Broadcast and Multicast Communication Enablers for the Fifth-Generation of Wireless Systems

Deliverable D4.2
Converged Core Network

Version 3.0 Date: 31/05/2019

Document properties:

Grant Number:	761498
Document Number:	D4.2
Document Title:	Converged Core Network
Editor(s):	Jon Hart (BT)
Authors:	David Gomez-Barquero (UPV); David Navratil (NOK); Andrew Murphy (BBC); Jon Hart (BT), Mael Boutin (Broadpeak); Roman Odarchenko (Bundeslab); Tuan Tran (Expway); Peter Sanders (one2many); Menno Bot (one2many).
Reviewers:	David Gomez-Barquero (UPV); David Navratil (NOK); Kevin Foster* (BT); David Thorne* (BT); Mael Boutin (Broadpeak); Roman Odarchenko (Bundeslab); Tuan Tran (Expway); Christian Menzel (IRT); Baruch Altman (LiveU); Peter Sanders (one2many); D'Aria Giovanna (TIM); Maria Pia Galante (TIM); Mauro Castagno (TIM), Toktam Mahmoodi* (King's College London).
Contractual Date of Delivery:	2019/06/30
Dissemination level:	CO[1]
Status:	Issue 3.0
Version:	3.0
File Name:	5G-Xcast D4.2_v3.0

Abstract

There is an increasing expectation from consumers to be able to use any service, at anytime, anywhere on any network access in a seamless manner. In order to meet this and other key requirements, the authors see the 5G System as an opportunity for industry to define a fully converged fixed and mobile architecture. This document describes the key drivers, benefits and use cases for full network convergence. An evaluation of current methodologies for providing partial network convergence is discussed. A number of key technologies for 5G were investigated including network slicing and caching mechanisms. As the work in standards bodies increases in this area, the current activities are outlined before highlighting some of the limitations of current mechanisms for partial convergence and how the 5G-Xcast architecture could address those gaps by fulfilling a set of key technical requirements. The document provides a number of alternative architectural solutions in order to address those key requirements.

[1] CO = Confidential, only members of the consortium (including the Commission Services)

PU = Public

* = External Reviewer

Keywords

5G, 5GC, fixed-mobile convergence, converged core, multilink, network slicing

Revision History

Revision	**Date**	**Issued by**	**Description**
1.0	2018/02/28	Jon Hart (BT)	Issue 1.0
2.0	2018/09/30	Jon Hart (BT)	External and internal review comments incorporated and final updates made.
3.0	2019/05/31	Jon Hart (BT)	Final version

Executive Summary

This document describes the key drivers for full network convergence between fixed and mobile networks. Full network convergence is defined in the introductory section as the "full integration of fixed and mobile access networks into a single end-to-end architecture, supporting a single set of identities, common authentication, seamless handover and simultaneous access across any access technologies". Seamless user experience and efficiency are the main benefits of full network convergence.

The document describes several of the existing mechanisms that achieve some form of (partial) network convergence, together with their main capabilities and examples of use cases where they can be applied.

The authors see the 5G System being an opportunity for industry to define a fully converged fixed and mobile architecture that can deliver, amongst others, the use cases presented in this paper. The document explains some of the ongoing efforts in standards bodies to fulfil that vision. A number of key technologies for 5G were investigated including network slicing and caching mechanisms.

A converged 5G system could enable operators to migrate their fixed and mobile customers onto a common platform if by doing so they are able to deliver both customer experience improvements and operational efficiencies.

Most importantly, this document describes a number of architectural solutions to enable a 5G converged core network architecture:

- A trusted architecture in which the UE and the UPF act as the points of convergence.
- A trusted architecture in which the 5G Residential Gateway and the UPF act as the points of convergence.
- An overlay solution which is less integrated with the core network than the previous two versions, which uses a Multi Access Proxy outside of the core and the UE as the points of convergence.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

3GPP	Third Generation Partnership Project
5G/RG	5G Residential Gateway
AF	Application Function
AGF	Access Gateway Function
AMF	Access and Mobility Management Function
AN	Access Network
ANDSF	Access Network Discovery and Selection Function
AP	Access Point
AR	Augmented Reality
ATSC	Advanced Television Systems Committee
ATSSS	Access Traffic Steering, Switching and Splitting
AUSF	Authentication Server Function
AWS	Amazon Web Services
BBF	Broadband Forum
BNG	Broadband Network Gateway
BNO	Broadcast Network Operator
BSP	Broadcast Service Provider
CDN	Content Delivery Network
CO	Central Office
CP	Control Plane
CP-AT3SF	Control Plane Access Traffic Steering Switching and Splitting Function
CPE	Customer Premise Equipment
DAS	Direct Attached Storage
DASH	Dynamic Adaptive Streaming over HTTP
DC	Dual Connectivity
DN	Data Network
DNS	Domain Name System
DSL	Digital Subscriber Line
DSLAM	Digital Subscriber Line Access Multiplexer
DTT	Digital Terrestrial Television
EC	European Commission
eMBMS	Evolved Multimedia Broadcast/Multicast Service
eNodeB	Evolved Node B
EPC	Evolved Packet Core
ePDG	evolved Packet Data Gateway
FAGF	Fixed Access Gateway Function
FMC	Fixed Mobile Convergence
FMIF	Fixed Mobile Interworking Function
FN/RG	Fixed Network Residential Gateway
FTTC	Fibre to the Cabinet
FTTH	Fibre to the Home
GPON	Gigabit Passive Optical Network
HD	High Definition
HLS	HTTP Live Streaming
HPHT	High Power, High Tower
HTTP	HyperText Transfer Protocol
KPI	Key Performance Indicator
IoT	Internet of Things
IP	Internet Protocol
IPSec	Internet Protocol Security
ISP	Internet Service Provider
IPTV	Internet Protocol Television
LAA	License Assisted Access
LAN	Local Access Network
LPLT	Low Power, Low Tower
LTE	Long Term Evolution
LWA	LTE Wi-Fi Aggregation

