3GPP TR 26.891 V16.0.0 (2018-12)

Technical Specification

3rd Generation Partnership Project;

Technical Specification Group Services and System Aspects;

5G enhanced mobile broadband;

Media distribution

(Release 16)

** 

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.

Keywords

5G enhanced mobile broadband, Media distribution

***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.

© 2018, 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 [5](#__RefHeading___Toc531432532)

Introduction [5](#__RefHeading___Toc531432533)

1 Scope [6](#__RefHeading___Toc531432534)

2 References [6](#__RefHeading___Toc531432535)

3 Definitions and abbreviations [7](#__RefHeading___Toc531432536)

3.1 Definitions [7](#__RefHeading___Toc531432537)

3.2 Abbreviations [7](#__RefHeading___Toc531432538)

4 5G System Overview [8](#__RefHeading___Toc531432539)

4.1 Introduction [8](#__RefHeading___Toc531432540)

4.2 Network Functions and Interfaces [8](#__RefHeading___Toc531432541)

4.3 Non-Roaming Architecture [11](#__RefHeading___Toc531432542)

4.4 Roaming Architecture [13](#__RefHeading___Toc531432543)

4.5 High Level Features of 5G System for Media Delivery [14](#__RefHeading___Toc531432544)

4.5.1 Introduction [14](#__RefHeading___Toc531432545)

4.5.2 Usage of UL Classifiers for PDU Session [14](#__RefHeading___Toc531432546)

4.5.3 Support for Local Area Data Network [15](#__RefHeading___Toc531432547)

4.5.4 Application Function Influence on Traffic Routing [15](#__RefHeading___Toc531432548)

4.5.5 Support for Edge Computing [16](#__RefHeading___Toc531432549)

4.5.6 Network slices in the 5G system [16](#__RefHeading___Toc531432550)

4.6 5G QoS Model [17](#__RefHeading___Toc531432551)

4.6.1 Qos Processing [17](#__RefHeading___Toc531432552)

4.6.2 QoS Parameters [18](#__RefHeading___Toc531432553)

5 Mapping of Existing Media Services [18](#__RefHeading___Toc531432554)

5.1 Media Distribution Systems [18](#__RefHeading___Toc531432555)

5.1.1 Introduction [18](#__RefHeading___Toc531432556)

5.1.2 Media Distribution Functions [19](#__RefHeading___Toc531432557)

5.1.2.1 Edge Media Delivery function [19](#__RefHeading___Toc531432558)

5.1.2.2 Delivery Core function [19](#__RefHeading___Toc531432559)

5.1.2.3 Content Origin function [19](#__RefHeading___Toc531432560)

5.1.2.4 Content Preparation [20](#__RefHeading___Toc531432561)

5.1.2.4.1 Live Linear Ingest [20](#__RefHeading___Toc531432562)

5.1.5.4.2 Control [20](#__RefHeading___Toc531432563)

5.1.5.4.3 VOD ingest [20](#__RefHeading___Toc531432564)

5.1.5.4.4 Other files [21](#__RefHeading___Toc531432565)

5.2 Identified Issues in Mapping to 5G System [21](#__RefHeading___Toc531432566)

5.2.1 Network Interface Considerations [21](#__RefHeading___Toc531432567)

5.2.2 5G QoS Considerations for Media Distribution [21](#__RefHeading___Toc531432568)

5.3 Potential Mapping to 5G System [22](#__RefHeading___Toc531432569)

5.3.1 Introduction [22](#__RefHeading___Toc531432570)

5.3.2 PSS Service Mappings [22](#__RefHeading___Toc531432571)

5.3.2.1 Identified PSS Functions [22](#__RefHeading___Toc531432572)

5.3.2.2 Mapping to the 5G system [23](#__RefHeading___Toc531432573)

5.3.3 Evolution of PSS and MBMS towards 5G [24](#__RefHeading___Toc531432574)

6 Mapping of New Media Services and Verticals [25](#__RefHeading___Toc531432575)

6.1 Introduction [25](#__RefHeading___Toc531432576)

6.2 Media Production [25](#__RefHeading___Toc531432577)

6.2.1 Introduction [25](#__RefHeading___Toc531432578)

6.2.2 FLUS Framework Usage in 5G [25](#__RefHeading___Toc531432579)

6.2.2.1 Identified FLUS Functions [25](#__RefHeading___Toc531432580)

6.2.2.2 Mapping to the 5G System [26](#__RefHeading___Toc531432581)

6.3 AR and XR Services – Use Cases and Context [27](#__RefHeading___Toc531432582)

6.3.1 Service Description [27](#__RefHeading___Toc531432583)

6.3.2 Potential Issues on AR/XR/VR (new) [28](#__RefHeading___Toc531432584)

6.3.3 Key Issue and Recommendations [29](#__RefHeading___Toc531432585)

6.4 Conditional and Dynamic Policies for Application Instances [29](#__RefHeading___Toc531432586)

6.4.1 Service Description and new issues [29](#__RefHeading___Toc531432587)

6.4.2 Potential Architecture enhancement [30](#__RefHeading___Toc531432588)

6.4.3 Example procedure for VoD Preloading case [31](#__RefHeading___Toc531432589)

6.4.4 Example procedure for Audio PodCast case (changing priority) [32](#__RefHeading___Toc531432590)

6.4.5 Example procedure for a FLUS session with QoS [32](#__RefHeading___Toc531432591)

6.4.6 Relation to other SA4 Network Assistance Functions [33](#__RefHeading___Toc531432592)

6.4.7 Interface to 3rd Party Service Provider [33](#__RefHeading___Toc531432593)

6.4.8 Summary of Proposals [33](#__RefHeading___Toc531432594)

6.5 Cloud Media Processing in 5G [33](#__RefHeading___Toc531432595)

6.5.1 Introduction [33](#__RefHeading___Toc531432596)

6.5.2 5G System Considerations [34](#__RefHeading___Toc531432597)

7 API Aspects [35](#__RefHeading___Toc531432598)

7.1 Introduction [35](#__RefHeading___Toc531432599)

7.2 Use Cases [35](#__RefHeading___Toc531432600)

7.2.1 Introduction [35](#__RefHeading___Toc531432601)

7.2.2 OTT Service Provider [35](#__RefHeading___Toc531432602)

7.2.3 Browser-based consumption [35](#__RefHeading___Toc531432603)

7.2.4 Ad Insertion [35](#__RefHeading___Toc531432604)

7.2.5 Live Services [35](#__RefHeading___Toc531432605)

7.2.6 VR Services [35](#__RefHeading___Toc531432606)

7.3 Interoperability Considerations [36](#__RefHeading___Toc531432607)

7.3.1 Background [36](#__RefHeading___Toc531432608)

7.3.2 Potential Solutions for Capability Discovery [38](#__RefHeading___Toc531432609)

7.3.2.1 Media Profile [38](#__RefHeading___Toc531432610)

7.3.2.2 3G-FF/ISO BMFF/DASH/CMAF Header/Initialization Segment [39](#__RefHeading___Toc531432611)

7.3.2.3 MIME Sub-parameters [39](#__RefHeading___Toc531432612)

7.3.2.4 Media Capabilities [39](#__RefHeading___Toc531432613)

7.3.2.5 Device Capability – Persistent Item Solution [40](#__RefHeading___Toc531432614)

7.3.2.6 Playback Capabilities [40](#__RefHeading___Toc531432615)

7.3.2.7 MBMS Feature Tags [40](#__RefHeading___Toc531432616)

7.3.2.8 PSS Capability Exchange [40](#__RefHeading___Toc531432617)

7.4 Key Issue and Recommendations [40](#__RefHeading___Toc531432618)

8 Summary and Conclusions [40](#__RefHeading___Toc531432619)

8.1 General Summary [40](#__RefHeading___Toc531432620)

8.1.1 General [40](#__RefHeading___Toc531432621)

8.1.2 Support from AFs [41](#__RefHeading___Toc531432622)

8.1.3 Addressing Application Servers [41](#__RefHeading___Toc531432623)

8.1.4 QoS Management [41](#__RefHeading___Toc531432624)

8.1.5 Other Functions [41](#__RefHeading___Toc531432625)

8.2 Conclusions [42](#__RefHeading___Toc531432626)

Annex A: Change history [43](#__RefHeading___Toc531432627)

# Foreword

This Technical Report 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.

# Introduction

With the introduction of the 5G System, the network core architecture underwent a complete overhaul. The 5G Core has been designed following a Function-based approach with separation of control and user plane. Concepts like network slicing, edge computing, dynamic QoS, northbound APIs, and traffic influencing are a core part of the 5G System.

This Technical Report documents the integration of media distribution services in the 5G System. It is organized in two parts: network aspects and UE API aspects. The network aspects cover the migration of existing media distribution services from LTE to 5G. It also discusses emerging media services and use cases and how they can be realized in 5G. The API section investigates requirements of existing and immersive media services and the related requirements and API needs.

# 1 Scope

The present document studies the evolution of media distribution with the development of 5G. It addresses the networking and the API aspects that effect the integration in the 5G System.

# 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".

[2] 3GPP TS 23.501: "System Architecture for the 5G System".

[3] 3GPP TS 23.502: "Procedures for the 5G System".

[4] ETSI TS 102 796: "Hybrid Broadcast Broadband TV".

[5] 3GPP TS 26.346: "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs".

[6] 3GPP TS 26.234: "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs".

[7] IETF RFC 4337: "MIME Type Registration for MPEG-4".

[8] IETF RFC 6381: "The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types".

[9] ETSI GS MEC 011: "Mobile Edge Computing (MEC); Mobile Edge Platform Application Enablement".

[10] ETSI GS MEC 009: "Mobile Edge Computing (MEC); General Principles for Mobile Edge Service APIs".

[11] 3GPP TR 23.722: "Study on Common API Framework for 3GPP Northbound APIs".

[12] 3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)".

[13] 3GPP TS 23.222: "Functional architecture and information flows to support Common API Framework for 3GPP Northbound APIs; Stage 2".

[14] 3GPP TR 26.918: "Virtual Reality (VR) media services over 3GPP".

[15] 3GPP TR 22.891: "Study on New Services and Markets Technology Enablers".

[16] 3GPP TR 28.803: "Study on management aspects of edge computing (Release 16)".

# 3 Definitions and abbreviations

## 3.1 Definitions

For the purposes of the present document, the terms and definitions 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].

