|  |  |
| --- | --- |
| 3GPP TR 26.803 V0.2.0 (2020-08) | |
| 3rd Generation Partnership Project;  Technical Specification Group Services and System Aspects;  Study on 5G Media Streaming Extensions for Edge Processing  (Release 17) | |
|  | |
|  |  |
|  | |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP. The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification. Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. | |

|  |
| --- |
|  |
| ***3GPP***  Postal address  3GPP support office address  650 Route des Lucioles - Sophia Antipolis  Valbonne - FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  http://www.3gpp.org |
| ***Copyright Notification***  No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.  © 2020, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).  All rights reserved.  UMTS™ is a Trade Mark of ETSI registered for the benefit of its members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® and the GSM logo are registered and owned by the GSM Association |

Contents

Foreword 4

Introduction 5

1 Scope 6

2 References 6

3 Definitions of terms, symbols and abbreviations 6

3.1 Terms 6

3.2 Symbols 6

3.3 Abbreviations 6

4 Overview of Relevant Architectures 7

4.1 General 7

4.2 SA6 Edge Architecture 7

4.3 SA2 Edge Support 8

4.4 SA5 Edge Management 9

4.5 5GMS Architecture 10

5 Use Cases for Edge Media Processing 10

5.1 General 10

5.2 Downlink Streaming Use Cases 10

5.2.1 Caching Downlink Streaming Content 10

5.2.2 Split Rendering 11

5.2.3 User-generated live streaming 14

5.2.4 Augmented Video Streaming 17

5.2.5 Generalized Split and Cloud Rendering and Processing 19

5.2.6 Photo-realistic AR Rendering in Network 21

5.2.7 Media Services in the Edge 23

5.2.8 Partial delivery of 3D content (point cloud, mesh) for AR/MR device 24

6 Potential 5GMS Architecture Extensions 26

6.1 General 26

6.2 Common Extensions 26

6.3 Extensions for Downlink 26

6.4 Extensions for Uplink 26

Annex <A> (normative): <Normative annex for a Technical Specification> 27

Annex <X> (informative): Change history 28

**Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.**Error! Bookmark not defined.Error! Bookmark not defined.**Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.**Error! Bookmark not defined.Error! Bookmark not defined.

# Foreword

This Technical Specification has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# Introduction

In this Technical Report, we study the architectural enablers and extensions to the 5G Media Streaming Architecture for edge media processing. It covers aspects such as discovery, configuration, execution, and management of media processing in 5G to address the needs identified for example by the XR5G and FLUS items.

# 1 Scope

This clause shall start on a new page.

The present document …

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

…

[x] <doctype> <#>[ ([up to and including]{yyyy[-mm]|V<a[.b[.c]]>}[onwards])]: "<Title>".

# 3 Definitions of terms, symbols and abbreviations

## 3.1 Terms

For the purposes of the present document, the terms given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

**example:** text used to clarify abstract rules by applying them literally.

## 3.2 Symbols

For the purposes of the present document, the following symbols apply:

<symbol> <Explanation>

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

<ABBREVIATION> <Expansion>

# 4 Overview of Relevant Architectures

## 4.1 General

In this section, we provide an overview of ongoing SA activities on edge computing architectures.

## 4.2 SA6 Edge Architecture

SA6 has taken significant steps towards the definition of normative edge computing architecture for 5GC in [3]. Starting from common scenarios, described in the Annex, a set of requirements is defined, and the following architecture is proposed:



The architecture defines the key nodes and functions as well as the interfaces between them.

The identified functions with a brief description is given here:

* Edge Enabler Server (EES): provides supporting functions needed for Edge Application Servers and Edge Enabler Client.
* Edge Enabler Client (EEC): provides supporting functions needed for Application Client(s).
* Edge Configuration Server (ECS): provides supporting functions needed for the Edge Enabler Client to connect with an Edge Enabler Server.
* Edge Application Server (EAS): the application server resident in the Edge Data Network, performing the server functions. The Application Client connects to the Edge Application Server in order to avail the services of the application with the benefits of Edge Computing.
* Application Client (AC): application resident in the UE performing the client function. Details of the Application Client are out of scope of this specification.

A typical sequence of steps to use edge computing services is as follows:

* EESs register with the ECS to publish their edge configuration capabilities
* The EEC is provisioned with a list of EESs, e.g. from the ECS
* The EEC registers with a selected EES.
* EASs register with EESs to publish their edge capabilities
* The EEC queries the EES to discover specific EASs. Different types of filtering can be used during this discovery phase
* The EAS may register for location and other notifications with the EES related to a specific session
* The AC is now able to connect to the EAS to consume edge computing services

Figure 2 represents the SA6 edge server architecture as defined in 23.558:



Figure 2 SA6 Edge data network architecture

The XR server capabilities needed for the split rendering use-case if is run on an edge server, is supported by an Edge Application Server (EAS) in the above figure.