LWIP	LTE WLAN Radio Level Integration with IPsec Tunnel
M&E	Media & Entertainment
MANO	Management and Orchestration
MBMS	Multimedia Broadcast Multicast Service
MC	Multi Connectivity
MEC	Multi-access Edge Computing
MIO	Mobile Infrastructure Operator
ML	Multilink
MMTP	Multimedia Multiplexing Transport Protocol
MPEG	Moving Picture Experts Group
MPTCP	Multi-Path Transmission Control Protocol
MSE	Multi-Service Edge
NEF	Network Exposure Function
NFV	Network Functions Virtualisation
NRF	Network Routing Function
NSSF	Network Slice Selection Function
OLT	Optical Line Terminal
ONT	Optical Network Terminal
OSS	Operational Support System
OTT	Over The Top
PC-AT3SF	Policy Control Access Traffic Steering Switching and Splitting Function
PCF	Policy Control Function
PON	Passive Optical Network
PoP	Point of Presence
PTM	Point-To-Multipoint
PTP	Point-To-Point
QoS	Quality of Service
RAM	Random Access Memory
RAN	Radio Access Network
RAT	Radio Access Technology
RCLWI	RAN-controlled LTE-WLAN interworking
RG	Residential Gateway
ROM	Receive Only Mode
SA	System Architecture
SAN	Storage Area Network
SAND	Server and Network Assisted DASH
SAS	Serial Attached SCSI
SATA	Serial Advanced Technology Attachment
SCSI	Small Computer System Interface
SD	Slice Differentiator
SDD	Solid State Drive
SDN	Software-Defined Networking
SIM	Subscriber Identity Module
SLA	Service Level Agreement
SMF	Session Management Function
SSC	Session and Service Continuity
UDM	Unified Data Management
UDR	Unified Data Repository
UDR-AT3SF	User Data Repository for Access Traffic Steering Switching & Splitting Function
UDSF	Unstructured Data Storage Function
UE	User Equipment
UE-AT3SF	UE Access Traffic Steering Switching and Splitting Function
UMTS	Universal Mobile Telecommunications Service
UP	User Plane
UP-AT3SF	User Plane Access Traffic Steering Switching and Splitting Function
UPF	User Plane Function
VAS	Value Added Service
VDSL	Very-High-Bit-Rate Digital Subscriber Line
VoD	Video on Demand

W-5GAN	Wireline 5G Access Network
WAN	Wide Area Network
Wi-Fi	Wireless Fidelity
WiMax	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Access Network

1 Introduction

The 5G-Xcast project is focused on content distribution and delivery over 5G networks to address the continued growth of content services and requirements for point-to-multipoint delivery. 5G-Xcast aims to define a unified delivery architecture which can dynamically exploit unicast, multicast and broadcast delivery modes as well as local caching. 5G-Xcast technologies will be fundamental in the progression towards a converged 5G architecture to provide a seamless user experience over fixed and mobile in this case for content delivery services.

"5G-Xcast Deliverable D2.1 Definition of Use Cases, Requirements and KPIs" [1] has defined use cases that demonstrate and specify the applicability and requirements. At the same time, 3GPP on the mobile side and fixed network standardisation bodies on the other side (BBF, DVB) have been working on convergence. This deliverable provides the framework for 5G-Xcast in view of this work and the project requirements. This document will focus on the single operator scenario with regard to convergence. In this deployment scenario the single operator owns and operates the infrastructure i.e. the fixed network and the mobile network. The CDN could be owned by either the same operator or by a 3rd party CDN provider depending on the scenario being discussed.

The telecoms industry is undergoing a major transformation towards 5G networks in order to fulfil the needs of existing and emerging use cases (e.g. as defined in [3]). Traditionally, fixed and mobile networks have been deployed as separate systems delivering independent but often similar services such as broadband, media and entertainment, messaging, voice and video communications, etc. The 5G architecture should seamlessly support both fixed and mobile access technologies under a fully converged end-to-end system in order to deliver future service requirements.

There are many different definitions of fixed mobile convergence, including:

- ***Products and services***: combining fixed and mobile broadband on a single bill or service package.
- ***Channels to market***: reusing the same sales channels, marketing or branding to sell fixed and mobile services.
- ***Customer service***: having common contact centres and support channels across fixed and mobile services.
- ***Platform convergence***: reuse of common infrastructure, e.g. virtualisation layer, hardware or transport network resources across multiple services.
- ***Network convergence***: full integration of fixed and mobile access networks into a single end-to-end architecture, supporting a single set of identities, common authentication, seamless handover and simultaneous access across any access technologies.

All forms of convergence above are expected to provide significant benefits to network operators. However, there are additional benefits that network convergence could offer, such as:

- ***Improved and Seamless User Experience*** – increased mobility, higher throughput (upstream and downstream), lower latency and jitter, higher reliability and faster provisioning are becoming key requirements to certain applications. The ability to seamlessly move between or simultaneously combine fixed and mobile access networks including indoor and outdoor deployments can help telecoms operators fulfil these requirements. Also, offering a unified set of identities and a consistent experience across all networks.

- ***Operator Benefits*** – Further network optimisation, platform convergence can provide significant benefits in this space, however, full network convergence can improve operators' ability to make the best use of network assets, including often scarce access network resources, thus helping them to become more efficient when delivering services to a wider range of customers. Other benefits include the best use of networks, improved reliability, asset reuse, simplified OSS, new service and revenue opportunities, etc.

[3] Provides further background information regarding the current approaches in regard to convergence.

The rest of the document is structured as follows:

Section 2 provides several user experience-driven service deployment scenarios that require network convergence.