## 3.2 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].

DASH Dynamic Adaptive Streaming over HTTP

HbbTV Hybrid Broadcast Broadband TeleVision

HEVC High Efficiency Video Coding

HTML-5 HyperText Markup Language version 5

HW HardWare

ISO BMFF ISO Base Media File Format

MIME Multipurpose Internet Mail Extensions

OMAF Omnidirectional MediA Formal

UL Up-Link

VR Virtual Reality

ABR Adaptive Bitrate

AMF Access and Mobility Management Function

ANIF Application to Network Interactions Function

API Application Programming Interface

ARP Allocation and Retention Priority

AUSF Authentication Server Function

CAPIF Common API Framework

CDN Content Distribution Network

CDR Call Detail Record

CENC Common Encryption

CMAF Common Media Application Format

DANE DASH-Aware Network Element

DNAI Data Network Access Identifier

DNN Data Network Name

DNS Domain Name Service

DRM Digital Rights Management

DVR Digital Video Recording

EMD Edge Media Delivery

ESAM Event Signaling and Messaging

FEC Forward Error Correction

FLUS Framework for Live Uplink Streaming

FQDN Fully Qualified Domain Name

GBR Guaranteed Bitrate

GPU Graphics Processing Unit

JITP Just In-time Packaging

LADN Local Aread Data Network

MBR Maximum Bitrate

MEC Mobile Edge Computing

MNO Mobile Network Operator

NAS Non-access Stratum

NEF Network Exposure Function

NRF Network Repository Function

NSSAI Network Slice Selection Assistance Information

NSSF Network Slice Selection Function

PCF Policy Control Function

POIS Placement Opportunity Information Service

QFI QOS Flow Identity

RAT Radio Access Technology

SDI Serial Digital Interface

SLA Service Level Agreement

SMF Session Management Function

SSC Session and Service Continuity

UDR Unified Data Repository

UDSF Unstructured Data Storage Function

UPF User Plane Function

# 4 5G System Overview

## 4.1 Introduction

3GPP SA2 has specified the 5G system architecture and procedures in TS 23.501 and TS 23.502 respectively. As specified in TS 23.501, a number of network functions interface with each other to provide connectivity services to a UE. Different network functions play different roles (e.g., control plane, data plane entities) and help setup sessions and route traffic from the UE to the intended destination. 5G system architecture is based on separation of control and user plane thus allowing independent scalability, evolution and flexible deployments e.g. centralized location or distributed (remote) location. Other key principles and concepts based on which the 5G system architecture is being defined are:

- Wherever applicable, define procedures (i.e. the set of interactions between network functions) as services, so that their re-use is possible.

- Enable each network function (NF) to interact with other NF directly if required.

- Minimize dependencies between the Access Network (AN) and the Core Network (CN). The architecture is defined with a converged core network with a common AN - CN interface which integrates different 3GPP and non-3GPP access types.

- Support a unified authentication framework.

- Support "stateless" NFs, where the "compute" resource is decoupled from the "storage" resource.

- Support capability exposure.

- Support concurrent access to local and centralized services. To support low latency services and access to local data networks, UP functions can be deployed close to the Access Network.

- Support roaming with both Home routed traffic as well as Local breakout traffic in the visited PLMN.

## 4.2 Network Functions and Interfaces

Table 1 shows different network functions currently defined in TS 23.501 [2]:

Table 1: Different Network Functions in 5G Architecture

|  |  |
| --- | --- |
| Network Function | Functional Description |
| Authentication Server Function (AUSF) | Provides AUSF authentication server functionality as specified by 3GPP SA3 |
| Access and Mobility Management Function (AMF) | Terminates NAS signalling from the UE and provides registration, connection, reachability, and mobility management between the UE and 5G CN. It also helps with lawful interception, transparent proxy for routing SM messages, access authentication and authorization |
| Unstructured Data Storage network function (UDSF) | An optional function that provides storage and retrieval of information as unstructured data by any NF |
| Network Exposure Function (NEF) | Provides means for exposure of services and capabilities of network functions and users; translation of information exchanged between AF and internal network functions |
| NF Repository Function (NRF) | Provides service discovery functions |
| Network Slice Selection Function (NSSF) | Helps with selection of network slice instances service the UE, determining Allowed NSSAI, and determining the AMF Set to be used to serve the UE |
| Policy Control function (PCF) | Provides policy framework to govern network behaviour and policy rules to control plane functions; implements a front end to access subscription information relevant for policy decisions in UDR |
| Session Management Function (SMF) | Provides session management functionality such as session establishment, modification, and release functionality; UE IP address allocation and management; selection and control of UP function; traffic steering and policy enforcement; downlink data notification; roaming functionality etc. |
| Unified Data Management (UDM) | Supports functionality such as authentication credential handling, user identification handling, access authorization, registration/mobility management, subscription management, and SMS management. |
| Unified Data Repository (UDR) | Supports storage and retrieval of subscription data by the UDM front end; storage and retrieval of policy data by the PCF; storage and retrieval of information as structured data, such as location data and application data (including Packet Flow Descriptions (PFDs) for application detection, application request information for multiple UEs), by the NEF |
| User Plane Function (UPF) | Provides data plane functionality such as anchor point for intra-/inter-RAT mobility, point of interconnect to data network, packet routing and forwarding, packet inspection, traffic usage reporting, uplink classifier and branching point functionality, QoS handling, traffic verification, packet marking, and downlink packet buffering etc. |
| Application Function (AF) | Interacts with 3GPP network to support application influence on traffic routing, accessing Network Exposure Function, and interact with policy framework for policy control |
| SMS Function (SMSF) | Provides functionality to support SMS over NAS such as SMS subscription checking, SMS related CDR, lawful interception etc. |
| Non-3GPP InterWorking Function (N3IWF) | Provides functionality to connect non-3GPP network to 5G CN including IPSec tunnel establishment with the UE, relaying uplink and downlink control-plane NAS signalling between UE and AMF, relaying uplink and downlink user-plane packets between UE and UPF, QoS enforcement for user-plane packets etc. |

The network functions described above interface with each other as shown in the system architecture below using a service based representation (as described in TS 23.501). In service based representation, the interaction between two network functions is viewed as an access of a service provided by one NF to another NF. With this representation, service based interfaces exist between network functions and can be shown as follows (as described in TS 23.501 [2]):



Figure 1: 5G System Architecture Using Service Based Interfaces

Table 2 lists the different service based interfaces defined in TS 23.501 [2]:

Table 2: Service based Interfaces

|  |  |
| --- | --- |
| Interface Name | Description |
| Namf | Service-based interface exhibited by AMF |
| Nsmf | Service-based interface exhibited by SMF |
| Nnef | Service-based interface exhibited by NEF |
| Npcf | Service-based interface exhibited by PCF |
| Nudm | Service-based interface exhibited by UDM |
| Naf | Service-based interface exhibited by AF |
| Nnrf | Service-based interface exhibited by NRF |
| Nnssf | Service-based interface exhibited by NSSF |
| Nausf | Service-based interface exhibited by AUSF |

The network functions within the 5G Control Plane described in Table 1 interact with each other using service-based interfaces described in Table 2. Using its service-based interface, each network function may offer multiple services to other authorized network functions. Network functions may interact with other network functions by either using a request / response model or a subscribe / notify model. Table 3 shows a brief description of services provided by few network functions in the service based reference architecture of Figure 1.

Table 3: Services provided by Different Network Functions

|  |  |  |
| --- | --- | --- |
| Network Function (Service Provider) | Services | Supported Service Operations and Service Consumers |
| AMF | Communication Service | This service enables an NF to communicate with the UE and/or the AN through the AMF. Consumers of such a service include SMF, SMSF, PCF, NEF, Peer AMF |
| Event Exposure Service | This service enables other NFs to subscribe or get notified of the mobility related events and statistics. Consumers of such a service include SMF, NEF, PCF, UDM |
| UDM | Subscriber Data Management | 1. Allow NF consumer to retrieve user subscription data when necessary  2. Provide updated user subscriber data to the subscribed NF consumer  Consumers include AMF, SMF |
| UE context management | 1. provide the NF consumer of the information related to UE's transaction information, e.g. UE's serving NF identity, UE status, etc.  2. allow the NF consumer to register, remove its information for the serving UE in the UDM  Consumers include AMF, SMF, SMSF |
| Authentication | Provide updated authentication related subscriber data to the subscribed NF consumer  Consumer of such information includes AUSF |
|  | |
| PCF | Policy Control | PCF provides all the operations related to policy rule to NF consumers such as AMF SMF |
| Event Notification | PCF provides the policy related information / event to subscribed NF consumer such as NEF, AF |
| Policy Authorization | PCF authorises an AF request and to create policies as requested by the authorised AF for the PDU-CAN session to which the AF session is bound to. Consumers of such information include AF, NEF |
| NRF | Management | Provides support for Discovery of NF, NF services. Known consumers include AMF, SMF, PCF, NEF, NRF, SMSF, AUSF, UDM |
| SMF | PDU Session | This service manages the PDU sessions and uses the policy and charging rules received from the PCF. The service operations exposed by this NF service allows the consumer NFs to handle the PDU sessions. Consumers of this service includes V-SMF, H-SMF, AMF |
| Event Exposure | This service enables other NFs to subscribe or get notified of the mobility related events and statistics. Consumers of such a service include SMF, NEF, PCF, UDM |
| SMSF | SM Service | This service allows AMF to authorize SMS and activate SMS for the served user on SMSF. Consumers of such service include AMF |

## 4.3 Non-Roaming Architecture

In addition to the service based representation as shown in Figure 1, the 5G core network reference architecture can also be shown using a reference point representation as shown in Figure 2.



Figure 2: Non-Roaming 5G System Architecture in reference point representation

Note 1: The network functions UDSF, NEF, NRF, and UDR are not shown in the architecture diagram as any network function can interface with them directly (or through other NFs) and use the services provided by them.

Note 2: The DN node in the above architecture diagram refers to the end data network to which the UE wishes to connect to.

As shown in Figure 2 above, different network functions interface with each other to provide connectivity services to the UE. Table 4 below provides a brief description of the functionality of few significant reference points.

Table 4: Reference Point Functionality

|  |  |
| --- | --- |
| Reference Point | Description |
| N1 | Used for exchanging NAS signalling between the UE and the core network. A separate NAS connection exists for each access the UE uses to connect to the core network. A single NAS connection is used for both Registration Management, Connection Management, and for SM-related procedures for a UE. |
| N2 | Reference point used to connect core network to a standalone 3GPP access or a non-3GPP access (through N3IWF). Used to exchange the above N1 requests and responses between the UE and AMF. |
| N3 | Reference point between the (R)AN node and UPF for transporting user plane traffic both in uplink and downlink direction from/to the UE. |
| N4 | Reference point between SMF and UPF and used for provisioning and configuring UPF network functions for data plane. Session requests and responses are exchanged in between SMF and UPF for exchanging data plane configuration information. |
| N5 | Reference point between AF and PCF and used by application functions to configure policy and traffic routing configuration so PDU sessions can be controlled as desired |
| N7 | Used by SMF and PCF for exchanging policy and QoS information to control PDU sessions from/to the UE. |
| N8 | Reference point between AMF and UDM to exchange information such as subscription profiles, service area restrictions, Subscribed S-NSSAIs etc. for use during registration management and connection management. |
| N10 | Reference point between SMF and UDM to exchange SMF level subscription data, internal group identification information, supported SSC modes and default SSC mode, default 5QI/ARP (to set QoS parameters for QoS flow) etc. This information along with policy and QoS information from PCF helps SMF configure user plane parameters for PDU sessions. |
| N11 | Reference point between AMF and SMF and is used to carry SM-related NAS messages from/to the UE. |

## 4.4 Roaming Architecture

TS 23.501 also defines roaming reference architectures using service based and reference point representations. There are two kinds of roaming scenarios that are defined in TS 23.501 – local break out and home-routed scenarios. In local break out scenario, when a UE is roaming and in the VPLMN, the PDU session traffic from the UE can be routed to DNN by the VPLMN. However, for home routed scenario, the PDU session traffic from the UE are routed to the DNN through the HPLMN. Figure 3 shows the roaming architecture for local breakout scenario with AF in VPLMN in service based representation.



