with the game and the moment at which the game engine processes the player’s response.
* End-to-End Latency (EEL): The latency for an action that is originally presented in the scene or captured by camera until its visibility on the remote display.
* Round-trip Interaction Delay (RID): The time of an action by the user until it sees the action reflected on its screen. This delay is the sum of the user interaction delay and End-to-end-Latency.



Figure 6.3.6-1: Architecture and latencies for interactive immersive service

The maximum RID depends on the type of scene. A typical example is the Stadia cloud gaming platform and an excellent introduction is provided here [52]. Some extracted high-level requirements on user experience for RID are provided in Figure 6.3.6-2.

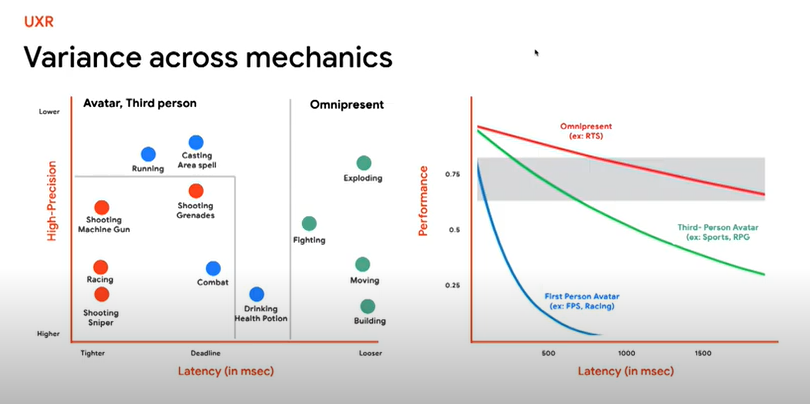


Figure 6.3.6-2: User experience depending on Round-trip Interaction Delay/Latency for different scenes (see [52])

Similar data is collected in TR 26.928 [2], clause 4.5. Typically, systems have maximum delay requirements between 60ms and 500ms. In terms of formats and bitrates, similar considerations as for clause 6.2.6 apply. However, note that in many cases a pre-rendering is applied in the network, such that data rates and formats are more similar to the split-rendering considerations. Similar considerations as for clause 6.2.6 apply on raster-based split rendering apply.

For EDGAR-based devices, raster-based split rendering based on clause 6.2.5 as defined TR 26.928 [2] applies. Similar considerations as for clause 6.2.6 apply.

The uplink is predominantly the pose information and user interactions. Data rates are several 100 kbit/s and the latency should be small in order to not add to the overall target latency.

**===== CHANGE =====**

### 6.3.7 Standardization areas

The list of potential standardization area that has been collected is provided in the following:

- Streaming of immersive scenes with 2D and 3D media formats and objects to STAR-based devices including

- Low-latency streaming protocols to support latencies in the range between 50 to 500ms, typically using RTP-based real-time streaming based on cloud rendering

- Scene description format, functionality, and profile as an entry point of immersive media

- Simplified 3D media formats and 2D media formats with integration for STAR-based devices

- Relevant subset of media codecs for different media types and formats

- RTP encapsulation of media formats

- 5G System and 5G Media Streaming support

- Split rendering delivery of immersive scenes to EDGAR-based devices

- simple 2D media formats that match AR glass display capabilities

- Media payload format to be mapped into RTP streams

- Capability exchange mechanism and relevant signalling

- Protocol stack and content delivery protocol

- Cross-layer design, radio and 5G system optimizations for QoS support

- uplink streaming of predicted pose information

- Required QoE metrics

**===== CHANGE =====**

6.4.6 KPIs and QoS

In the downlink this scenario is equivalent to clause 6.3.6 and similar KPIs and QoS aspects apply.

For the uplink, the above scenarios relate to the following cases in TR 26.928 [2], clause 6. In particular:

* XR distributed computing based on clause 6.2.7 as defined TR 26.928 [2].

Details on uplink streaming of sensor data is ffs.

**===== CHANGE =====**

### 6.4.7 Standardization areas

The list of potential standardization area that has been collected is provided in the following:

- Similar functionalities as identified in clause 6.3.7 for downlink

- For the uplink, streaming of sensor information to the network

- Low-latency streaming protocols to support latencies in the range between 50 to 500ms, typically using RTP-based real-time streaming

- simple 2D media formats to stream match AR sensor data

- Payload format to be mapped into RTP streams

- Capability exchange mechanism and relevant signalling

- Protocol stack and content delivery protocol

- Cross-layer design, radio and 5G system optimizations for QoS support

- Required QoE metrics