Section 3 describes several of the existing mechanisms that achieve some form of (partial) network convergence. It also provides a summary of these, together with their main capabilities and examples of use cases where they can be applied.

Section 4 provides an overview of the fixed network architectures for both Fibre to the Cabinet (FTTC) and Fibre to the Premises (FTTP) deployments.

Section 5 explains some of the ongoing efforts in standards bodies to fulfil the convergence vision. A number of key technologies are then considered; network slicing, caching technologies and seamless content delivery.

Section 6 outlines the main architectural options for the 5G-Xcast converged network.

Section 7 provides an overview of the considerations and approach required in order to achieve a converged broadcast network.

Note that this document is written in conjunction with "5G-Xcast Deliverable 4.1 Mobile Core Network" [4] and "5G-Xcast Deliverable 4.3 Session Control and Management" [5].

2 Deployment Scenarios

There are several user experience-driven service deployment scenarios that require full “network convergence” across fixed and mobile technologies. This section outlines some examples; Operators may have a number of different scenarios and drivers for convergence. The authors do not necessarily see these examples being “mass-market” today, however, some Operators may want to deploy these or other converged services to their subscribers in the future. These scenarios assume the case of a single network operator across fixed broadband and mobile networks.

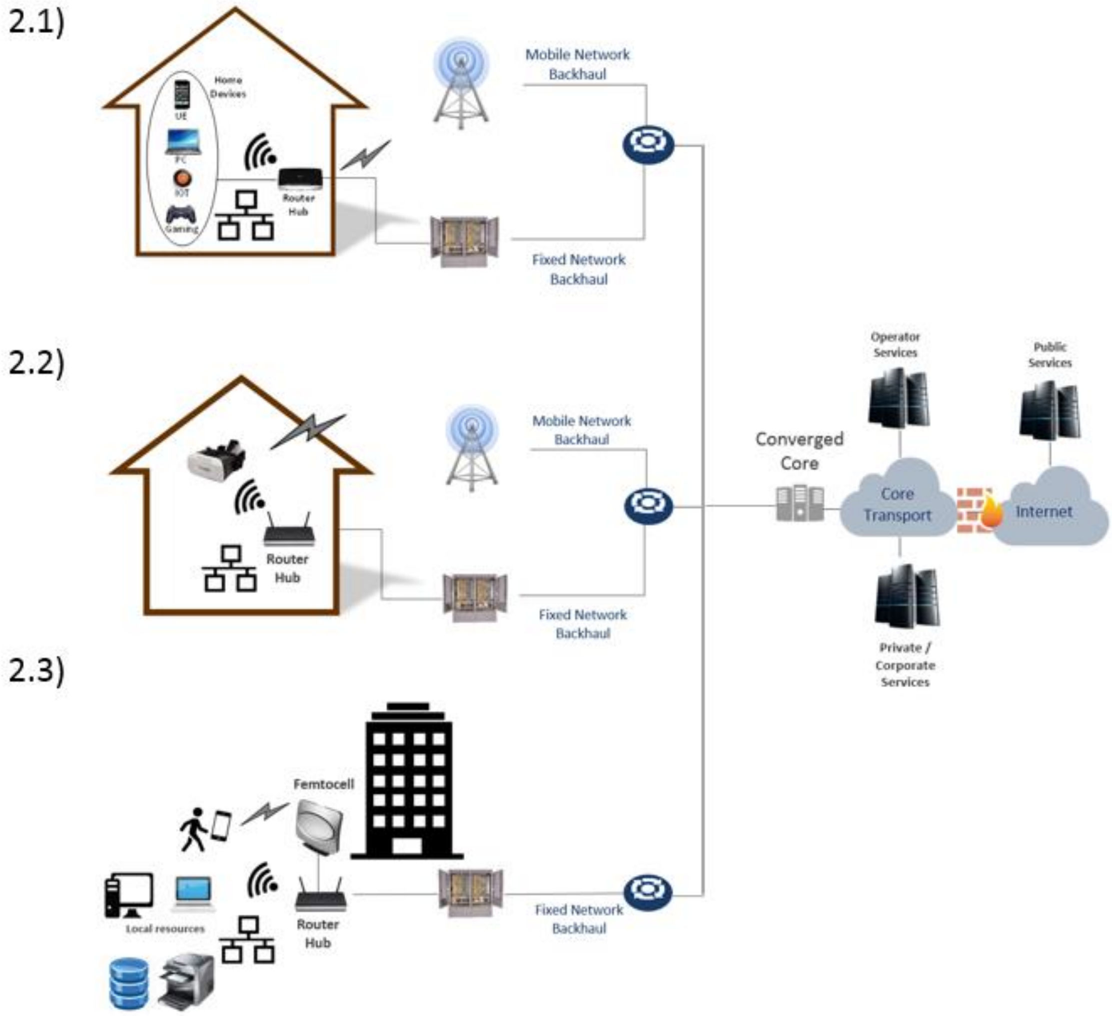


Figure 1 – Convergence Use Cases

The aim of these deployment scenarios is to deliver a seamless user experience. This includes the ability for the device to seamlessly move across different available access networks, without any user intervention, in order to deliver the best possible service experience in the most efficient manner for the Network Operator.

2.1 Fixed Hybrid Access

In this scenario the customer Home Gateway device has two WAN interface connections to the core network: a fixed broadband interface (e.g. xDSL, fibre, cable, etc.) and a cellular radio access connection. All home devices connect to the network via the Home Gateway's LAN connections (e.g. Wi-Fi, Ethernet, etc.) and they are not visible to the Network Operator.

The key advantages and value-add features that can be provided to the end user in this scenario are:

- Simultaneous use of multiple access networks - This traffic scenario allows constant and simultaneous use of both fixed broadband access and cellular access in such a way that the end result data rate is close to the dynamic sum of both fixed and cellular data rates in a continuous manner. A common implementation of this functionality would be to only use the cellular access when it is required i.e. for traffic peaks, as a load-balancing mechanism, or as a top-up as per Operator policy.