Figure 3: Roaming 5G System architecture- local breakout scenario with AF in VPLMN in service-based interface representation

In Figure 3, the AF is inside the VPLMN. It is also possible that the AF can be inside the HPLMN. Figure 4 shows the roaming architecture for local breakout scenario with AF in HPLMN in service based representation.



Figure 4: Roaming 5G System architecture- local breakout scenario with AF in HPLMN in service-based interface representation

Figure 3 and Figure 4 show the roaming architecture for local break out scenario. Figure 5 below shows the roaming architecture for the home routed scenario in service-based interface representation.



Figure 5: Roaming 5G System architecture - home routed scenario in service-based interface representation

In addition to the service-based representation, TS 23.501 provides 5G system roaming architectures for both the local breakout and home routed scenarios using the reference point representation. In addition, TS 23.501 also describes roaming architectures for both the local breakout and home routed scenarios when the UE accesses the 5G core network using untrusted Non-3GPP access.

## 4.5 High Level Features of 5G System for Media Delivery

### 4.5.1 Introduction

This clause describes few of the high level features of the 5G system that are significant for media delivery and transport.

### 4.5.2 Usage of UL Classifiers for PDU Session

The Session Management Function (SMF) may insert an UL CL (uplink classifier) in the data path of the PDU session. The UL CL is a functionality supported by an UPF that aims at diverting (locally) some traffic matching traffic filters provided by the SMF. The SMF may dynamically insert and remove an UL CL inside a UPF for a PDU session. The SMF may include more than one UPF supporting the UL CL functionality in the data path of a PDU session.

When a UPF with UL CL is inserted in the data path of the PDU session, the UL CL forwards PDU session traffic to different PDU session anchors of the PDU session. Each PDU session anchor provides a different access to the same DNN for the PDU session. As a result, the UL CL provides forwarding of uplink traffic to multiple PDU session anchors and merge of downlink traffic from multiple PDU session anchors.

The insertion of an UPF with a UL CL in the data path of a PDU session is depicted in Figure 6.



Figure 6: User plane Architecture for the Uplink Classifier

### 4.5.3 Support for Local Area Data Network

As described in TS 23.501, LADN provides local access to a DN (data network) for a PDU session from the UE. The availability of LADN is informed to the UE by the network and the UE is allowed to access the LADN only if the UE is in the service area of the LADN. The LADN service areas are managed by the network and the AMF inside the 5G core network keeps track of the mobility information of the UE. The SMF gets to know about the UE location information from the AMF and LADN service area, and it allows the UE to use the LADN as long as it is in the LADN service area.

The UE can request a PDU session to the LADN and perform session establishment procedures as described in TS 23.502. However, whenever the UE leaves the LADN service area, the UE may not release the PDU session and use the same PDU session when it gets back into the LADN service area.

### 4.5.4 Application Function Influence on Traffic Routing

The Application Functions (AF) may influence SMF routeing decisions made by the SMF for traffic of PDU sessions. A trusted AF interacts directly with other network functions inside the network. However, for an untrusted AF, all traffic from the AF goes through the NEF before reaching the network functions inside the network.

The requests from the AF are sent to the PCF which translates AF requests into policies which are then configured in other network functions as appropriate and ultimately applied to the PDU sessions.

An Application Function may:

- Influence UPF (re)selection and allow routeing user traffic to a local access to a Data Network.

- Issue requests on behalf of applications not owned by the PLMN serving the UE.

- Be in charge of the (re)selection or relocation of the applications within the local DN.

The requests from AF to influence routeing decisions may contain at least:

- Information to identify the traffic to be routed (e.g., slicing information, AF-Service-Identifier).

- Information about the N6 traffic routing requirements for traffic identified. This is provided in the form of a list of routing profile IDs, corresponding each to a DNAI. Based on the routing profile ID the PCF determines traffic steering policy IDs sent to SMF that each corresponds to a steering behaviour which is preconfigured on the SMF or UPF.

- Potential locations of applications towards which the traffic routing should apply. The potential location of application is expressed as a list of DNAI(s).

- Information on the UE(s) whose traffic is to be routed.

- Information on when (temporal validity condition) the traffic routing is to apply.

- Information on where (spatial validity condition) the UE(s) are to be when the traffic routing applies.

- AF subscription information to certain events such as notifications about UPF path management events.

Based on the information received from the AF, the SMF may use that information to:

- (re)select UPF(s) for PDU sessions

- activate mechanisms for traffic multi-homing or enforcement of an UL Classifier (UL CL)

- inform the Application Function of the (re)selection of the UP path

### 4.5.5 Support for Edge Computing

TS 23.501 describes 5G system support for edge computing by describing how 3rd party services can be hosted closer to the edge of the core network and the UE. The 5G system provides different functionality to support edge computing such as the following:

- User plane (re)selection

- Local routing and traffic steering

- Session and service continuity

- AF influence on UPF (re)selection and traffic routing

- Network capability exposure

- QoS and charging

- Support for Local Area Data Network (LADN)

Edge Computing offers application developers and content providers cloud-computing capabilities and an IT service environment at the edge of the network. This environment is characterized by ultra-low latency and high bandwidth as well as real-time access to radio network information that can be leveraged by applications. Operators can open their Network edge to authorized third-parties, allowing them to flexibly and rapidly deploy innovative applications and services towards mobile subscribers, enterprises and vertical segments.

3GPP is evaluated ETSI MEC APIs (application enablement aspects in ETSI GS MEC 011 [9]) and API principles in ETSI GS MEC 009 [10]) specifications) in the 3GPP environment, as documented in 3GPP TR 23.722 [11] on 'Study on Common API Framework for 3GPP Northbound APIs'. 3GPP TS 23.222 [13] describes the relationship between ETSI MEC APIs and Common API Framework for 3GPP Northbound APIs (CAPIF).

### 4.5.6 Network slices in the 5G system

5G systems introduce the concept of network slice, i.e. a set of network resources for usage from a certain service [2]. Currently, TS 23.501 specifies 3 types of network slices, as mentioned below:

Table 5: Standardised Slice/Service Type (SST) values

|  |  |  |
| --- | --- | --- |
| Slice/Service type | SST value | Characteristics |
| eMBB (enhanced Mobile Broadband) | 1 | Slice suitable for the handling of 5G enhanced Mobile broadband, useful, but not limited to the general consumer space mobile broadband applications including streaming of High Quality Video, Fast large file transfers etc. It is expected this SST to aim at supporting High data rates and high traffic densities as outlined in Table 7.1-1 "Performance requirements for high data rate and traffic density scenarios" in TS 22.261. |
| URLLC (ultra- reliable low latency communications) | 2 | Supporting ultra-reliable low latency communications for applications including, industrial automation, (remote) control systems.  This SST is expected to aim at supporting the requirements in Table 7.2.2-1 "Performance requirements for low-latency and high-reliability services." in TS 22.261 related to high reliability and low latency scenarios. |
| MIoT (massive IoT) | 3 | Allowing the support of a large number and high density of IoT devices efficiently and cost effectively. |

Support for all these types is not mandatory, and TS 23.501 [2] does not mandate the exclusion of additional slice types.

Additionally, for a specific network slice type, multiple network slice instances may be offered in an operator's 5G network, and there is no normative specification on the number of network slice instances that may be provided, nor on which applications should use which network slice instances. TS 23.501 [2] further asserts that a network slice instance for a specific UE's application is selected by the NSSF on the basis of the following information provided by the UE to the 5G network during registration: (i) a SST and, optionally (ii) a Slice Differentiation (SD), indicating a specific slice instance. This information may be pre-configured in the UE or may have been obtained by the UE during a previous registration process with the network.

## 4.6 5G QoS Model

### 4.6.1 Qos Processing

The 5G core network applies QoS rules on QoS Flows as defined in [2]. As part of a PDU session, a QoS Flow is identified by a unique QoS Flow ID (QFI) in the 5G System. All User Plane traffic within a PDU session with the same QFI will receive the same QoS treatment, i.e. traffic forwarding, scheduling, and admission control. All QoS Flows are controlled by the SMF. A QoS Flow can be pre-configured, established during the PDU session establishment procedure, or by the PDU session modification procedure.

For handling of uplink traffic, the UE follows QoS rules to classify and mark the packets. The QoS rules may be explicitly provided to the UE as part of the PDU session establishment or modification procedures or they may be deduced through reflective QoS.

The SMF assigns a QFI to a QoS Flow and derives its QoS profile and the QoS rules from the policies provided by the PCF. The UPF then ensures the user plane traffic on the downlink is mapped correctly and the UE performs similar functionality for the uplink traffic.

To identify the traffic, the UPF uses the associated IP Packet Filter Set to match to the correct QoS profile. The Packet Filter Set may include the following parameters:

- Source/destination IP address or IPv6 prefix.

- Source / destination port number.

- Protocol ID of the protocol above IP/Next header type.

- Type of Service (TOS) (IPv4) / Traffic class (IPv6) and Mask.

- Flow Label (IPv6).

- Security parameter index.

- Packet Filter direction.

### 4.6.2 QoS Parameters

The 5G QoS model supports both GBR and non-GBR flows. A QoS flow is assigned a QoS Profile with some of the following QoS parameters:

- A 5QI identifier: A set of static 5QI characteristics are specified. The characteristics include values that indicate whether it is a GBR or non-GBR QoS flow, the priority level, delay budget, packet error rate and averaging window, and the maximum data burst volume. In addition to the standardized 5QI values, non-standardized 5QI that are tailored to specific services can be defined dynamically.

- Allocation and Retention Priority to indicate the pre-emption priority of the traffic in the QoS flow.

- Guaranteed Flow Bitrate for the uplink and downlink.

- Maximum Flow Bitrate for the uplink and downlink.

- Reflective QoS Attribute.

Maximum Packet Loss Rate for the uplink and downlink (limited to GBR QoS flows used for VoLTE).

# 5 Mapping of Existing Media Services

## 5.1 Media Distribution Systems

### 5.1.1 Introduction

The figure 7 provides a potential extension to the 5G reference architecture (as defined in TS 23.501) with media delivery related functions like CDN edge and origin functions.

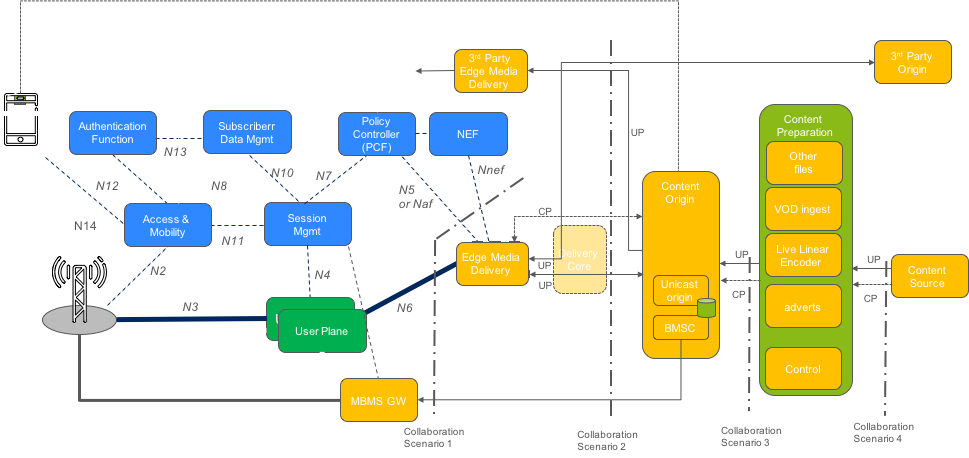


Figure 7: Media on 5G System Architecture

Most media distribution on 5G is based on Adaptive Bit Rate streaming with HTTP 1.1 to deliver file based video content. A very common video container format is fMP4 (also called ISO-BMFF) and MPEG2-TS. The new CMAF format is a profile of fMP4 and can be used with different manifest formats.