TS23.558 provides Edge Application Server KPI discovery as shown in Table 1:

* Table 1: Edge Application Server Service KPIs

|  |  |  |
| --- | --- | --- |
| Information element | Status | Description |
| Maximum Request rate | O | Maximum request rate from the Application Client supported by the server. |
| Maximum Response time | O | The maximum response time advertised for the Application Client's service requests. |
| Availability | O | Advertised percentage of time the server is available for the Application Client's use. |
| Available Compute | O | The maximum compute resource available for the Application Client. |
| Available Graphical Compute | O | The maximum graphical compute resource available for the Application Client. |
| Available Memory | O | The maximum memory resource available for the Application Client. |
| Available Storage | O | The maximum storage resource available for the Application Client. |
| Connection Bandwidth | O | The connection bandwidth in Kbit/s advertised for the Application Client's use. |

## 4.3 SA2 Edge Support

Edge Computing has been identified as a key feature of 5G early on and has been specified as part of the 5G System Architecture in clause 5.13 of [1]. It describes the selection of a close UPF for non-roaming or LBO-connected UEs. The specification lists different options to enable access to edge computing:

* User plane (re)selection: the 5G Core Network (re)selects UPF to route the user traffic to the local Data Network
* Local Routing and Traffic Steering: the 5G Core Network selects the traffic to be routed to the applications in the local Data Network
* Session and service continuity to enable UE and application mobility
* An Application Function may influence UPF (re)selection and traffic routing via PCF or NEF
* Network capability exposure: 5G Core Network and Application Function to provide information to each other via NEF or directly
* QoS and Charging: PCF provides rules for QoS Control and Charging for the traffic routed to the local Data Network;
* Support of Local Area Data Network: 5G Core Network provides support to connect to the LADN in a certain area where the applications are deployed

In addition, SA2 has started a new release 17 study item on enhancement of support for edge computing in 5GC [2] to address identified gaps to enable edge computing. The findings will be documented in TR 23.748.

So far, the identified key issues are:

1. Discovery of Edge Application Server
2. Edge relocation
3. Network Information Provisioning to Locate Applications with Low Latency
4. (Consecutive traffic steering in different N6-LAN)
5. Activating the traffic routing towards local data network per AF request

Section 6 of 23.748 documents a set of solutions for the previously identified key issues.

## 4.4 SA5 Edge Management

SA5 has recently agreed a new study item [4] on the management aspects of edge computing. The focus of the study item will be on the following scenarios:

* How 3GPP management solutions support the 3rd party service provider to deploy and manage Edge Configuration Server.
* How 3GPP management solutions support Edge Computing Service Provider(s) to deploy and manage Edge Enable Server(s).
* How 3GPP management solutions support Edge Application Provider(s) to deploy and manage Edge Application Server(s).

As can be seen, SA5 adopts the SA6 architecture and will study ways of enabling it through providing the necessary management functions.

The scope has been set to the following items:

* + The lifecycle management of EAS, EES and ECS needed to support edge computing, by considering the various deployment scenarios.
  + Deployment and provisioning of 5GC network functions needed to support the edge computing.
  + Provisioning of EES, and ECS needed to support edge computing.
  + Performance assurance of EES, ECS and relevant 5GC functions needed to support edge computing.
  + Fault supervision of EES, ECS and relevant 5GC functions needed to support edge computing.
  + Mechanism(s) to enable and support EAS deployment on a particular edge data network.
  + Mechanism(s) to enable and support
    - ECSP to deploy and manage EES and ECS.
    - ASP to deploy and manage EAS.
  + Studying the need of providing management provisions to create and manage communication service(s) at a particular edge data network.

## 4.5 5GMS Architecture

# 5 Use Cases for Edge Media Processing

## 5.1 General

Editor’s Note: only relevant use cases should be documented.

