Source: Samsung Electronics Co., Ltd.

**Title: FS\_5GSTAR: Unidirectional media flow from server to UE**

**Agenda Item: 10.9**

**Document for: Discussion and Agreement**

# **Introduction**

During the Video SWG telco in Jan. 19th, it was agreed to have several device types and their detailed functional architectures. Based on those agreements and the use cases identified, this contribution is proposing the basic architecture and the immersive service flow.

# **Immersive Media Flow**

The use cases of this study, as described in Table 5.1 of TR 26.998, fall into one of the three immersive media flow scenarios, depending on the use case’s source and destination:

1. Unidirectional : from server to UE
2. Unidirectional : from UE to server
3. Bi-directional

We have identified that a majority of the collected use cases contain scenarios where immersive media is exchanged between 2 UEs in a bidirectional manner, mostly in a real-time communication or conversation scenario such as an AR call or AR conference (3)). Also, there are some typical cases like AR streaming which falls into 1), and 3D image messaging which can be a combination of 1) and 2).

For example, the use case #2 (AR sharing), described in Table 5.1 of TR 26.998 is the story about buying a new couch at home. A user would receive 3D model of the couch from the server of furniture store and enjoy an AR experience by augmenting this virtual model into the living room. Additionally, this augmented scene (virtual couch surrounded by real living room) would be captured to share with the friends or to upload to the server. In this case, the exchanged media from the server to UE should be immersive in order to maximize the sense of augmented reality. On the other hand, the captured scene may be simply 2D picture/image or even 3D (depending on the capturing device capability). Therefore, depending on the detailed scenario, this use case #2 can be solely the case of 1) (Unidirectional: from server to UE) or the combination of 1) and 2).

This contribution is covering the case of 1).

## 2.1 Scenario 1: Extension to 5G Media Streaming Downlink for AR

As described above, this clause covers the cases where the immersive media – compared to the traditional 2D media – is transmitted in a unidirectional path, which is from the server (e.g., AR/MR Application Provider) to UE. Additionally, the reverse path (from UE to the server) can be exploited to deliver the metadata and/or traditional 2D media, depending on each use cases and scenario.

Figure X identifies this unidirectional immersive media flow according to each of device type, described in Table Y of clause 4.2 in TR26.998.



Figure X: Unidirectional immersive media flow – from server to UE

The immersive media was assumed to be generated and stored before the service is initialized. Then it is transmitted to UE (AR device) throughout 5G system via Uu interface (represented as red arrow in the figure). In the figure, three different media path were shown depending on each of device type. In general, Path 1 maps to Type 1, Path 2 maps to Type 2, and Path 3 maps to Type 3.

However, Type 1 and Type 3 can particularly have another variant to split AR/MR functions to Cloud/Edge, as described in Table Y of clause 4.2. In this case, the media pipeline becomes to be identical to that of Path 2.

Therefore, each device type can be mapped as follows;

* Path 1 : Type 1 (when AR/MR functions are available in AR device)
* Path 2 : Type 2 & Type 1/Type 3 (when AR/MR functions are split to cloud/edge)
* Path 3 : Type 3 (when AR/MR functions are split to tethered device)

Each of media path has its characteristics / properties in terms of QoS such as required data rate and latency. Below shows some of key observations

* Path 1 requires higher bandwidth than Path 2 as it directly sends the immersive media (e.g., 3D), rather than pre-rendered 2D. However, it can be advantageous in terms of latency as it does not need additional 2D encoder/decoder process.
* Path 2 additionally requires uplink traffic to deliver metadata for pre-processing (e.g., 6DoF pose information for pre-rendering)
* Media data and its relevant metadata for Path 3 are almost identical to those of Path 1, but more latency caused by extra link between phone to AR glasses (e.g., 2D encoder/decoder, sidelink delay) should be taken into account.

2.1.1 Case Study – UC #18 Streaming of volumetric video for glass-type MR devices

This clause will document a case study how the immersive media flow looks like for a specific use case. Here, use case #18 as a typical example of unidirectional case is chosen as reference.

The use case #18 in Table 5.1 of technical report is about streaming volumetric media, captured by a gym instructor in advance. A user at home wearing an AR device receives the high-quality volumetric media in a streaming manner by enabling to play, pause, and rewind. While the user is walking around with 6DoF, he/she can see the different side of the virtual instructor and even can see the occlusion if it is blocked by real-world object.

Assuming the Type-2 device and pre-rendering at the cloud/edge, the basic procedure for this service is as follows;

1) A user selects the desired volumetric video. Information of device capability (e.g., display resolution, processing capability) is exchanged between cloud/edge and UE. Cloud/edge prepares the immersive media pipeline for pre-rendering

2) UE sends to the cloud/edge the metadata information for pre-rendering including 6DoF pose. Additionally, supplemental information for immersive renderer may also be sent to support the required processing for occlusion (e.g., Spatial mapping information to provide a detailed representation of real-world surfaces around the device)

3) Cloud/edge starts sending the pre-rendered 2D video from the received 6DoF pose. Additional metadata such as information for pose correction (e.g., depth), augmentation (e.g., object size, default orientation) may also be delivered as well as information used for pre-rendering (e.g., timestamp, pose info used)

4) Volumetric video is represented based on the extracted information from vision engine in UE.

In short, the following information may be required in this scenario; (example, but not limited to)

* Downlink
  + Media data
    - Pre-rendered coded RGB 2D video
    - Information for pose correction: Coded depth video
  + Metadata
    - Information for representation: timestamp, pre-rendered info (pose used for the render)
    - 3D model information for augmentation: object size, default orientation, etc
* Uplink
  + Metadata
    - 6DoF pose information
    - Supplemental information for immersive renderer: Spatial mapping information to be rendered, light source information, etc.

If UE is using path 1 or 3 in Figure X, then some parts of procedure above need to be changed or even dropped out. For example, the immersive media is changed to have 3D or volumetric format (e.g., point cloud or mesh) and the metadata for rendering can be for internal purpose and may not need to be exchanged via uplink interface. The detailed format and context for media and metadata is FFS.

[Editor’s Note] For next steps,

a. The basic procedures and call flows for the two architectures

b. identify the relevant content formats and rendering requirements

c. identify the relevant QoS and QoE Parameters

d. identify the potential standardization needs

# **Proposal**

We propose to include the text and table in section 2 of this document into the Permanent Document as a baseline for detailed discussions and the way forward.