Typically, media segments are addressed with URLs where the domain name indicates the content provider name, i.e. the domain name of Content Origin.

The major components of media distribution are Content Preparation, Content Origin, and Delivery. The media delivery network elements and their functions are described in summery herein.

### 5.1.2 Media Distribution Functions

#### 5.1.2.1 Edge Media Delivery function

Edge Media Delivery is part of a CDN facing the clients and connects to 5G packet core UPF through the N6 reference point. The function is typically a HTTPS Reverse proxy/cache serving the UEs with content pulled from the CDN. It also caches content and functions as a HTTPS server when serving UEs from the pre-cached store. Content pre-caching is performed either on-demand when requested by clients or by CDN operator management instruction that pre-fills Edge Delivery with new content prior to the official launch of content. The management and the content life cycle management of pre-cached assets are handled via the Manager and Controller.

In case of cache miss, content is pulled with HTTP(s) from the CDN. When the content is not in the cache, the Edge Delivery connects up-stream to a Delivery Core that finally forwards the HTTP GET request to Content Origin. Request Routing part of the CDN Control routes the request with HTTP redirection to the location of the content.

Being HTTPS reverse Proxy / cache, Edge Delivery maintains the Certificate for the CDN domain name.

Edge Media Delivery function are capable of redirecting client requests to other Edge Media Delivery functions, based on various conditions. This procedure is called in the following Request Redirection.

Beside the basic HTTP revers proxy /cache functions, an edge delivery function may perform (depending on the collaboration scenario) manifest manipulation e.g. as specified by SCTE-130 Ad Decision Manager where the decision to insert ads into the manifest done by the SCT-130 Ad Decision Service. Manifest Manipulation allows for the creation of custom manifests per content and client in realtime.

Edge Delivery may provide differentiated media delivery on 3GPP access based on type of content (live or VOD), streaming connection state, and type of client device. Edge Delivery is state-full by keeping states on all connected clients and their state in streaming.

The edge deliver may consider using the path knowledge between Edge Delivery function and the UE in the transport protocol configuration.

#### 5.1.2.2 Delivery Core function

Delivery Core provides a mid-tier caching level in the CDN. Cache miss at Edge Delivery is routed to Delivery Core, so that the Origin is shielded. Cache miss in Delivery Core is routed to Content Origin (local or remote) one. Routing upstream to Delivery Core or Origin is supported by content aware Request Routing.

A remote Origin is configured with either origin FQDN or IP address, through which Delivery Core can fetch the content managed remotely by Content Provider.

In large CDNs, Delivery Core is required to shield Origin servers, both local and remote.

Caching at Core Delivery is either on demand or prepositioned. The management and the content life cycle management of pre-cached assets are handled via the Manager and Controller.

#### 5.1.2.3 Content Origin function

Origin is the entry point to place content onto the CDN and provides typically a permanent store for the managed assets of the CDN. It is the central access point within the CDN. It is responsible for ABR delivery of the content functioning as HTTP Server with a large file based storage.

VOD, Live-to-VOD and live / Linear assets are stored at Content Origin along with their metadata (manifests). Live content is placed on Origin for a fairly short-time. The management and the content life cycle management of assets in Origin are handled via the Manager and Controller.

Assets at Origin are prepared beforehand and stored in fMP4 files, i.e. transcoded, packaged in MBR streams for multi-screen video delivery. If ingested VOD content from CP is in a different format, Origin repackages the content into fMP4 format. To this end, Just in Time packaging (JITP) might be used, i.e. packaging is done when clients request the content.

Origin for Server Side Ad Insertion provides manifest manipulation e.g. as specified in SCTE-130 Ad Decision Manager. This is done for DVR services particularly for private copy DVR.

#### 5.1.2.4 Content Preparation

##### 5.1.2.4.1 Live Linear Ingest

The main function of Live Linear Ingest is to prepare live streams for ABR delivery. Those are the Encoder functions.

Ingest – ingesting live channels including Audio, video, subtitles and metadata. It supports ingestion of H.264 and H.265 video over IP as well as other inputs like SDI inputs.

Transcoding – Perform transcoding from single bit stream (MPEG2, H.264, H.265) to multiple H.264 or H.265 ABR streams. Currently, H.264 codec is widely used. However, H.265 is introduced more and more supporting UHD and offering about double the data compression ratio at the same level of video quality, or substantially improved video quality at the same bit rate.

Packager – Perform content packaging for ABR multi-screen delivery. Packager includes both segmentation and Encryption. Packager segments each bit rate stream into segments. It also generates the manifest for the live stream. Each segment is encrypted for DRM with Common Encryption (CENC). The encrypted content as well as the manifest is pushed to Content Origin.

Packager creates manifests and may also insert Event Signalling and Messaging (ESAM) or other markers. Packager may work in conjunction with a Placement Opportunity Information Service (POIS).

##### 5.1.5.4.2 Control

5.1.5.4.2.1 Request Routing

Request Routing function is responsible for end-user request routing within the CDN. It is essentially provided by the Request Router or A-DNS depending on whether it is HTTP based or DNS based request routing.

Request Router is stateless and supports rule-based HTTP redirection with multiple redirection with multiple features like (Load balancing, node health-check, Node bandwidth usage, GeoIP policy, Content awareness…).

In simple terms, the client requests content, the request hits first the Request Router that selects the suitable Edge Delivery for the client and route the request to it. In addition, Request Routing is used for steering the internal traffic within CDN and up to the Content origin providing resiliency and content awareness.

5.1.5.4.2.2 Authoritative DNS

A-DNS is an authoritative DNS server serving zones with content to Delivery Edge mappings. This works in tandem with the Request Router in order to fulfil external DNS resolution requests from the internet.

##### 5.1.5.4.3 VOD ingest

In charge of viewing and controlling the ingestion of VOD assets into the Content Origin through distribution of individual tasks that detail all the information associated with a particular asset throughout its processing.

VOD ingest provides service exposure to securely ingest the asset onto Content Preparation and then into the Content Origin. This operation supports the direction of the assets to Origin or Edge Delivery. Service Exposure handles content life cycle management.

Metadata Ingest manages the ingestion of VOD metadata into the CDN.

VOD ingest may provide quality check, transcoding and packaging of VOD content. VOD content is encrypted for DRM with Common Encryption (CENC). The encrypted content is pushed to the Content Origin.

##### 5.1.5.4.4 Other files

The content preparation function may also prepare other files like HTML based metadata and asset descriptions and asset images. These content files may require dedicated preparation and processing functions.

## 5.2 Identified Issues in Mapping to 5G System

### 5.2.1 Network Interface Considerations

In mapping media distribution services to the 5G system, the following issues have been identified, for which a set of solutions will be investigated as part of the present document:

- Resolving the appropriate / closest-by edge cache: the ability of supporting edge caches for media services is one of the new functionalities offered by 5G. There will likely be several edge caches in a 3GPP 5G network, and it might be desirable to redirect UEs to one of these edge caches, e.g. based on proximity to the UE or for load balancing or other reasons. Hence the main question is how to redirect the user to the most appropriate edge cache. DNS could be used for resolution (assuming a DNS server is present in the 5G network). However, how long should a DNS entry be valid in this case? It seems hard to decide on a fixed validity time for a DNS entry, since the validity of a certain DNS entry may depend on the network the UE is using (3GPP or not) and on how fast the UE is moving (and hence a new edge cache may become more appropriate). UE mobility is maintained by the AMF, could the AMF help here?

- 5G NF service exposure via the NEF: to enable media services, a number of 5G network function services need to be exposed by the NEF. For example, when new edge caches are deployed or switched off, the corresponding DNS entries to reach them may need to be updated in the DNS of the 5G network. The NEF will likely expose a functionality to update the DNS entries. Additional information that may be exposed is an overview of the available DANEs and other functions or information relevant for media applications.

- AF influence traffic routing in the 5G core: one of the new aspects introduced by the 5G architecture is that AFs have the ability to influence traffic routing in the 5G core, e.g. based on some application's requirement. How are 3rd parties (e.g. CDNs) envisioned to use this functionality? Is it useful for media services?

When SAND [12] is deployed for over-the-top VoD, several gaps can be identified for the case when a DANE is managed by a third party. For what concerns the interfaces of the DANE to other network elements, the following gaps are highlighted:

1) What information does the DANE need to obtain in order to advice/inform the media clients and how can the information be provided to it, and how should the information be exposed to the DANE? For example: information on the bandwidth / latency in the Access Network, Core Network and Data Network towards the content source (may be origin or cache).

2) How can be ensured that once the DANE has decided e.g. a Shared Resource Allocation, that the resources are also actually available/reserved for those clients?

### 5.2.2 5G QoS Considerations for Media Distribution

Content providers may negotiate certain QoS profiles for their traffic and even for specific users. These can include standardized 5QI values or they may define non-standardized 5QI values with tailored QoS characteristics and priority levels. These profiles may be stored by the UDM to be accessed later by the PCF.

With the emergence of new media services such as FLUS for live broadcast and VR/AR/MR, service requirements will vary greatly. For instance, in the former, QoS allocation cannot be reflective and more bandwidth and delay requirements are needed for the Uplink. For the latter, latency requirements are essential. For other new services such as media for autonomous driving, extremely high reliability in addition to low latency are critical. Some of these applications may initially be covered by standardized 5QI but as the services and applications evolve more customization will be required.

The NEF is used among others for provisioning and policy/charging requests by external entities. Content and service providers may use the NEF to pre-configure or dynamically control the QoS handling of its traffic. This functionality may be offered through CAPIF APIs. As of now, TS 23.502 defines an NEF service for influencing traffic, allowing the content provider to influence the routing of traffic. However, it is not clear if this enables influencing the QoS profile assignment as well.

## 5.3 Potential Mapping to 5G System

### 5.3.1 Introduction

In this clause, a set of potential mapping options of existing media distribution functions to the 5G system are described.

### 5.3.2 PSS Service Mappings

#### 5.3.2.1 Identified PSS Functions

3GPP offers Packet Switched Streaming Services (PSS) as a streaming solution for mobile devices. It supports RTP-based as well HTTP-based streaming, in addition to progressive download. These services usually consist of a set of functions that are composed together to provide the media streaming service.

The following functions can be identified:

- Streaming server: the server that acts as the origin server for the content and which serves the content. The server may be a DASH server or an RTSP server. Alternatively, it could be a regular HTTP server that hosts the media content for progressive download.

- Specified Support Functions:

- Capability exchange: a server that host device profiles and provides access to device information to the streaming server.

- QoE reporting server: a server that collects quality of experience measurements based on a set of QoE metrics that it agrees with the client through the session setup procedure.

- A Networked bookmark server: a server that stores bookmarks for specific content and for a specific user and makes it available across devices.