- Bandwidth boost - This traffic scenario allows the on-demand use of cellular access to provide bandwidth boost to fixed broadband access. Users can trigger the bandwidth demand according to number and type of applications, time of day and type of UEs. This allows users to control their tariffs and increase their QoE. For Operators it enables the ability to increase revenues by upselling dynamic and temporary on-demand bandwidth for a specific time duration with a higher QoE of content (e.g. AR+ 4K video game streaming vs. full HD). As a minimum, users should have the ability to select on-device turbo boost purchases via a smartphone app/webpage portal.

- Failover with session continuity - Use of cellular access as failover mechanism in the case where fixed broadband goes out of service and vice versa.

- Fast provisioning - Use of cellular access as fast provision service whilst users wait for their fixed broadband to be deployed or activated.

- Symmetric bandwidth - This traffic scenario provides the end user with the same high data rates in the upstream direction as in the downstream direction. The faster uplink speeds can be used for cloud services, media and photo upload, etc.

2.2 Mobile Hybrid Access

In this scenario, it is the end device that has the two WAN interface connections to the core network. A radio access connection via the cellular network access and a Wi-Fi/Ethernet connection via a fixed broadband gateway. In this scenario, the UE is fully visible to the core on both the cellular and fixed connections simultaneously.

The key advantages and value-add features that can be provided to the end user are similar to the previous use case:

- Simultaneous use of multiple access networks – as above.
- Bandwidth boost – as above.
- Failover with session continuity – as above.
- Symmetric bandwidth – as above.

- Seamless mobility – The device should be able to seamlessly transfer across different available access networks, without any user intervention, in order to deliver the best possible service experience in the most efficient manner for the Operator.
- User centric policy implementation.
- Bandwidth Utilisation.

2.3 Hybrid Access over a Common Backhaul

In this scenario a femtocell is physically connected to a fixed broadband gateway. The femtocell provides a cellular radio access connection and the fixed broadband gateway provides a Wi-Fi/Ethernet connection. Both access networks share the same fixed broadband backhaul. Users are able to access both Local Area Network services as well as network-provided private corporate services from either the Femtocell-provided cellular radio access or the Wi-Fi connection with the same user experience. In this scenario, the UE is fully visible to the core on both the cellular and wireless connections (i.e. the Residential Gateway is acting as a Relay Node).

The key advantages and value-add features that can be provided to the end user are:

- Unified set of identities - In this scenario an Operator shall be able to manage a unified set of identities for a single user in order to consolidate subscriber data to allow seamless access across fixed and cellular access networks, including cellular access over fixed broadband.

- Consistent set of policies - In this scenario an Operator may apply a consistent set of policies (e.g. QoS and traffic management policies) to traffic belonging to a particular user across all networks, including cellular access over fixed broadband, in order to deliver a consistent user experience.

- Same set of services - In this scenario once an Operator has identified a user, it may then grant that user access to a single set of services (e.g. VAS, parental controls, content, content optimisation and access to local services when appropriate) when using cellular, fixed or cellular over fixed broadband access. This has the advantage for the Operator to consolidate services, and for the end user with regard to constancy of service offering across all access types.

- Access to Local Area Network services - Access to Local Area Network services (e.g. peripherals, content servers) shall also be possible from the Small Cell deployed in the customer premises.

- Seamless mobility – as above.

Note: for the deployment scenarios outlined above there may be an additional delay introduced by the UE / CPE when processing multiple streams, a process that is non-trivial.

3 Existing mechanisms for partial network convergence

Standards bodies have either defined or are in the process of defining a number of architectural approaches to achieve different forms of network convergence between fixed and mobile technologies. This section outlines some of these approaches.

Figure 1 below shows how the mobile hybrid access deployment scenario can be partially addressed by current technical solutions.