## 5.2 Downlink Streaming Use Cases

## 5.2.1 Caching Downlink Streaming Content

|  |
| --- |
| **Use Case Name** |
| Caching downlink streaming content |
| **Description** |
| A Mobile Network Operator that deploys a downlink streaming service or supports the delivery of media content from a third-party service wants to offer that content in the highest possible quality to all of its users. The MNO also notices that video streaming already accounts for a large part of the traffic on the backhaul network. For these reasons, the MNO wants to offload (part of the) content hosting from the CDN to caches near or within its network. Users of the service may access the content from the edge, allowing them to select higher quality renditions of the content (e.g., DASH representations) and play it back without interruptions. The MNO may improve the hit ratios of the cache by employing intelligent caching. Furthermore, to ensure that clients access the content from the optimal edge, the network operator may want to direct clients to this edge. |
| **Categorization** |
| **Type: CDN**  **Delivery: Download, Live Streaming, On Demand Streaming**  **Device: Phone, tablet, HMD, TV** |
| **Preconditions** |
| End user devices should be able to stream, decode, and display the video streams. Modern smartphones already have these capabilities. |
| **Requirements in terms of Capabilities and QoS/QoE Considerations** |
| The capabilities of edge nodes are similar to regular CDN nodes distributing video content, although at smaller scale. This means that edge nodes should have storage and HTTP serving capabilities, and UEs should have high-bandwidth connectivity to edge nodes. Higher video quality, less playback interruptions, and shorter loading times improve the QoE. |
| **Feasibility and Industry Practices** |
| TBD |
| **Nominal Cost Analysis** |
| Using edge computing for video content caching is the next step in distributing video delivery. It will allow MNOs and streaming services to further scale up and serve more users, while reducing load on the backhaul network. As in a regular CDN node, a node at the edge can be used by many users at the same time and servers scale horizontally. MNOs can use existing facilities at PoPs or points further in the network with serving capabilities. |
| **Benefits and Impact** |
| The major benefit is expected for MNOs and service providers, who are able to serve more users with high quality video while significantly reducing the load on the backhaul network, thus improving the efficiency of the network infrastructure. End-users are expected to benefit as it will increase the access to content in a high video quality, also enabling demanding streaming applications including VR, and delivering those applications with shorter loading times and with fewer interruptions. |
| **Potential Technical Requirements** |
| * It should be possible for edge caches to be operated either by the MNO or by a third-party service such as a 5GMSd Application Provider. * It should be possible for the network to steer clients to a certain edge or CDN. * It should be possible for (third-party) services to specify caching directives. * It should be possible for DASH clients to send hints (e.g., about anticipated upcoming requests) to the network enabling intelligent caching on the edge. * It should be possible for the network to send hints to clients regarding the delivery of content from the edge (e.g., about availability or bandwidth). |
| **Potential Standardization Status and Needs** |
| TBD |

## 5.2.2 Split Rendering

|  |
| --- |
| **Use Case Name** |
| Split Rendering |
| **Description** |
| The system design for split rendering follows the discussion and requirements from TR26.928, clause 6.2.5. The architecture us shown in Figure 1.    Figure Split Rendering with Asynchronous Time Warping (ATW) Correction  Raster-based split rendering refers to the case where the XR Server runs an XR engine to generate the XR Scene based on information coming from an XR device. The XR Server rasterizes the XR viewport and does XR pre-rendering.  According to Figure Figure 1, the viewport is pre-dominantly rendered in the XR server, but the device is able to do latest pose correction, for example by asynchronuous time-warping (see clause 4.1) or other XR pose correction to address changes in the pose.  - XR graphics workload is split into rendering workload on a powerful XR server (in the cloud or the edge) and pose correction (such as ATW) on the XR device  - Low motion-to-photon latency is preserved via on device Asynchronous Time Warping (ATW) or other pose correction methods.  The following call flow highlights the key steps:  1) An XR Device connects to the network and joins XR application  a) Sends static device information and capabilities (supported decoders, viewport)  2) Based on this information, the XR server sets up encoders and formats  3) Loop  a) XR Device collects XR pose (or a predicted XR pose)  b) XR Pose is sent to XR Server  c) The XR Server uses the pose to pre-render the XR viewport  d) XR Viewport is encoded with 2D media encoders  e) The compressed media is sent to XR device along with XR pose that it was rendered for  f) The XR device decompresses video  g) The XR device uses the XR pose provided with the video frame and the actual XR pose for an improved prediction using and to correct the local pose, e.g. using ATW.  According to TR 26.928, clause 4.2.2, the relevant processing and delay components are summarized as follows:   * **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. 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 such a player response. * **Age of content** is defined as the time duration between the moment a content is created and the time it is presented to the user. In the context of gaming, this is the time between the creation of a video frame by the game engine and the time at which the frame is finally presented to the player.   The **roundtrip interaction delay** is therefore the sum of the *Age of Content* and the *User Interaction Delay*. If part of the rendering is done on an XR server and the service produces a frame buffer as rendering result of the state of the content, then for raster-based split rendering (as defined in clause 6.2.5) in cloud gaming applications, the following processes contribute to such a delay:   * User Interaction Delay (Pose and other interactions)   + capture of user interaction in game client,   + delivery of user interaction to the game engine, i.e. to the server (aka network delay),   + processing of user interaction by the game engine/server, * Age of Content   + creation of one or several video buffers (e.g. one for each eye) by the game engine/server,   + encoding of the video buffers into a video stream frame,   + delivery of the video frame to the game client (a.k.a. network delay),   + decoding of the video frame by the game client,   + presentation of the video frame to the user (a.k.a. framerate delay).   As ATW is applied the motion-to-photon latency requirements (of at most 20 ms) are met by XR device internal processing. What determines the network requirements for split rendering is time of pose-to-render-to-photon and the roundtrip interaction delay. According to clause TR 26.928, clause 4.5, the permitted downlink latency is typically 50-60ms. |
| **Categorization** |
| **Type: XR, Cloud Computing, GPU**  **Delivery: Interactive, Split, Gaming**  **Device: Phone, HMD, Glasses** |
| **Preconditions** |
| On the device, a gaming application may be installed  On the network, an XR Server is installed, that runs a gaming application as well GPU based rendering and an encoding. |
| **Requirements in terms of Capabilities and QoS/QoE Considerations** |
| <provides a summary on potential requirements as well as considerations on KPIs/QoE as well as QoS requirements> |
| **Feasibility and Industry Practices** |
| <How could the use case be implemented based on technologies available today or expected to be available in a foreseeable timeline, at most within 3 years?  - What are the technology challenges to make this use case happen?  - Do you have any implementation information?  - Demos  - Proof of concept  - Existing services  - References  - Could a reduced experience of the use case be implemented in an earlier timeframe or is it even available today?  >  Steam  Boundless XR: https://zerolight.com/de/news/press-releases/worlds-first-boundless-xr-over-5g-retail-experience |
| **Cost Analysis** |
| < How does the use case scale with an increasing number of users? > |
| **Potential Standardization Status and Needs** |
| <identifies potential standardization needs> |