- DRM protection: a DRM license server is used to verify eligibility of a user to consume content and to provide decryption keys in form of a license for a limited period of time to allow decryption and consumption of the content.

- Scene description: an HTML5 document and its related resources are hosted on an HTTP server to provide access to the content as part of an enriched scene.

- Network support: information may be provided by servers in the network to assist the client in operating the streaming session.

- Implicit Support Functions:

- DNS address resolution: a DNS server provides DNS resolution for the content URL. The resolution may take into account geographic proximity, load, and other aspects for optimized support.

- Load balancing: a load balancer may be used to redirect the requests from the client to the closest and/or least loaded server to improve the quality of experience.

- Content distribution: a CDN may be used to scale the distribution of content to a large number of users.

- QoS management: the network may gather information about the pending or ongoing streaming session and allocate resources to the session to meet some target QoS targets.

Over the top streaming sessions may provide some of these functions through the network and provide others externally, i.e. outside the mobile operator's network.

#### 5.3.2.2 Mapping to the 5G system

Following the design principles of the 5G system, the aforementioned functions should be designed according to the service-based design model and to enable the functions to be virtualized.

The following table provides an initial mapping of the identified functions and the 5G system functions:

Table 6: Mapping of Media Streaming Functions to 5G system functions

| Media Streaming Function | 5G System Function | 5G System Interface to | Description |
| --- | --- | --- | --- |
| Streaming Server -  Control Plane | AF | NEF | The streaming server may reside within the MNO's network as a dedicated application function or it may reside externally and interact with the network through the NEF. |
| Streaming Server – User Plane | AF | UPF | Media data flows through the DN to the UPF directly or through one or more AFs. In the latter case, the AF may act as if it were the origin server. |
| Capability exchange | AF/UDR |  | The user data repository function may be used to store device profiles and user preferences. Alternatively, this may be performed by an AF. |
| QoE reporting | AF |  | A reporting server may be implemented as an application function within the operator's network. |
| Networked bookmark | UDR/AF |  | Bookmarks may be stored as part of a user profile that is accessed through a dedicated networked bookmark application function. |
| DRM protection | AF/AUSF |  | The license server will usually be implemented as a dedicated application function. The authentication server function may fulfil this functionality or support the DRM AF. |
| Scene description | AF |  | Scene description may be offered through an HTTP server that is implemented as an AF. |
| Network support | AF | NSSF | Congestion marking and SAND functionality may be offered through an AF in concert with the network slice selection function. The NSSF will assign a dedicated slice that understands the nature of the service and offers adequate network support. |
| DNS |  | NRF | The repository function may be configured to perform appropriate DNS resolution to locate media resources and serve the content through the closest edge server. |
| Load balancing |  | SMF/NRF | Session management is used to assign a session to the most appropriate end point. NRF may also be used to perform load balancing. |
| Content distribution | AF |  | CDN nodes are implemented as application functions that operation at the application layer and support the requested protocols (e.g. HTTP). |
| QoS management |  | PCF | The PCF in conjunction with other functions will ensure appropriate QoS allocation to the network slice that is assigned for the session. |

It is to be noted that the reference points between the stated network functions will need appropriate adjustments to fulfil the requirements of media streaming services. These adjustments are subject to further study.

### 5.3.3 Evolution of PSS and MBMS towards 5G

The PSS Architecture (Stage 2) is defined in TS 26.233 and contains unicast delivery architecture for RTSP, Progressive Download and 3GP DASH.

The MBMS User Service Architecture (Stage 2) is defined in TS 26.346.

The figure below starts decomposing the MBMS User Service Architecture and the PSS Architecture into smaller functions.

The UE functions are ffs and currently only drawn for completeness.

The MBMS related transport functions are drawn in red, orange and yellow.

The xMB reference point considers content preparation on the content provider side, so that 3GP DASH formatted segments and MPDs are provided by the content provider. In PSS, the PSS server receives 3GP-DASH format from a content preparation function.

MBMS and PSS support QoE reporting. In case of MBMS, the QoE reporting is part of the associated delivery function (ADF).

MBMS supports Byte Range File Repair as part of the associated delivery function (ADF-FR HTTP-BR). Such a function can be hosted together with 3GP DASH unicast segments and MPDs on an HTTP server.

The MBMS provisioning and more control plane functions are depicted in organ. Unicast and MBMS Session and Transmission Function (SnT F) for file delivery can be used to offer Service Announcement to clients.



Figure 8: Decomposing MBMS and PSS functions

# 6 Mapping of New Media Services and Verticals

## 6.1 Introduction

In this clause, a selected set of new media services that are either expected to be enabled or significantly enhanced by the 5G system are investigated. The analysis covers the integration with 5G as well potential extensions to leverage the new 5G features.

## 6.2 Media Production

### 6.2.1 Introduction

The 5G System offers content providers a robust, high bitrate, low latency uplink that can be leveraged to offer a contribution link for user generated and professional content providers. As part of Release 15, a Framework for Live Uplink Streaming (FLUS) has been specified to provide a flexible framework for content providers to stream live content to a FLUS Sink. The FLUS Sink can be discovered by the content provider based on that Sink's capabilities.

A professional content provider may want to make use of the FLUS framework to stream content with a high fidelity to the network and instruct the receiving FLUS Sink to perform initial processing and then stream that content to the content providers servers or store the content at the Sink for deferred access.

The content provider acts as a FLUS Source and discovers a FLUS Sink that has certain capabilities such as:

- High bandwidth for very high quality video streams

- Capabilities to recover from losses in the Uplink through support for FEC and/or QoS

- Storage capability or capability to stream content live to content provider's network

- Pre-processing capabilities, e.g. to encrypt content at the Sink

The FLUS Sink can be operated as an Application Function that has the capabilities to communicate with the PCF to define the QoS policy for the uplink stream. This AF may be configurable through the NEF and accessible directly through the Data Network.

The Generic FLUS framework may be appropriate for this scenario, where the content provider uses the RESTful API to create and manage the FLUS session.

## 6.2.2 FLUS Framework Usage in 5G

### 6.2.2.1 Identified FLUS Functions

The Framework for Live Uplink Streaming (FLUS) is an enabler for live media streaming from a source entity to a sink entity, which supports an IMS-based and a non-IMS-based instantiation. The system usually consists of a set of functions that are composed together to provide the media streaming service.

The following functions can be identified:

- FLUS Source: The FLUS source receives media content from one or more capture devices. The capture devices are considered as parts of a UE or are connected to it.

- FLUS Sink: When the FLUS sink is located in a UE, the FLUS sink forwards media content to a decoding and rendering function. When the FLUS sink is located in the network, the FLUS sink may forward media content to a processing or distribution sub-function.

- Specified Support Functions:

- FLUS control plane: FLUS control plane functionality including the associated processing by FLUS sink of the uploaded media for subsequent downstream distribution, and FLUS media instantiation selection.

- Implicit Support Functions:

- Rendering: a function that transforms media signals into appropriate formats that can be displayed or played back, and therefore be watched or listened by a user, e.g., by taking the direction of user's head into account.

- Processing: a function that changes the format of media signals for further transmission or generation of a media signal from multiple input media streams, e.g., for network based stitching or mixing.

- Distribution: a function that forwards the media into other network or devices.

### 6.2.2.2 Mapping to the 5G System

These functions of FLUS functions can be mapped to the 5G system functions as follows:

Table 7: Mapping of FLUS Media Streaming Functions to 5G system functions

| Media Streaming Function | 5G System Function | 5G System Interface to | Description |
| --- | --- | --- | --- |
| FLUS Source | UE | UPF | FLUS Source may consist of a conventional UE supporting modem capability and a media front-end that captures audio and video signals. The media front-end, such as a 360 camera, may be physically separated to UE and connected via interfaces that are out of the scope of 3GPP. |
| FLUS Sink | UE | UPF | FLUS Sink resides in a conventional UE that supports connectivity and a media front-end for rendering and playback. The media front-end, may be a head mounted display, that is physically separate but connected to the UE via interfaces that are out of the scope of 3GPP. |
| AF | NEF | FLUS Sink in network may forward media content to a Processing and Distribution functions. |
| FLUS Control | UE | IMS | FLUS control is used to establish and control the FLUS session between FLUS Source and FLUS Sink.  When the FLUS Sink is a UE, the control happens through the IMS Core. |
| AF | PCF | A FLUS Sink in the network may connect to the PCF to assign appropriate QoS allocation to the FLUS session. |
| Rendering | UE |  | Media signals may be rendered at a UE or a media front-end connected to the UE via interfaces that are out of the scope of 3GPP. |
| Processing | AF | NEF (optional) | An application function (AF) that offers to host different types of media processing, e.g. VR stitching, transcoding, or re-formatting the video or audio.  The processing may optionally be exposed through the NEF, if the service provider is a 3rd party. |
| Distribution | AF | NEF (optional) | An application function (AF) facilitating the distribution of the FLUS media, e.g. through the PSS or MBMS or to external sources.  The distribution may optionally be exposed through the NEF, if the service provider is a 3rd party. |

## 6.3 AR and XR Services – Use Cases and Context

### 6.3.1 Service Description

Augmented and Extended Reality is expected to grow significantly over the next few years, based on different market studies. Many use cases and applications are expected to be wireless on mobile and portable devices (including new form factors such as AR glasses) requiring many different enablers that play together to create immersive services and experiences. This includes, but is not limited to:

- highest visual quality in order provide realistic presentation in mixtures of real world and extended realities;

- highest sound quality in order provide in mixtures and interaction of real world and virtual sounds;

- intuitive and seamless interaction with real and virtual objects for a full immersive experience.

Whereas AR and XR are expected to share certain requirements with Virtual Reality as discussed in TR 26.918 [14], especially eXtended reality (XR) is considered an umbrella of VR and AR use cases, enabling seamless integration of different worlds as shown in Figure 9.

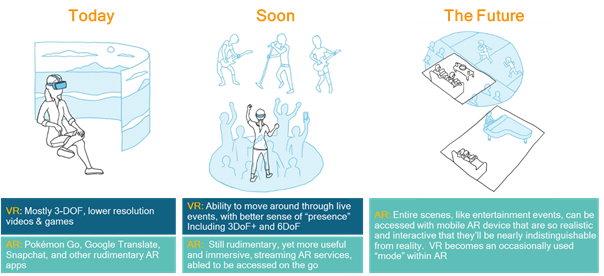


Figure 9: From VR and AR to Extended reality (XR)

Use cases for extended reality include entertainment, infotainment, education, healthcare, industrial use cases, etc. Also, in the SMARTER TR 22.891 [15], six use cases discuss Augmented Reality at least as being a part of the considerations. Also it is expected that extended realities will be more and more supported by new form factors such as AR glasses with a multitude of sensors, trackers, cameras, and microphones, as shown for example in Figure 10.



Figure 10: Glance of a first responder glass

A substantial difference of XR to VR as studied in TR 26.918 [14] is the addition of significant uplink traffic, potentially associated with very low latency requirements.

### 6.3.2 Potential Issues on AR/XR/VR (new)

While the growth perspectives are huge, studying the relevancy of interoperability enablers for such new services requires detailed understanding of use cases and requirements to address these issues. It also requires understanding how emerging 5G core network and radio technologies can successfully contribute to AR and XR services, for example to support latency and/or bitrate requirements. Before initiating detailed normative work a detailed analysis of use cases, requirements and especially the interop points for are relevant. In particular, the following aspects should be considered:

- Analysing the different technologies and equipment in place that provide an Extended Reality experiences.