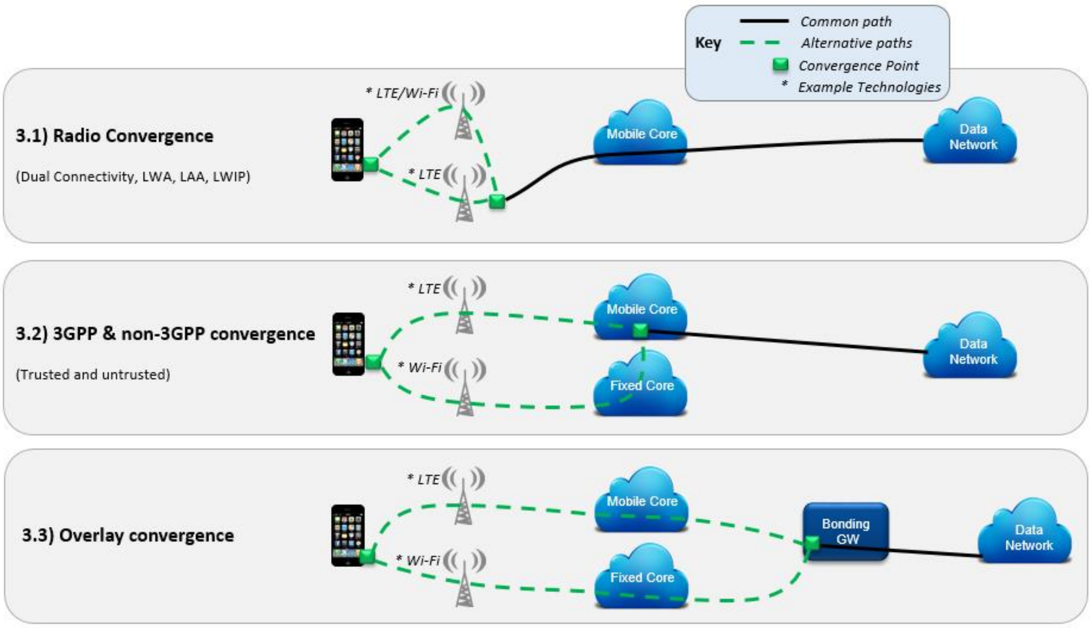


Figure 2 – Existing Architectural Options for Partial Network Convergence

3.1 Radio Access Convergence

A number of technical solutions have been defined to bring together different radio technologies, such as cellular (e.g. LTE) and wireless (e.g. Wi-Fi) to deliver improvements to the end user as a result of radio level "diversity". These include increased capacity, throughput or improved mobility within the coverage area of the single cell or access point. This set of solutions requires a combination of fixed and mobile technologies together with some form of backhaul solution between the two radio access nodes:

- ***Dual Connectivity*** is a capability where a UE can simultaneously connect to, and deliver traffic over, two different radio cells, typically a LTE macro and LTE small cell. The macro cell operates as the master, dynamically controlling how data is delivered to the UE over the two paths.
- ***LTE Wi-Fi Aggregation (LWA)*** is very similar to Dual Connectivity except that the small/secondary cell uses Wi-Fi radio technology instead of LTE [6]. Typically the LTE base station (eNodeB) and Wi-Fi Access Point (AP) will be co-located. Both the Wi-Fi AP and the UE need to be LWA capable.
- ***Licence Assisted Access (LAA)*** is a form of Carrier Aggregation (CA) where LTE protocols are carried over both the LTE and 5GHz Wi-Fi spectrum [6]. This needs to behave fairly with respect to Wi-Fi users.

- ***LTE WLAN Radio Level Integration with IPsec Tunnel (LWIP)*** is a similar approach to LWA except that the Wi-Fi infrastructure does not require upgrading [6]. Here, a Wi-Fi access point will connect to a LWIP gateway which then delivers traffic over an IPSec tunnel to the LTE eNodeB. Due to the nature of the interconnect between the Wi-Fi and LTE access, this can only operate at the IP flow level and not lower levels of the protocol stack like other approaches.
- ***RAN-controlled LTE-WLAN interworking (RCLWI)*** allows the 3GPP RAN to instruct the UE to connect to the WLAN upon receipt of measurement information including Wi-Fi signal strength [6].
- **RAT multi-connectivity (MC)** is envisioned as a way to tackle effects of increased losses and a worse coverage by connecting to multiple radio access technologies simultaneously, allowing users to benefit from each air interface's advantages (e.g. LTE + 5G NR) [6].

3.2 Convergence of 3GPP and Non-3GPP Access

3GPP has already defined a number of mechanisms to allow non-3GPP access networks (typically Wi-Fi) to be connected into an existing 3GPP core (e.g. EPC). These mechanisms allow for many different levels of integration between the two networks ranging from simple SIM based authentication on to a Wi-Fi network through to anchoring non-3GPP user plane traffic in the EPC including policy-based traffic steering. The 3GPP standards describe two types of non-3GPP access:

- ***Trusted non-3GPP Access*** - where the fixed access network provides security of the traffic from the UE through to the mobile core. This approach requires some integration between the 3GPP and non-3GPP networks and so typically both will be owned by the same operator.
- ***Untrusted non-3GPP Access*** - where a UE can access the mobile core through any non-3GPP network. Here, it is the UE's responsibility to ensure security through the non-3GPP access, which is achieved by delivering all traffic through a secure tunnel between itself and an evolved Packet Data Gateway (ePDG) residing in the mobile core. Typically, this approach is used at the application level (e.g. voice over Wi-Fi) where the application establishes the secure connection.

3GPP has also defined a mechanism for operators to provide policy rules to a device in order to assist the process of discovering and selecting the most appropriate access network to connect to, as well as how to route traffic across multiple networks when applicable. The biggest challenge is finding a way to assist the mobile node to intelligently determine and select the most preferable Point of Service, anytime, anywhere so that users can get best access network for different individual data service. The function element involved in this mechanism is Access Network Discovery and Selection Function (ANDSF). ANDSF is a Core Network function within the EPC (Evolved Packet Core) which assists UEs in the discovery/selection of non-3GPP access networks (e.g. Wi-Fi) in their vicinity, and provides them with rules policing the connection to these networks.

This functional element can send a set of policy rules with the UE over a dedicated interface, which the UE can use to decide which network(s) it should connect to from those that are currently visible.

As mentioned above, there are also rules to assist the UE when routing traffic over different access networks. It is important to note that ANDSF rules for access network selection are at present focused on 3GPP versus non-3GPP access (e.g. 3GPP access takes priority over WLAN), allowing for prioritisation amongst WLAN networks available (e.g. WLAN_X takes priority over WLAN_Y); however, it does not provide prioritisation between different 3GPP radio access technologies and specific WLAN networks (e.g. it does not allow for LTE to take priority 1, WLAN_X priority 2 and UMTS priority 3). It is also mainly based on static policies and does not provide a mechanism to transmit network conditions to the UE, relying on the UE to determine these for itself.

ANDSF provides the following benefits:

- Seamless Connectivity:

  A) Supports intelligent offload between 3G/4G to WI-FI and vice – versa, by the operator that manages both networks

  B) Enables operators to support location based, time, UE profile, battery, signal strength based intelligent offload

  C) Intelligent enough to dynamically select optimal access points or select between the cellular or Wi-Fi network as the user moves to different location

- Improved performance:

  A) Enable smart device to discover non 3GPP access networks (Wi-Fi & WIMAX) to enforce the user policies

  B) Saves battery life

  C) Help define roaming policy at operator / network level in co-ordination with HS 2.0

- Enhance Customer Experience:

  A) User centric policy implementation

  B) Efficient bandwidth utilisation

In addition to the above, the 3GPP RAN can provide **RAN assistance information** to the UE [7], [8] including 3GPP and Wi-Fi access rules and offload preference indicators. These can also be used by the UE to select the most appropriate access network to connect to. The way those policies interact with ANDSF policies is defined in [9].