## 5.2.3 User-generated live streaming

|  |
| --- |
| **Use Case Name** |
| User-generated live streaming |
| **Description** |
| A social influencer starts a live captured media session similar to Facebook Live and publishes the content through 5G Media Uplink Streaming. The content is then distributed live to several or many viewers through 5G Media Downlink Streaming.  The application provides several usage scenarios with various configuration options which may change during the live session:   1. The social influencer may:    1. be static, occasionally moving or highly mobile, for example in a vehicle, on skis or on a bicycle,    2. produce different quality of content, depending on lighting conditions, speed, as well as based on the quality of the camera and the available uplink bandwidth,    3. produce highly-valuable content that requires extra content protection,    4. want to capitalize on the stream by allowing ad-insertion in the content (targeted pre-roll and/or mid-roll). 2. The viewers may be quite diverse and changing because:    1. they may be dynamically joining or leaving the live stream,    2. their number might be just a few or they may quite many, in range of tens of thousands or more, for example for a popular influencer,    3. they may be geographically spread with different densities in various areas, and their densities may change during the session,    4. they may consume the service on different devices, for example on 4K TV sets, mobile phones, in-car receivers or tablets, with different operating systems, DRM capabilities as well as different codec hardware capabilities,    5. some of them may be mobile, i.e on a car or public transport,    6. they may react (smilies, comments, likes, audio dubs, images, avatars, and animations) to the content or previous reactions by the viewers who watched the content earlier. 3. The service requirements may be quite different. The content may need to be:    1. available for live and/or on-demand consumption,    2. only available for consumption after the end of the live uplink session, i.e. it is uploaded entirely before being made available to followers,    3. available with a required target latency with ranges in between capture and display of as low as 1 second up to several tens of seconds,    4. dubbed into the same or different language,    5. provided with automatic extraction and addition of captions/subtitles from the audio,    6. post-processed to improve the audio and/or visual quality,    7. processed by adding overlays and content tags and other augmented material,    8. indexed, including the addition of thumbnail navigation in real time,    9. provided to regional proxies with specific metadata such as black-out information, ad insertion opportunities, language settings or other service metadata,    10. profanity checked and appropriately altered before being distributed,    11. available for viewing for some time period (from a few minutes to forever) after the end of the live session.   In a simple reference scenario of 5G Media Uplink and Downlink Streaming, the following aspects are supported:   1. An application manages the service. The server application may be run by external application providers, by MNO, or by a joint collaboration of application provider and MNO. 2. Uplink streaming is provided through 5G media uplink streaming to the application. 3. The application may perform one or more of the following processes:    1. It decodes the received content    2. It applied the various processes the content such as:       1. Upscaling       2. Light correction       3. Stabilization       4. Dubbing       5. Captioning       6. Overlaying and tagging       7. Indexing       8. Navigation improvements.    3. It encodes the uploaded content to 5GMS downlink streaming formats.    4. It packages the content and adds appropriate ad metadata.    5. It applies content protection and DRM. 4. As soon as the content becomes available, the server application uses downlink streaming for distribution.    1. It provisions a 5GMS downlink streaming service.    2. It ingests the content into a 5GMS streaming service.    3. It gets feedback from consumption reporting on what content is consumed and may change the encoded streaming formats.   The following aspects may be critical for the service:   1. Can the service be provided throughout the session without any interruption considering the dynamic aspect of the service? 2. Can the desired end-to-end latency be met with the reference scenario and if not, what are possible ways to realize this? 3. Is the content generation flexible and fast enough to address different user population, yet highly utilized, i.e. the variations of the content is consumed by one or more viewers and is shared as much as possible by many viewers? 4. Is there a benefit to push certain processing closer to the influencer, certain viewers, or in-between (in terms of bandwidth, latency, and processing requirements)? |
| **Categorization** |
| **Type: 2D**  **Delivery: 5GMS uplink streaming, 5GMS downlink streaming, live media streaming**  **Device: Phone, HMD, Glasses** |
| **Preconditions** |
| On the sending UE, 5GMS Uplink streaming is provided.  On every receiving UE, a 5GMS downlink streaming client is provided.  An Application Service Provider supports network interfaces for 5GMS Uplink Streaming (M1u and M2u) as well as 5GMS Downlink Streaming (M1d and M2d).  Extended functionalities in uplink streaming and downlink streaming may be supported such that the Application Service Provider can delegate certain tasks to the 5G Media Streaming System. |
| **Requirements in terms of Capabilities and QoS/QoE Considerations** |
| Potential new required capabilities of 5G Media Streaming System:   1. Decoding of content received through uplink streaming 2. Processing of the content with functionalities such as    1. Upscaling    2. Light correction    3. Stabilization    4. Dubbing    5. captioning    6. Etc. 3. Transcoding of the uploaded and processed content to 5GMS downlink streaming formats 4. Packaging/Encrypting the content and adding appropriate ad metadata   Relevant KPIs:   1. Subjective/objective quality of the encoded streams in terms of bitrates/quality. 2. Latency for the end to end service. 3. Latency of each processing step. 4. Seamless integration of ads with the main content. 5. Scalability and handling of different number of devices. 6. Quality of the delivered service in terms of streaming metrics. 7. Computational efficiency in terms of:    1. Encoding and Decoding workflow    2. Encryption and D    3. Manifest generation    4. Target-based advertising |
| **Feasibility and Industry Practices** |
| Today, the above processing may be achieved by running cloud processing on the application server for all the functionalities mentioned above. Alternatively, third-party cloud-based transcoding services exist.  The number of streams, support of multiple codecs, the amount of delay, the number of representations in the encoding ladder varies depending on the service.  Codec support evolves with the availability of new codecs.  Services like Facebook Live encode the content on demand and on-the-fly.  Services like Youtube Live provide channels for professional content providers. They can insert advertisement clips, but the ad insertion occurs using two video elements.  Services like TikTok allow the user to add a lip-sync to a song, to add his/her reactions on a small window over the video, and to make short movies using filter and speed change of the video. The user also can capture a video next to a pre-recorded video. |
| **Nominal Cost Analysis** |
| The cost of service increases linearly with the number of ingest streams.  The cost of service increases less than linearly with the number of download streaming clients as the encoding and caching requirement will be common with the large number of viewers.  Cost analysis needs to take into account if there a benefit or necessity to push certain processing closer to the users/edge (in terms of bandwidth, latency and processing requirements)? |
| **Benefits and Impact** |
| Potential Benefits and Impacts when running services within 5GMS network   1. tbd   Potential Benefits and Impacts when running services on the edge   1. tbd |
| **Potential Technical Requirements** |
| Potential Requirements:   1. Establish workflows as defined in the use case with the addition of service parameters. 2. Sistribute the tasks across 5G System and 5GMS components. |
| **Potential Standardization Status and Needs** |
| 1. tbd |