- Collecting the associated use cases and identifying the 3GPP service(s) they map to

- Analysing and identifying the media formats (including audio and video), interfaces and delivery procedures between client and server required to offer such an experience

- Identifying relevant client and network architectures and APIs that support AR and XR use cases

- Collecting key performance indicators for relevant AR and XR services and the applied technology components.

- Conducting subjective tests to estimate the audio and video formats and encoding parameters required for ensuring the quality of experience as considered necessary

- Studying the processing requirements (both audio and video) and associated issues such as spatial resolutions, frame rate, latency and accuracy of field of "view" rotation.

- Collecting information on market and standardization status and communication with relevant 3GPP groups and external organizations

- Drawing conclusions on the potential needs for standardization in 3GPP.

It is considered important that 3GPP initiates a study on this matter.

### 6.3.3 Key Issue and Recommendations

Based on discussion in clause 6.3, it is proposed that 5G Media addresses different device APIs:

- APIs that enable access to functions that are defined by 3GPP and are device internal.

- APIs that enable to abstract complex 3GPP network functions with abstracted APIs that can be accessed by third-party applications.

It is also recommended that 3GPP devices in 5G media provide a consistent support for capability discovery. Suitable options may be:

- Media Profile capability API as defined in clause 7.3.2.1.

- ISO BMFF API as defined in clause 7.3.2.2.

Media Capability API work in 3GPP should be followed. Also, any work on MIME parameters extensions may be considered if progressed.

Selection and profiling of the APIs that are needed to support immersive media on the UE will be needed. In particular, the group needs to study requirements on the UE to support the following functionalities:

- Graphics processing APIs for processing and rendering media

- Any potential new or updated Codec requirements to support immersive media

- New media types such as Meshes and Point Clouds

- Study different Scene Graph solutions as a potential addition to scene description in 3GPP

## 6.4 Conditional and Dynamic Policies for Application Instances

### 6.4.1 Service Description and new issues

Mobile Operators are offering more and more special services within a data plan subscription. For example, that some 3rd party services will be charged differently, when a certain policy is followed. An example might be a video pre-loading service, which result in lower data volume cost, when the download bitrate remains below a certain threshold. Another example might be a video on demand preload service, which follows different reception rules than other traffic to the same user.

An important aspect here is the dynamicity of selecting different policies. The policy should not be applied to all traffic between the application and the content provider back-end. Only some selected traffic is being policed and different policies may be possible. For example, browsing the content catalogue of the content provider should not be policed, while the policy should be active when pre-load the content in background.

Note, the application is still elastic and capable of adapting to different link bitrates.

Today's solutions often require the provision of IP addresses to allow identification of eligible traffic for policing. The disadvantage of such static detection schemes is that all content on that server is eligible to the same policy rule.

The 3rd party content provider may have different agreements with the different national mobile network operators. Integration effort on the content provider side should be minimized. Deployment of operator specific servers should be avoided.

It can be generally said, that content providers start collaborating with mobile network operators to offer a new type of service on top of the IP connectivity service. Procedures to on-board new content providers should be very efficient and highly automated. Preferably, a "self-on-boarding" procedure should be possible so that a content provider can leverage the new "different charging policy" services. The 3rd party content provider may prefer to use the same application server instances for different policies and for different MNOs (i.e. dynamic, flexible and fast policy activation is needed).

Many applications are executed potentially simultaneously on the UE. These applications may also interact with some servers, thus, use the 3GPP radio interface (Uu) for communication. Even the application of one content provider may execute multiple application service instances simultaneously. For example, the user may watch a video stream at the same time as a VOD pre-loading service is executing. These application service instances may be very dynamic.

In today's CDN architectures, CDN Edge Media Delivery functions are often realized as very simple and stateless HTTP servers or HTTP reverse proxies. There is often the strategy to separate static content from simple CDN Edges and more dynamic content (like personalized recommendations or comments) from different CDN nodes.

In order to activate a certain policy for a video preloading or a video streaming service, the 5G Core needs to identify the traffic flow or flows for the policy. To continue supporting simple and stateless HTTP servers, a basic approach would be to trigger the policy for a certain service instance from the client application side.

It is here assumed that the media service is operated within a single network slice.

### 6.4.2 Potential Architecture enhancement

The architecture from clause 5.1.1 has been extended (and simplified) by an additional user plane function for the Application to Network Interactions for policy activation (ANIF). The function is responsible to activate different policies for traffic flows.

This function provides a user plane end-point for interacting with the application. The function is located on N6 reference point. The function could be controlled by the mobile operator (trusted function outside of the Core Network) and could therefore be authorized to communicate directly with PCF or other functions. Alternatively, the function could be controlled by the 3rd party content provider, thus, needs to interact with the NEF. The application from the UE side may interact with ANIF to activate or deactive a policy for a certain traffic flow. The ANIF includes an Application Function (AF), which allows the ANIF to interact with e.g. the Policy Control Function (PCF), directly or via NEF.

A content provider may have loaded one or more policies for this application. The Application may then dynamically active / deactivate policies for a given flow.

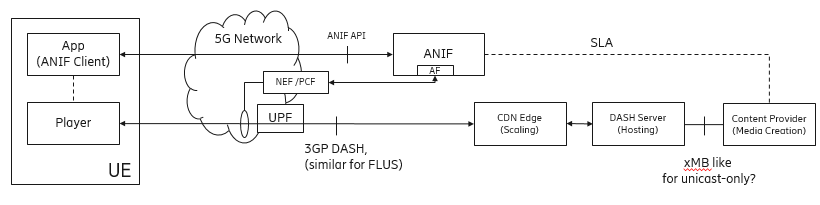


Figure 11: Extended Architecture

Figure 11 depicts media plane functions for 3GP DASH (cf. TS 26.247), which includes a CDN Edge (for scaling), the DASH Server (Content hosting) and the Content Provider (which includes media creation). The content provider can have a Service Level Agreement (SLA) with the MNO.

### 6.4.3 Example procedure for VoD Preloading case

Considered Architecture (additional functions compared to architecture):

- PLC: Preload Client provides means to decide on preloading of a certain VoD asset. Note, the MNO is not aware about the actual VoD item, i.e. privacy is ensured.

- Edge Media Delivery: EMD hosts the different VOD assets for pre-loading. The EMD may be a simple HTTP Server or an HTTP Reverse Proxy (maybe with CDN like token access).

- ANIF: Application Network Interaction Function authorizes application policy requests. An application may have the choice for different policies and the application decides based on a need bases, which policy to activate for the upcoming transaction.

- Cont Op: The 3rd Party Operator, which offers the content for pre-loading. Different policies may allow the operator to select a different data tariff.



Figure 12: Example VoD Preloading Procedure

Precondition:

1: The Content Operator agrees with a network operator to use different policies for content delivery. The UE application of the Content Provider (called UE application afterwards) could have a list of possible policies for different use-cases. Application level signalling (out of scope here) may provision the UE applications with the policy possibilities.

When a user selects to pre-load a VoD asset

2: The UE application of the content provider understands that a VoD asset should be fetched e.g. within a certain time frame. The UE application sends a Policy Activation Request for a certain policy to ANIF.

3: The ANIF function notifies the UPF/PCF about the selected policy

4: The ANIF function sends a response, indicating that the policy request was granted and expiration information of the policy.

5: The UE application sends a request for the VoD asset.

6: The UE application receives a positive response, indicating the start of the download procedure. The download procedure may take some time.

7: The UE application client has received the full VoD asset. Note, the asset might be fully encrypted due to copy right reasons.

### 6.4.4 Example procedure for Audio PodCast case (changing priority)

Another example is a PodCast example. In case of audio PodCast, a client fetches a stream syndication file (once or periodic), which describes a sequence of additional audio media assets. The client downloads each audio file (using 3gp file format) and starts playing it, once a first file is loaded.

This example assumes a podcast client, which is just downloading the needed audio file some short time in advance. The client is fetching here the next audio file, while a first audio file is playing. Thus, the download of the next audio file needs to finish, while the first file is playing (i.e. download deadline).

The podcast client leverages a different download policy, which decreases the traffic priority (called here "Lower than best effort") in order to save transmission resources.

Considered Architecture (additional functions compared to architecture):

- PC: Podcast Client decides, which audio file should be downloaded next and at what time the download should be completed at latest.

- Edge Media Delivery: EMD hosts the different audio files for downloading. The EMD may be a simple HTTP Server or an HTTP Reverse Proxy (maybe with CDN like token access).

- ANIF: Application Network Interaction Function authorizes application policy requests. A client may have the choice for different policies and the client decides based on a need bases, which policy to activate for the upcoming transaction.

- Cont Op: The 3rd Party Operator, which offers the audio podcast feed. Different policies may allow the operator to select a different data tariff.

Precondition:

1: The Content Operator agrees with a network operator to use different policies for content delivery. The UE application of the Content Provider (called UE application afterwards) could have a list of possible policies for different use-cases. Application level signalling (out of scope here) may provision the UE applications with the policy possibilities.

When a user selects to pre-load an audio file (which should be fetched until a certain deadline)

2: The Podcast client understands that the audio file should be fetched within a certain time frame. The client sends a Policy Activation Request for a certain policy to ANIF.

3: The ANIF function notifies the UPF/PCF about the selected policy

4: The ANIF function sends a response, indicating that the policy request was granted and expiration information of the policy.

5: The UE application sends a request for the audio file.

6: The UE application receives a positive response, indicating the start of the download procedure. The download procedure will take some time.

7: The podcast client detects the risk, that the audio file cannot be downloaded in time. In steps 8 to 10, the podcast client increases the priority by interacting with ANIF.

11: The podcast client has received the full audio file.

### 6.4.5 Example procedure for a FLUS session with QoS

Another example can be a live uplink video case, leveraging the FLUS framework. With ANIF, the FLUS source can interact with an ANIF function to get temporarily either different traffic treatment. The call flow would look very similar to the previous examples.

### 6.4.6 Relation to other SA4 Network Assistance Functions

The ANIF concept adds media related transactions with a UE based client and a new network function. TS 26.247 already includes similar type of transactions in the "Server and Network Assistance DASH" (SAND) section (Clause 13). Here, the DASH Player can interact with a DASH aware Network element (DANE) for assistance. TS 26.247 supports for Network Assistance (e.g. for rate recommendations), Proxy Caching, Consistent QoE/QoS and Multi-Network Access support. Functions like network assistance can be of benefit for other media services like progressive download or FLUS as well.

### 6.4.7 Interface to 3rd Party Service Provider

To address the gaps identified in clause 5.2 on OTT SAND deployments, it may be desirable to enable mechanisms that expose real-time network information from the operator network to an 3rd Party DANE. As one potential solution to realize this goal, ETSI APIs GS MEC 009 - GS MEC 011 can be considered. An 3rd Party DANE may then use the real-time network information towards deriving the relevant SAND PER messages, as defined in 3GPP TS 26.247 [12].

### 6.4.8 Summary of Proposals

The Application to Network Interaction concept (ANIF) introduces the possibility that a media client can interact with the system in order to change the delivery policy. It has been identified, that the new policy related function can leverage similar interaction concepts such as Network Assistance for DASH.