Many of the mobile devices sold today are dual radio (i.e., include both cellular and Wi-Fi radio) and are capable of using both radios simultaneously. However, most manufacturers do not enable this functionality by default due to the impact on battery life. As the battery life problems are resolved, simultaneous access to Wi-Fi and 3GPP means that there is the opportunity to direct certain services to Wi-Fi, and others to 3GPP access. However, as more and more devices are capable of operating on multiple technology types (e.g., 3G, 4G, Wi-Fi), the more important intelligent network selection and traffic steering becomes.

To tackle this issue, ANDSF and its enhancement aims to provide policy driven intelligent network selection and traffic steering. A key enabler of this approach is the ability to implement an ANDSF client in devices that communicates with an ANDSF server in the network and supplies the ANDSF policy to the functionality in the device that performs network selection and traffic steering decisions. By distributing tailored policies to the ANDSF client, operators are able to steer traffic between Wi-Fi and cellular for better user experience and to allow better utilisation of network and radio resources.

Currently, it is the understanding of the authors that ANDSF procedures have had very limited adoption in the devices market. The main limitations of the ANDSF approach are that it is very detached from the real time application and content delivery direct performance needs (bandwidth, latency, reliability). However, there are some examples of “application-based” ANDSF implementations offering a set of limited functions.

In addition, as mobile devices have become more intelligent, they are the entity that is most aware of actual network connectivity conditions (e.g. radio conditions, throughput over existing connectivity, etc.), and real-time conditions in the device (e.g. type of pending traffic, active applications, status of device including battery levels, etc.). With the combination of operator policy and network/device intelligence, it is now the device that is in the unique position to make the best determination of which traffic should be transported over what access type (e.g., Wi-Fi or cellular).

The main drawbacks with this type of convergence is that still it very much depends on single operator, dual networks only, mostly resolves offloading but limited bandwidth and reliability benefits, non-awareness of application level performance and therefore the matching between bandwidth reliability and latency to actual application and content needs, limited mobility and implementation within the network(s) elements themselves.

3.3 Overlay Convergence

A number of solutions have emerged in industry looking to provide some convergence over the existing Fixed and Mobile core network infrastructure. In most solutions, a gateway is required to aggregate the traffic and can be deployed in either in the operator’s network or in a 3rd Party Network. An application is needed in the UE to perform the aggregation function. A simple example of this type of convergence is through the use of Multi-Path TCP (MPTCP) as used by Apple’s Siri. There are varying levels of complexity and integration possible with this approach. The main reason to refer to these as “overlay” solutions is that they are not necessarily integrated into the existing fixed or mobile core (even if deployed inside the operator’s network) but work as an overlay layer, typically using some form of tunnelling mechanism over existing, unmodified networks. Further, at these layers the real time application performance needs are visible to the algorithms and can be matched by the network connections real time performance.

Other solutions provide convergence on top of the IP stack, just below the application layer. For example, solutions in which the bonding mechanism handles any IP connection and matches the actual application level performance, to the overall performance of all links in real time.

For more information about overlay convergence, multilink and multi-connectivity please see 5G-Xcast D4.1 ‘Mobile Core Network’ deliverable [4].

3.4 Summary of Existing Partial Network Convergence Solutions

Table 1 below provides a summary of existing solutions to deliver partial network convergence, including the capabilities that they deliver as well as some example deployment scenarios that they can be applicable to.

Technology	Capabilities	Example deployment scenarios
Trusted non-3GPP Access	Offload traffic on to cheaper access, whilst providing same end user services and policy control. No specific changes required to UE	**Carrier Wi-Fi** Fixed access network requires some integration with mobile core. Policy control of underlying fixed network is independent of mobile network
Untrusted non-3GPP Access	Offload traffic on to cheaper access, whilst providing same end user services and policy control. Works over any fixed access network.	**Voice over Wi-Fi** UE requires client to secure traffic to core network. Client configuration not straightforward, hence use tends to be application specific.

Dual Connectivity	Carrier aggregation provides greater download speeds and resilience.	**Urban deployment** Requires overlapping cells i.e. small cell deployment. Increased speeds only in downlink. Increased network and handset complexity.
LWA	Carrier aggregation provides greater download speeds. Offload of traffic onto cheaper access	**Enterprise deployment** eNodeB and Wi-Fi AP need to be in close proximity. Increased speeds only in downlink. Increased network and handset complexity.
LAA	Carrier aggregation provides greater download speeds. Offload of traffic onto cheaper access	**Enterprise deployment** Needs to behave fairly with existing Wi-Fi deployments. Increased network and handset complexity.
LWIP	Carrier aggregation provides greater download speeds. Offload of traffic onto cheaper access. No changes required to Wi-Fi access network.	**Enterprise deployment** Increased speeds only in downlink. Increased network and handset complexity. Can only route traffic at the IP flow level.
RAT multi-connectivity	Carrier aggregation provides greater download speeds.	Requires overlapping cells Increased speeds only in downlink. Increased network and handset complexity.
Multi-link	Carrier aggregation provides greater download speeds, higher reliability. Offload of traffic onto cheaper access. No changes required to Wi-Fi access network.	Requires new mechanisms of bonding in core network.
Overlay	Deployment is independent of network operator infrastructure. Dual connected nature offers resilience especially for mobility. Increased download speeds.	**Boosted broadband.** **Fast provisioning of residential Gateways.** **Personal Assistants (Siri)** May only support certain types of traffic depending on approach e.g. MPTCP. Limited integration with underlying networks likely to affect performance.
ANDSF	Operator policies of assisting UEs in the discovery/selection of most appropriate access networks to provide them with rules policing the connection to these networks	**Public Wi-Fi** Rules define prioritisation between 3GPP and non-3GPP access, Prioritisation of different WLAN access networks also supported. No support for priorities between specific 3GPP RATs and non-3GPP access networks.
RAN-assisted WLAN interworking	RAN-provided information to help the UE select the most appropriate access network to connect to	**Private Wi-Fi** The cellular network may provide rules for the UE to select WLAN or cellular when appropriate. Two separate networks (WLAN and mobile) still have to be maintained.
RAN-controlled LTE-WLAN interworking	The Radio access network can steer the UE to connect the WLAN upon measurement report information	**Private Wi-Fi** The cellular network may instruct the UE to connect to WLAN upon receipt of measurement reports. Two separate networks (WLAN and mobile) have to be maintained.