## 5.2.4 Augmented Video Streaming

|  |
| --- |
| **Use Case Name** |
| Augmented video streaming |
| **Description** |
| Augmented video streaming will help to understand the video and enhance the interest of the video. In live streaming scenes, it can analyze the video content with AI technology, refine the video information, understand the video content and enhance the video quality, such as improving the resolution and color. It may also superimpose AR effect to tag the identified video information interestingly, such as the AR effect of the pass path in a football game or the AR effect on human face or background during streaming, etc. Processing at the edge can reduce the time required for video enhancement processing and network transmission.   1. Augmented video streaming with AI.   Due to limited shooting conditions or the need for new video features, augmented video streaming with AI can provide a service that edge nodes is used to enhance the captured video to improve the user experience.  1) An encoded video streaming is compressed and uploaded to an edge node, and the video resolution, color gamut and other parameters should be specified in configuration.  2) The edge node server decodes the video streaming, then with the installed AI capabilities, sets up the enhancement functions, such as resolution enhancement, color enhancement, signal-to-noise ratio enhancement, video feature recognition, subtitle, etc., for video editing.  3) Augmented Video is re-encoded and distributed to users.   1. Augmented video streaming with AR.   It can provide AR experience to its users.  1) An encoded video streaming is compressed and uploaded to an edge node.  2) The edge node server decodes the video streaming, identifies and understands the video features with AI capability, and loads the corresponding AR special effects on the understanding results. The AR special effects are overlapped with the video streaming.  3) The edge node distributes it to users.   1. Augmented video streaming with interaction.   AI augmented content may consider the user’s preferences. AR augmented content will output according to the user's FOV. Therefore, the workflow of the edge node can involve interactions. According to the user's preference, the edge node server selects the corresponding augmented video streaming to distribute, or processes the video according to the user's FOV.    When there is a high requirement for QoE of enhanced video, the terminal device cannot meet this requirement and cloud or edge servers are needed for processing. And in the case of local video streaming transmission, the video streaming can be processed directly at the edge. Compared with uploading to the cloud-based processing, it can reduce roundtrip time delay (RTT) of network and avoid the waste of resources on the core network.  For AI video enhancement, users can choose the function of the video enhancement. If the AI video enhancement processing functions are deployed on the cloud, downlink transmission needs to involve multi-video streams with different enhancement effects for users to choose from. The pressure of the downlink will be multiplied. Therefore, it is better for edge nodes to support different video enhancement processing.  For AR effects enhancement, users need to use AR devices to watch enhanced AR video and have higher requirement on delay, like MTP (Motion To Photons) latency requirement. It is necessary for split rendering of edge processing to ensure this. |
| **Categorization** |
| **Type: XR, AI, Cloud Computing**  **Delivery: Interaction**  **Device: Phone, HMD, TV** |
| **Preconditions** |
| On the device side, a video application or a browser supporting XR player with interaction is needed.  On the network side, a video server is installed for video data structuring process and its key techniques like deep learning, big data technology and cloud storage. Edge nodes need to store some AR effects in advance. |
| **Requirements in terms of Capabilities and QoS/QoE Considerations** |
| <provides a summary on potential requirements as well as considerations on KPIs/QoE as well as QoS requirements>   * Requirements   + Real-time adaptation of encoding and transcoding   + Overlay graphics.   + CDN and content caching   + 6DOF interaction * KPI   + - * + MTP         + Operation response time         + The number of users |