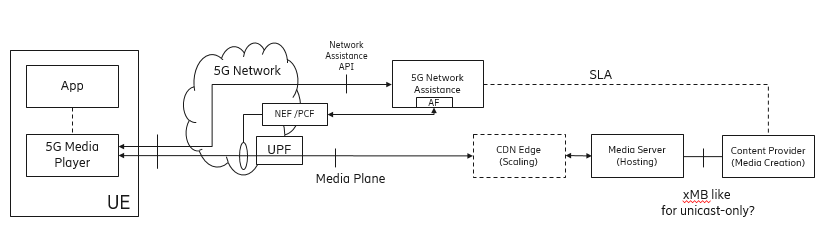


Figure 13: Suggested Architecture

It is suggested to identify common network assistance functions, which are relevant for several media services such as 3GP DASH, 3GP Progressive Download, FLUS, etc.

Common functions might be rate recommendations (SAND Network Assistance, cf. TS 26.247, clause 13.6) or also dynamic policy configuration (introduced here). The new 5G Network Assistance Functions should contain identified common function. A component in the UE (independently whether FLUS or 3GP DASH or 3GP progressive download) can access Network Assistance Functions via this API.

Summary of proposals:

1: It is proposed to define a new Network Assistance API, which includes Dynamic Policy Control functions.

2: It is proposed to review the 3GP DASH specification for common SAND Network Assistance functions.

## 6.5 Cloud Media Processing in 5G

### 6.5.1 Introduction

With the introduction of the 5G system, a new range of services becomes possible. Offloading media processing to the mobile edge cloud enables realizing advanced XR and immersive media use cases on mobile handheld devices in a way that would otherwise not be possible with traditional client/server architectures or even with classical cloud platforms. Note that the media and XR aspects of this service/use case will be handled in the XR5G study item.

One such service is Split Rendering, where rendering of large 3D scenes for gaming or entertainment, which requires a large number of GPUs to render in real-time, would be offloaded to the cloud. A render farm [https://en.wikipedia.org/wiki/Render\_farm] would usually have an architecture as depicted by the following diagram:

In this architecture, content is pre-rendered on the cloud by the master renderer or by a set of renderers that are controlled by the master renderer. The baked form of the content is then forwarded to the UE for final rendering.

Other applications of cloud media processing include processing of images and video for distribution. Possible operations include scaling and transcoding to match the UE's capabilities. More advanced processing, e.g. including object recognition and tracking, are also possible and are usually performed using deep learning algorithms.

Currently these services are already being performed in the cloud, e.g. for professional content creation and for online gaming. A 5G System that offers an edge cloud computing platform, will enable deploying these services in the 5G edge cloud(let), thus bringing the rendering as close as possible to the UE and guaranteeing QoS treatment of the application traffic.

### 6.5.2 5G System Considerations

The 5G core currently offers the ability to route traffic to a specific DN (e.g. an LADN) that is close to the UE. Provisions are taken to enable mobility of this traffic while maintaining session continuity. The DNN will usually connect to an edge compute environment, where Application Servers (AS) are running.

Third party content providers will be able to host their AS in proximity of users. The AS is also able to launch an application on the UE and assist the UE in selecting the appropriate DNN (e.g. the LADN) to reach the AS. This procedure is described in TS 23.501 [2] as follows:

*"PDU Sessions are established (upon UE request), modified (upon UE and 5GC request) and released (upon UE and 5GC request) using NAS SM signalling exchanged over N1 between the UE and the SMF. Upon request from an Application Server, the 5GC is able to trigger a specific application in the UE. When receiving that trigger message, the UE shall pass it to the identified application in the UE. The identified application in the UE may establish a PDU Session to a specific DNN, see clause 4.4.5."*

3GPP currently does not standardize a cloud platform for the edge compute environment. ETSI MEC [9] or any other cloud platform may be used for this purpose. 3GPP may also decide to define a standardized cloud platform or to integrate another standardized one. SA5 has started a study on the management of edge computing in [16]. If one edge compute platform is standardized or officially specified, the APIs offered by that cloud platform will be used by 3rd party service providers to launch compute engine instances in the edge. These compute engines will then run the service provider's AS.

Another alternative is to have agreements between service providers and operators to pre-configure and host AS from the service providers in the operator's compute environments. This on the other hand will negate the advantage of elastic computing that is the promise of cloud computing.

There is need for standardizing media handling and processing functions that can be deployed as AS. For instance, the split rendering functions can be deployed as a set of AS in the edge. The split rendering AS will direct the application in the UE to launch and connect to the AS e.g. through the appropriate LADN. Split rendering can then be performed between the AS and the UE renderers.

The following flow chart diagram shows a simplified procedure to initiate and perform media processing such as split rendering in the edge.

1) Session Establishment: the app on the UE contacts the service provider to establish a media session, e.g. an immersive 6DoF media streaming session.

2) The Service Provider instantiates a new AS in proximity of the UE if no instance is already running. It then exchanges information about the media session with the AS.

3) Application Trigger: the AS uses the Application Trigger mechanism to request the UE to establish a new PDU session to the DN (e.g. the LADN) on which the AS is residing.

4) PDU Session Establishment: the UE sends a NAS message with a PDU Session Establishment Request to the AMF to create a new PDU session to the LADN or the DN where the AS is located.

5) Media flow and processing can start from the service provider to the UE through the AS.

# 7 API Aspects

## 7.1 Introduction

This clause addresses questions on the evolution of immersive media and its deployment in a 5G system. It analyses the drawbacks of existing media APIs and the interaction with the network and identifies gaps and potential enhancements.

## 7.2 Use Cases

### 7.2.1 Introduction

This clause documents a few use cases relevant for media distribution in the context of 5G that are relevant for Device API aspects.

### 7.2.2 OTT Service Provider

An OTT service provider wants to distribute content to mobile devices. For this purpose, it develops an application, but for optimized performance the application wants to use broadly available capabilities on mobile phones. In particular, the OTT service provider wants to deploy only a single content encoding library. The OTT service provider reads the 3GPP user service specifications and is interested in reusing:

- the codecs that seem to be broadly available on devices and well tested, but is unclear how to use and access the codecs defined in 3GPP specifications.

- A DASH client as it offers the service as DASH content library.

The content provider is unclear how he can develop applications for Android and iOS reusing the above 3GPP device capabilities.

### 7.2.3 Browser-based consumption

In an extension to the use case above, the service provider does not want to rely on user installing app, and wants to make sure that the content plays back also on browsers. For this, it develops a progressive web application, but wants to use also use the same content library. The content provider is unclear how he can develop progressive web applications for browser to reuse 3GPP device capabilities.

### 7.2.4 Ad Insertion

In an extension to the use cases above, the content provider wants to do targeted and personalized ad insertion. The service provider has a library of content, but the ads are provided through a different library, typically encoded with different codecs and formats, despite these codecs and formats are also included in the 3GPP device capabilities.

The content provider is unclear how he can ensure that ad insertion with different codecs works to it satisfaction.

### 7.2.5 Live Services

The service provider uses now the 3GPP codecs, but for certain live services the latency of the content distribution is too high. The service provider wants to use network functionalities to improve the latency of Live streaming services without changing the streaming protocol. The service should be available on different platforms including Android, iOS and web browsers.

### 7.2.6 VR Services

The service provider now introduces VR Service distribution. In order to enable access and playback of the content, the app developer needs the following functions, either from the service platform with proper APIs, or integrated into the app:

- A DASH client, unless the DASH client is part of the application, (type 3), possibly with configuration APIs to supported optimized playback and rendering.

- OMAF metadata functionality to parse and extract the relevant information or at the minimum to instruct playback in the media pipeline.

- File format parsing functionality for video playback.

- Decryption module, if the content is encrypted.

- HEVC video decoder to decode the video content.

- Secure rendering and GPU functionalities to generate viewports.

- Sensors for viewport tracking.

- Generally, the content provider prefers to use HW supported functionalities to optimize speed, latency, power consumptions and overall performance. Each of those above functions may be accessed with APIs.

## 7.3 Interoperability Considerations

### 7.3.1 Background

Whereas 3GPP services typically address interoperability on network interfaces, in nowadays media architectures the protocols and formats are typically decoupled. Figure 14 provides a basic idea on the OTT application. Content conforms such that it can be played back on the 3GPP device. The app uses instructions to download and access the content and controls the presentation.



Figure 14: OTT application using 3GPP capabilities

An interop specification may primarily deal with the capability requirements of 3GPP Conforming Device that is used by an OTT application using well defined APIs in order to playback 3GPP Conforming content. The APIs may be abstract in order to support different application and device interface models. The media APIs also differentiate between devices supporting different playback variants, primarily "type 1" and "type 3":

- Type 1 Playback: The 3GPP device receives a manifest and downloads and plays back the contained media based on the information in the manifest. An OTT application may control the playback with limited control features.

- Type 3 Playback: The OTT application receives a manifest, downloads the media and uses media APIs in order to playback individual tracks of the media experience. The application is controls of the download and playback of the media using the well-defined track buffers.

The two models are shown in Figure 15 and Figure 16, respectively.



Figure 15: Type 3 type device



Figure 16: Abstracted device model for type 1

Initially focusing on type 3 playback, the application has access to the streaming manifest and parses and processes the manifest. Based on the information in the manifest, the application makes use of two primary high-level APIs:

- The control API: This API is primarily responsible for capability discovery of the device platform, establishing and tearing down track buffers for specific codecs, as well as to control the playback of media through track buffer controls.

- The media API: this API consists of one or multiple track buffers where the track buffers can be dynamically established and removed. The track buffers enable playback of content by the device platform.

It is expected that the application deals with manifest updates as well as with providing/downloading the segments that are then provided using the media APIs to the device platform for playback.

For 3GPP, the definition of such APIs as well as the necessary capability exchange and so on would be beneficial, in an abstract manner. Important instantiation would be:

- HTML-5 Media APIs: [https://www.w3.org/TR/html51/semantics-embedded-content.html#the-media-elements](https://www.w3.org/TR/html51/semantics-embedded-content.html)

- Android Framework APIs: <http://developer.android.com/reference/android/media/MediaCodec.html>

In addition, such an application may be embedded into a framework of network and device APIs as shown in Figure 17. The application may use proprietary protocols and functions to connect and communicate with the app server. The application may also use device internal APIs that abstract functionalities that are provided by network functions on the device. Finally the application may define well defined 3GPP codecs and local functions. Whereas many of the orange parts already exist and are documented in 3GPP specs, the green parts require new work and the blue parts are assumed out of scope.



Figure 17: Network and Device APIs

### 7.3.2 Potential Solutions for Capability Discovery

#### 7.3.2.1 Media Profile

The application uses the media profile for capability discovery. The media profile may be provided in the manifest or the 3G-FF/CMAF/ISO BMFF Header in the ftyp box. The application queries the device of the media profile that can be played back using the MIME Type parameter as defined in IETF RFC 4337 [7] and the profile sub-mime parameter as defined in RFC 6381 [8].

<mediatype>/mp4 profile="<media profile 4CC>"

The device may provide one of the following answers:

- Yes: If yes is provided, then the playback requirements for this media profile as documented in the present document are expected to be fulfilled.

- No: If no is provided, then the playback of the media profile is not supported by the device and the application will not playback this media profile.