Table 1 - Summary of Existing Partial Network Convergence Solutions

4 Fixed Network Architecture

This section provides an overview of the fixed network architectures for both Fibre to the Cabinet (FTTC) and Fibre to the Premises (FTTP) deployments.

4.1 Fibre to the Cabinet (FTTC)

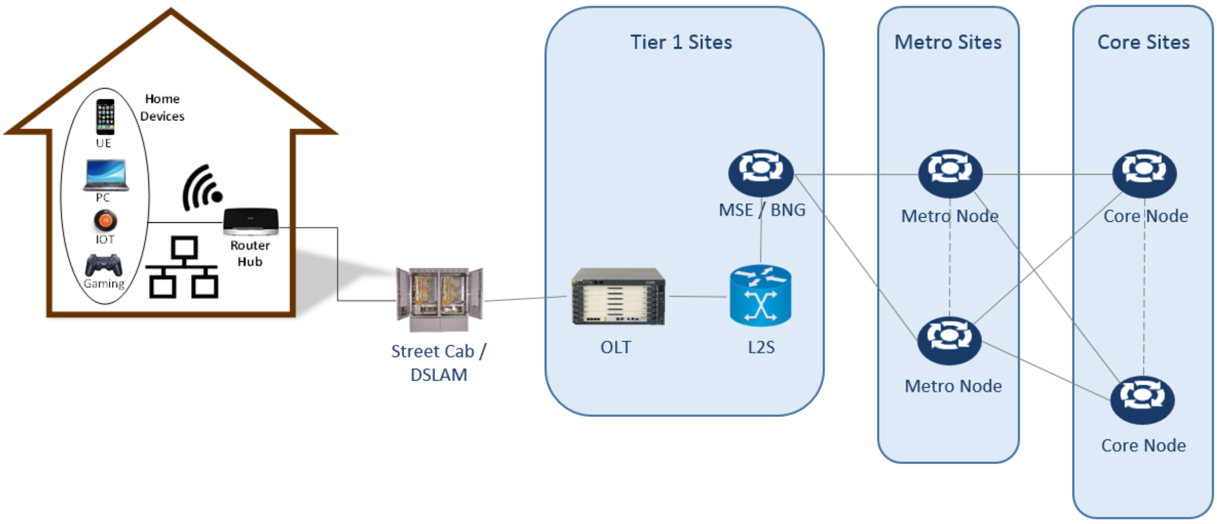


Figure 3 – FTTC Architecture

Figure 3 shows the FTTC architecture which is comprised of the following components:

- ***Home Gateway (Router/Hub)*** – This device has a WAN interface (VDSL) connection to the DSLAM. All home devices connect to the network via the Home Gateway's LAN connections (e.g. Wi-Fi, Ethernet, etc.)
- ***DSLAM*** – Digital Subscriber Line Access Multiplexer. The operator's DSLAM equipment will collect data from several home gateways and multiplexes the synchronised composite signal of both voice information and connection data. This traffic is routed through the backbone switch through an Access Network which connects through the Internet Backbone. The DSLAM is designed to simply network the connections as appropriate and provides an access point between the customer and the ISP.
- ***OLT (Optical Line Terminal)*** – This acts as the operator endpoint of a passive optical network (PON). It provides two main functions:
  - Converts the standard signals used by the operator's equipment to the frequency and framing used by the PON system.
  - Coordinates the multiplexing between the conversion devices on the other end of that network.
- ***MSE / BNG – Multi-Service Edge / Broadband Network Gateway*** - The access point for subscribers, through which they connect to the broadband network. When a connection is established between the Home Gateway and the BNG, the subscriber can access the broadband services provided by the operator. The BNG connects and operates subscriber sessions. During an active session, the BNG aggregates traffic from various subscriber sessions from an access network, and routes it to the operator's network. The BNG is deployed by the operator and is usually located at the first aggregation point in the network. The BNG manages subscriber access, and subscriber management functions such as:
  - Authentication, authorisation and accounting of subscriber sessions

- Address assignment
- Security
- Policy management
- Quality of Service (QoS)