| **Feasibility and Industry Practices** |
| <How could the use case be implemented based on technologies available today or expected to be available in a foreseeable timeline, at most within 3 years?  - What are the technology challenges to make this use case happen?  - Do you have any implementation information?  - Demos  - Proof of concept  - Existing services  - References  - Could a reduced experience of the use case be implemented in an earlier timeframe or is it even available today?  >  Intel edge computing solutions  At the 2018 China Mobile Global Partner Conference, China Mobile and Intel jointly released a set of video processing unit and intelligent video edge computing solution. <https://www.intel.sg/content/www/xa/en/cloud-computing/visual-cloud.html>   * Media Analytics      * AR solutions |
| **Cost Analysis** |
| < How does the use case scale with an increasing number of users? > |
| **Potential Standardization Status and Needs** |
| <identifies potential standardization needs> |

## 5.2.5 Generalized Split and Cloud Rendering and Processing

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Use Case Name** | | | | | |
| Generalized Split and Cloud Rendering and Processing | | | | | |
| **Description** | | | | | |
| See several use cases in TR 26.928 should be added to the already agreed split rendering use case by reference to TR 26.928. | | | | | |
| 3 | Streaming of Immersive 6DoF | VR | 3DoF+, 6DoF | Streaming Interactive Split | HMD with a controller |
| 4 | Emotional Streaming | 2D, AR and VR | 2D, 3DoF+, 6DoF | Streaming Interactive, Split | Phone and HMD |
| 5 | Untethered Immersive Online Gaming | VR | 6DoF | Streaming, Interactive, Split | HMD with a Gaming controller |
| 6 | Immersive Game Spectator Mode | VR | 6DoF | Streaming, Split | 2D screen or HMD with a controller |
| 21 | Immersive 6DoF Streaming with Social Interaction | VR and Social VR | 3DoF+, 6DoF | Streaming Interactive Conversational Split | XR5G-V3  XR5G-V4 |
| 22 | 5G Online Gaming Party | VR | 6DoF | Streaming, Interactive, Split, D2D | XR5G-V3  XR5G-V4 |
| 23 | Spatial Shared Data | AR | 6DoF | Streaming Interactive Conversational Split | XR5G-AX |
|  | | | | | |
| **Categorization** | | | | | |
| See several use cases in TR 26.928 | | | | | |
| **Preconditions** | | | | | |
| See several use cases in TR 26.928 | | | | | |
| **Requirements in terms of Capabilities and QoS/QoE Considerations** | | | | | |
| See several use cases in TR 26.928 | | | | | |
| **Feasibility and Industry Practices** | | | | | |
| See several use cases in TR 26.928 | | | | | |
| **Nominal Cost Analysis** | | | | | |
| See several use cases in TR 26.928 | | | | | |
| **Benefits and Impact** | | | | | |
| See several use cases in TR 26.928 | | | | | |
| **Potential Technical Requirements** | | | | | |
| * See several use cases in TR 26.928 | | | | | |
| **Potential Standardization Status and Needs** | | | | | |
| The following information should be added to the agreed use case.  Edge/Cloud processing and rendering is a promising technology to support online gaming in power- and resource constrained devices. Relevant aspects for generalized cloud/split rendering include:   * A generalized XR cloud and split rendering application framework based on a scene description * Support for 3D formats in split and cloud rendering approaches * Formats and protocols for XR Pose information delivery and possibly other metadata in the uplink at sufficiently high frequency * Content Delivery protocols that support generalized split/cloud rendering * Distributions of processing resources across different resources in the 5G system network, in the application provider domain (cloud) and the XR device. * Supporting the establishment of Processing Workflows across distributed resources and managing those * 5QIs and other 5GS/Radio capabilities that support generalized split/cloud rendering by coordination with other groups * Edge computing discovery and capability discovery based on work in SA2 and SA6 (see clause 4.3.6) | | | | | |