- unknown: In this case the application should find other options to identify if the media profile can be played back.

Note that the media profile does not support the configuration signalling and requires an additional capability mechanism on which configuration is preferably used.

#### 7.3.2.2 3G-FF/ISO BMFF/DASH/CMAF Header/Initialization Segment

In this case an API between the app and the platform exists, such that the application queries the device if the content described in the 3G-FF/ISO BMFF/DASH/CMAF Header/Initialization Segment header can be played back. This has the advantages of being complete, accurate, future-proof, but the drawback of not being human readable and possibly requires transmitting more information than the other approaches.

Again, the device may provide one of the following answers:

- Yes: If yes is provided, then the playback requirements for this media profile as documented in the present document are expected to be fulfilled.

- No: If no is provided, then the playback of the media profile is not supported by the device and the application will not playback this media profile.

- unknown: In this case the application should find other options to identify if the media profile can be played back

If a no or an unknown is provided, the response should provide an indication based on what feature the device rejected the playback.

#### 7.3.2.3 MIME Sub-parameters

This option consists in using one or more MIME sub-parameters to describe the different required capabilities (pre-decoding, decoding, and post-decoding). It is the mostly used options today because it has the advantages of enabling a progressive, detailed, compact and almost human readable signalling.

Post-decoding requirements are indicated in the ISO base media file format with restricted schemes. For example, the 'resv' sample entry type can be used for video tracks that require certain post-decoding operations. Similarly, pre-decoding requirements are indicated in the ISO base media file format with the protected scheme.

In this case, the application uses the detailed MIME type string as defined in RFC 6381 [8] for the communication with the device platform. The application queries the device of the media profile can be played back using:

- <mediatype>/mp4 mime-subparameters

The device may provide one of the following answers:

- Yes: If yes is provided, then the playback requirements for this media profile as documented in the present document are expected to be fulfilled.

- No: If no is provided, then the playback of the media profile is not supported by the device and the application will not playback this media profile.

- unknown: In this case the application should find other options to identify if the media profile can be played back

If a no or an unknown is provided, the response should provide an indication based on what feature the device rejected the playback.

#### 7.3.2.4 Media Capabilities

Aligned with the Media Capabilities API (https://wicg.github.io/media-capabilities/), an APIs exposing information about the decoding and encoding capabilities of a device platform for a given format, but also output capabilities of the current device to find the best match based on the device's display may be used. The application would query a vector of capabilities and if all required capabilities are supported, then the playback may be initiated.

#### 7.3.2.5 Device Capability – Persistent Item Solution

Another approach is to standardized device capabilities in a known format for storage by the manufacturer and recall in some manner available to the application. The preferred approach uses standardizes key-value pairs for relevant player characteristics that can be communicated to servers via JavaScript APIs or Objects. One proposal is to use the HTTP User Agent String as described in HbbTV [4], clause 7.3.2.4. However, this approach is considered problematic in terms of scalability and was therefore dispensed during the development of the present document.

#### 7.3.2.6 Playback Capabilities

3GPP may define a dedicated capability code for the platform that matches against the full requirements in the present document. While such an approach may provide the most stringent interoperability, at the same time adding yet another option to the already complex world of capability signalling was dispensed during the development of the present document.

#### 7.3.2.7 MBMS Feature Tags

MBMS feature tags as defined in clause 11.9 of TS 26.346 [5] may be used. In this case the MBMS client would, based on the information on the USD, query the media platform on the support for features included in the feature tags.

#### 7.3.2.8 PSS Capability Exchange

PSS has defined a capability exchange framework (UAProf) in clause 5.2.3.3 of TS 26.234 [6] that allows the PSS server to identify the client device and then download a device capability profile that is associated with that device. The profile document may be extended with additional capabilities either permanently or during the session setup.

## 7.4 Key Issue and Recommendations

Based on discussion in clauses 7.2 and 7.3, it is proposed that 5G Media addresses different device APIs:

- APIs that enable access to functions that are defined by 3GPP and are device internal.

- APIs that enable to abstract complex 3GPP network functions with abstracted APIs that can be accessed by third-party applications.

It is also recommended that 3GPP devices in 5G media provide a consistent support for capability discovery. Suitable options may be:

- Media Profile capability API as defined in clause 7.3.2.1.

- ISO BMFF API as defined in clause 7.3.2.2.

Media Capability API work in 3GPP should be followed. Also, any work on MIME parameters extensions may be considered if progressed.

# 8 Summary and Conclusions

## 8.1 General Summary

### 8.1.1 General

As part of this study item, the impact of the new 5G system and core network on existing and emerging media services has been studied. A mapping of the media distribution functionality to the new 5G network functions has been done. It became clear that the 5G system offers a rich set of capabilities that can be leveraged to enhance existing media distribution services or to enable completely new immersive media services. However, not all enablers are fully in place.

The following is a summary of the 5G aspects and functions that are relevant for media distribution and a discussion of the gaps/missing enablers to leverage the full potential of the 5G system.

### 8.1.2 Support from AFs

Application Functions play a key role in media distribution over 5G. Different media streaming functionality such as session control, media data caching and delivery, QoE reporting, MooD, etc. may be realized as Application Servers with control functionality in AFs. Network assistance, such as SAND and congestion marking, are also supported by AFs in collaboration with other 5G functions such as the PCF or NEF. It is however not clear how the AF is instantiated and controlled by the content provider in a standardized and secure way.

For example, an Application to Network Interaction (ANIF) may be introduced to offer the possibility for UEs to interact with the 5G system in order to influence the delivery policy. It has been identified, that the new policy related function can leverage similar interaction concepts such as Network Assistance for DASH.

It is suggested to identify common network assistance functions, which are relevant for several media services such as 3GP DASH, 3GP Progressive Download, FLUS, etc.

Common functions might be rate recommendations (SAND Network Assistance, cf. TS 26.247, clause 13.6), Ad Insertion, Low Latency Support, or also dynamic policy configuration (introduced here). The new 5G Network Assistance Functions should contain identified common function. A component in the UE (independently whether FLUS or 3GP DASH or 3GP progressive download or a third-party service) can access Network Assistance Functions via this API.

### 8.1.3 Addressing Application Servers

To enable some new media distribution functionalities (e.g. edge computing and caching), traffic needs to be routed through the UPF, which can for instance be achieved through appropriate DNS resolution. It should be possible for the content provider or the AF to influence the DNS resolution of the DNS entries that belong to it to process data locally. Currently, AF is only able to influence traffic forwarding and routing by sending requests to route UP data to a DNN with a particular DNNI.

The feasibility of alternative solutions to DNS configuration, e.g. IP Anycast or Application Trigger, may also be considered and studied to address this gap.

### 8.1.4 QoS Management

the 5G System offers a set of standardized 5QI profiles and allows for non-standardized 5QI definition. QoS profiles can be matched to traffic and special treatment will then be applied to the traffic. Media distribution services should leverage the ability to define and apply QoS profiles to their traffic based on agreements with the operator.

### 8.1.5 Other Functions

Media distribution services over 5G systems may benefit from plenty of new functionalities offered by the 5G system. However, this requires that the third-party media service provider is able to influence the way its traffic is handled, as well as which media functionalities (e.g. the AFs mentioned above) should be deployed, where and in which capacity. For this, appropriate NEF APIs are required to enable the service provider to influence traffic routing to their ASs, traffic handling, traffic processing, and QoS. On the management plane (i.e. slice management API), the service provider should be able to monitor and scale (up or down) the network slice it is using (e.g. total bandwidth capacity, amount and capacity of deployed AFs).

In order to follow the 5GS design principles, SA4 should consider splitting Control and User Planes of the BM-SC and the xMB interface.

It is anticipated that collaboration with SA2 will be required to address the identified gaps. Enablers and APIs to offer the new 5G functionality to content and service providers as well as to UEs will also be needed.

## 8.2 Conclusions

Based on discussion in clause 7, it is proposed that for 5G Media different device APIs are considered:

- APIs that enable access to functions that are defined by 3GPP and are device internal.

- APIs that enable to abstract complex 3GPP network functions with abstracted APIs that can be accessed by third-party applications.

It is also recommended that 3GPP devices in 5G media provide a consistent support for capability discovery of 3GPP defined functionalities on the device using common capability discovery API as defined in Android or Browser environments. The capability aspects include codecs, formats, protocols, rendering capabilities and other aspects that are specified by 3GPP terminal capabilities.

Several aspects may have an impact on our specification structure. Here are suggestions for discussion:

- Invite key industry players and SDOs that are interested in 5G to share their usage scenarios and needs to be addressed in the 5G Media architecture. This includes interfaces and functions. This includes classical broadcast content providers, emerging media service providers, as well as mobile network operators running media services for mobile networks.

- Address different collaboration scenarios between third party service providers and mobile network operators for media distribution over 5G.

- Define the 5G Media Streaming architecture based on a function-based design and to leverage 5G features for existing and new media services. The architecture is expected to be supported by a function and API-based client design.

- Define the concept of a service and a session in the context of 5G Media Distribution, especially for third-party media services. Continue studying the development of corresponding network and client APIs to establish, announce and discover the assistance capabilities.

- A Media ingest point independent of the media distribution (e.g. Unicast or Broadcast) would simplify content and service provider interfacing to the 3GPP system. xMB TS 26.348 is obviously a candidate for such evolution. Client APIs would also benefit from transport-agnostic definitions.

- Identify relevant stage 3 interfaces and functions that need to be maintained in the 5G media distribution context for different ingest scenarios. Initial examples include well-defined profiles for DASH and ISO BMFF, potentially based on TS 26.247 3GP-DASH and File format TS 26.244. However, such efforts should also include and enable harmonization with external efforts, such as in MPEG DASH and CMAF, CTA WAVE or DASH-IF.

- Evolve the QoE framework and architecture to support different collaboration scenarios between the MNO, third-party content providers, and metrics collection service providers. Such work should align and harmonize with common industry practices.

- Support network assistance functions that permit dynamic updates of the policies, e.g. QoS, based on information provided by the service provider and/or the client.

- Enable the support of different distribution-related functions such as Low-Latency DASH, Ad Insertion, DRM framework, Event notifications, presentation format etc, in different collaboration scenarios.

- The RTSP/RTP part of the PSS service is no longer required. PSS TS 26.234 RTP/RTSP protocol need not be maintained. The PSS codec part could be ported in a new base media profile for Media Distribution.

- Define service agnostic media profiles used for download, progressive download and streaming based on TS 26.116 and TS 26.118, but also in alignment with other industry efforts such as CMAF. Identify the necessity of new media profiles and new media types to be defined in 3GPP based on different collaboration scenarios, for example the necessity of new audio or subtitle profiles.

- Extract the non-IMS FLUS APIs and define the corresponding architecture where FLUS Sinks are in the network. Evolve the FLUS APIs for the Contribution service vertical.

Annex A:  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Change history** | | | | | | | |
| **Date** | **Meeting** | **TDoc** | **CR** | **Rev** | **Cat** | **Subject/Comment** | **New version** |
| 2018-03 | #79 | SP-180027 |  |  |  | Presented to TSG SA #79 for information | 1.0.0 |
| 2018-12 | #82 | SP-180982 |  |  |  | Presented to TSG SA #82 for approval | 2.0.0 |
| 2018-12 | #82 |  |  |  |  | Approved at SA#82 | 16.0.0 |