## 5.2.6 Photo-realistic AR Rendering in Network

|  |
| --- |
| **Use Case Name** |
| Photo-realitic AR Rendering in Network |
| **Description** |
| Bob accesses wants to position a complex 3D object with significant amount of data (materials, roughness, texture, transparency) and place it into a real-world environment using an AR device. For this purpose, he uses a local rendering engine, but he is very unsatisfied as natural lights are not well handled and the object looks unrealistic. A cloud service offers a PBR-based cloud rendering, and he gives it a try. He is really happy with the rendering, he gets sparkling effects and starts to move around the object and wants to put it into different angles and positions. However, moving, rotating, interacting the object makes him annoyed again as the latency of the interaction is only executed after several hundred of milliseconds.  However, he gets the offering from his 5G Operator that he can get cloud-based rendering in the edge, promising latencies that makes the 3D objects *present* and *interactive*.  The 5G Operator permits to ingest popular 3D objects into the network and provides an improved rendering for their consumers. Parts of the objects may be static, others may be timed. In certain cases, even several objects may be rendered at the same time.  Example:    <https://docs.microsoft.com/en-us/azure/remote-rendering/overview/features/pbr-materials> |
| **Categorization** |
| **Type: AR, 3D**  **Delivery: Split rendering, edge rendering**  **Device: Phone, AR devices** |
| **Preconditions** |
| On the UE, a 5G modem is available, an AR rendering and a depth camera is available.  In the network, advanced rendering functionalities are available that are accessible with the 5GS and 5G Edge enablers and also a data storage to host complex 3D objects.  An Application Service Provider supports network interfaces for ingesting popular 3D objects into the network. |
| **Requirements in terms of Capabilities and QoS/QoE Considerations** |
| * Potential new required capabilities of 5G Media Streaming System   + Fast uplink streaming of natural environments including protocols and formats to included depth and light conditions.   + Discovery and establishment of photo-realistic rendering functionalities that can work with uplink camera information.   + Fast Downlink streaming to support interaction with objects including new light conditions, etc.   + Formats and protocols that allow post-processing to the latest pose * Relevant KPIs   + Quality of rendered scene   + Immersiveness and presence of the scene   + Latency of each processing step   For many details, please refer to TR 26.928. |
| **Feasibility and Industry Practices** |
| tbd |
| **Nominal Cost Analysis** |
| tbd |
| **Benefits and Impact** |
| Potential Benefits and Impacts when running services within 5GMS network   * tbd   Potential Benefits and Impacts when running services on the edge   * tbd |
| **Potential Technical Requirements** |
| Potential Requirements   * tbd |
| **Potential Standardization Status and Needs** |
| * tbd |
|  |

## 5.2.7 Media Services in the Edge

|  |
| --- |
| **Use Case Name** |
| Media Services in the Edge |
| **Description** |
| In the media production and distribution, the use of cloud services get more and more popular. Examples for standardized approaches are for example documented here:   * <https://www.fims.tv/>, * <https://tech.ebu.ch/groups/mcma> * <https://www.smpte.org/microservices-media-and-entertainment-key-benefits-and-challenges>   These applications include Artificial Intelligence (AI) enabled services, such as: speech-to-text, automatic subtitling, translation, text-to-speech, face recognition and identification, emotion detection, speech-to-text evaluation, FFMPEG media transformation.  Is there a benefit to provide some applications on the edge rather than the cloud? |
| **Categorization** |
| **Type: 2D**  **Delivery: Cloud and edge processing**  **Device: phone, TV, tablet** |
| **Preconditions** |
| * Potential new required capabilities of 5G Media Streaming System   + Ingest   + Distribution * Relevant KPIs |
| **Requirements in terms of Capabilities and QoS/QoE Considerations** |
| * Potential new required capabilities of 5G Media Streaming System   + tbd * Relevant KPIs   + tbd |
| **Feasibility and Industry Practices** |
| See above links |
| **Nominal Cost Analysis** |
| tbd |
| **Benefits and Impact** |
| Potential Benefits and Impacts when running services within 5GMS network   * tbd   Potential Benefits and Impacts when running services on the edge   * tbd |
| **Potential Technical Requirements** |
| Potential Requirements   * tbd |
| **Potential Standardization Status and Needs** |
| * tbd |

## 5.2.8 Partial delivery of 3D content (point cloud, mesh) for AR/MR device

|  |
| --- |
| **Use Case Name** |
| Partial delivery of 3D content (point cloud and mesh) for AR/MR device |
| **Description** |
| Service provider provides various types of contents to be overlaid over real environment for augmented reality experience. Certain 3D content is huge as the earth, which is called digital twin of past, current and future of the world. The content is live and continuously updated even though it is past. Therefore, delivering one entire content doesn’t make any sense nor possible. In this case, a partial delivery of 3D content serves user’s demand. For example, user wants to travel oversea. The user can select places to stay and place to visit, then digital twin of hotel room with a view and sightseeing points will be partially delivered to the user. More practically, only the viewport can be delivered then rendered upon user’s position and view direction.  Another example is as follows. Certain 3D content has hundreds of millions of vertexes and faces to provide the finest detail as much as possible. For online catalogs of cars, selecting options to try a look and feel of each individual order is always enjoyable. User may hear engine sound and touch dashboard through the playback of such content. Since the level of details for rendering of the vertexes depends on a distance from user’s viewpoint, proper level of density and the only vertexes related with the view can be delivered. The 3D content is encapsulated for the purpose, and MEC does the required post processing to generate partial view then encoding the content. |
| **Categorization** |
| **Type: AR, MR, VR, XR, 2D, 3D, Cloud Computing**  **Delivery: Downlink, Interactive, Live streaming, On Demand Streaming, Conversational**  **Device: Phone, HMD, Glasses, Laptop** |
| **Preconditions** |
| On the content provider side, delivery or live streaming of 3D content are available.  On the UE side, consumption of 3D content and reporting of relevant viewport information are available.  On the network side, an MEC server is installed. Performing partial extraction, encapsulation and encoding of the 3D content are available. |
| **Requirements in terms of Capabilities and QoS/QoE Considerations** |
| <provides a summary on potential requirements as well as considerations on KPIs/QoE as well as QoS requirements>  Potential requirements   * Real-time partial extraction and control level of details of 3D content * Point cloud and mesh encoding/decoding * UE viewport information report   KPI   * The number of voxels and faces per second for partial extraction * The number of voxels and faces per second for encoding |
| **Feasibility and Industry Practices** |
| <How could the use case be implemented based on technologies available today or expected to be available in a foreseeable timeline, at most within 3 years?  - What are the technology challenges to make this use case happen?  - Do you have any implementation information?  - Demos  - Proof of concept  - Existing services  🡪 VR and OMAF have similar cases on sending the only part that user is viewing. For AR and 3D content, there is no such service announced so far. However, we believe it is a matter of maturity of AR ecosystem.  - References  Could a reduced experience of the use case be implemented in an earlier timeframe or is it even available today?  > |
| **Cost Analysis** |
| < How does the use case scale with an increasing number of users? >  Similar with split rendering since position and direction of user’s viewport would require different part of the 3D content. |
| **Potential Standardization Status and Needs** |
| <identifies potential standardization needs>  How to exchange UE viewport information in 3D space.  How to define and signal delivered part out of the total 3D content space.  How to measure MEC performance index of the partial extraction, and quality of level of details. |

# 6 Potential 5GMS Architecture Extensions

## 6.1 General

Editor’s Note: This clause will document the identified gaps. Identified gaps may go beyond the scope of the SA4 work, in which case the responsible groups will be contacted.

Editor’s Note: The potential extensions are mainly based on the identified relevant use cases in section 5.

## 6.2 Common Extensions

## 6.3 Extensions for Downlink

## 6.4 Extensions for Uplink

Annex <A> (normative):  
<Normative annex for a Technical Specification>

Annex <X> (informative):  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Change history** | | | | | | | |
| **Date** | **Meeting** | **TDoc** | **CR** | **Rev** | **Cat** | **Subject/Comment** | **New version** |
|  |  |  |  |  |  |  |  |