4.2 Fibre to the Premises (FTTP)

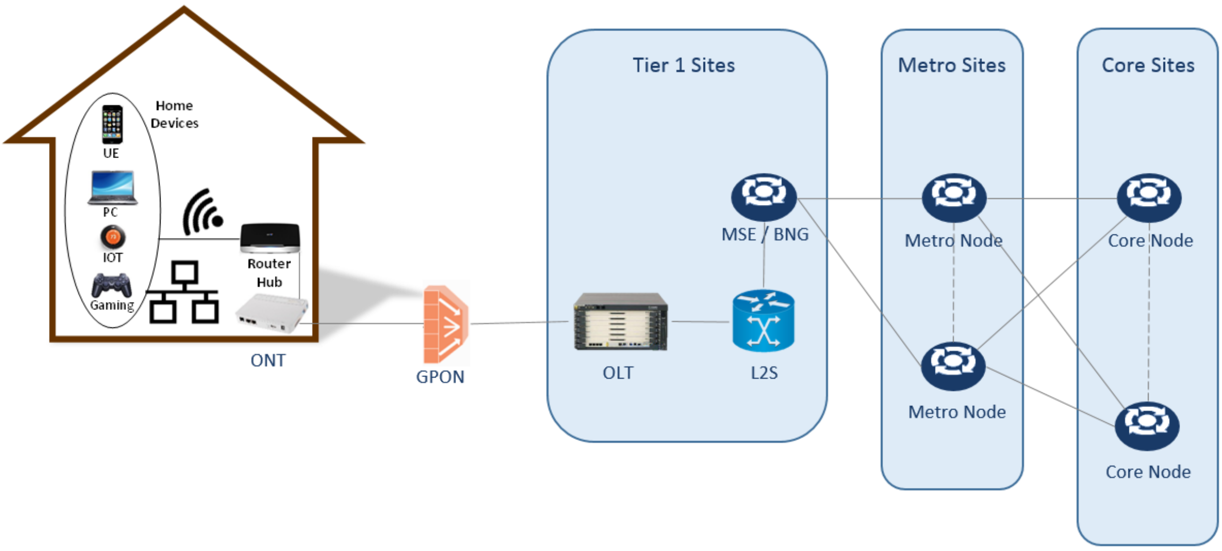


Figure 4 – FTTP Architecture

The FTTP architecture is comprised of the following components:

- ***Home Gateway (Router/Hub)*** – In the FTTP scenario, this device has a WAN interface (Ethernet) connection to the ONT. Both the Home Gateway and the ONT are collocated at the customer premises. All home devices connect to the network via the Home Gateway's LAN connections (e.g. Wi-Fi, Ethernet, etc.)
- ***ONT*** (Optical Network Terminal) – is used to terminate the fibre optic line, demultiplex the signal into its component parts (voice telephone and Internet access), and provide power to customer telephones.
- ***GPON Splitter*** – Splitters are installed in each optical network between the OLT and ONT that the OLT serves. The splitter can be deployed in the central office (CO) alongside the OLT, or it may be deployed in a cabinet closer to the subscribers. Passive optical splitter plays an important role in Fibre to the Home (FTTH) networks by allowing a single PON network interface to be shared among many subscribers.
- ***OLT (Optical Line Terminal)*** – as above.
- ***MSE / BNG – Multi-Service Edge / Broadband Network Gateway*** – as above.

5 5G Network Convergence Architecture

5.1 Ongoing Work in 3GPP

There are a number of Release 15 3GPP activities in this space and this document will focus on the following activities:

- Study Item on Media Distribution in 5G
    - Study of existing and new 3GPP media services and how they are mapped on the new 5G System and Architecture. Baseline architecture is Release 15 (no Wireless / Wireline convergence considered).
- Study on the Wireless and Wireline Convergence
    - Study on how to enhance the common 5G core network defined in TS 23.501 [10] and TS 23.502 [11] in order to support wireline access networks and Trusted N3GPP.
- Study on Access Traffic Steering, Switching and Splitting
    - Study on harmonised traffic handling across 3GPP and non-3GPP accesses, considering both untrusted and trusted non-3GPP access.

5.1.1 Study Item on Media Distribution in 5G

5.1.1.1 Use Cases

3GPP SA#76 (June 2017) has approved an SA4 Study Item on Media Distribution in 5G (acronym: FS_5GMedia_Distribution). The study item is based on the Feasibility Study on New Services and Markets Technology Enablers (SMARTER) which was documented by SA1 in 3GPP TR 22.891 [12]. The results of the study, that is currently still in progress, will be published in 3GPP TR 26.891 v16.x.y .

Among others, the following use cases from TR 22.891 will have a significant impact on the media handling procedures to enable them:

1. Broadcast
    a. Broadcasting Support (5.56)
    b. Ad-Hoc Broadcasting (5.57)
    c. Broadcast/Multicast Services using a Dedicated Radio Carrier (5.70)
2. Mission Critical Communications (5.3)
3. Virtual Presence (5.11)
4. Vehicular Internet and Infotainment (5.53)
5. Other Generic Use Cases that are motivated by media services
    a. Network Slicing (5.2)
    b. Best Connectivity per Traffic Type (5.26)
    c. On-demand Networking: streaming video in a stadium (5.7)
    d. Flexible Application Traffic Routing: AR application with edge computing (5.8)
    e. Multi-Access Network Integration: multi-path e.g., with Wi-Fi and 5G simultaneously (5.27)
    f. In-network and device caching (5.36)
    g. ICN-based Content Retrieval (5.38)

Emerging from the SMARTER study, 3GPP TR 22.863 [13] documents a feasibility study on New Services and Markets Technology Enablers and then specifically on Enhanced Mobile Broadband (eMBB). Section 5.6 in 3GPP TR 22.863 describes the scenario for fixed-mobile convergence, which is also described in section 5.3 in the present document.

3GPP TS 22.261 [14] specifies the 5G system for release 16, which contains only the requirements intended for media distribution for linear TV services (6.13).

Most of these requirements have already been met in the 4G system and none of these requirements address convergence. Other requirements for media distribution will follow in a later release.

At the time of writing, the SA4 study item on Media Distribution in 5G will address the service aspects, both on the network side and the UE side, and also architecture aspects in liaison with SA2, based on the SA1 study in TR 22.891. The study on Media Distribution in 5G is documented in 3GPP TR 26.891.

5.1.1.2 Mapping to 5G System Issues

Section 4 of 3GPP TR 26.891 provides a general overview of the 5G architecture, while section 5 analyses the mapping of existing media services based on the potential extended architecture as shown in Figure 4.

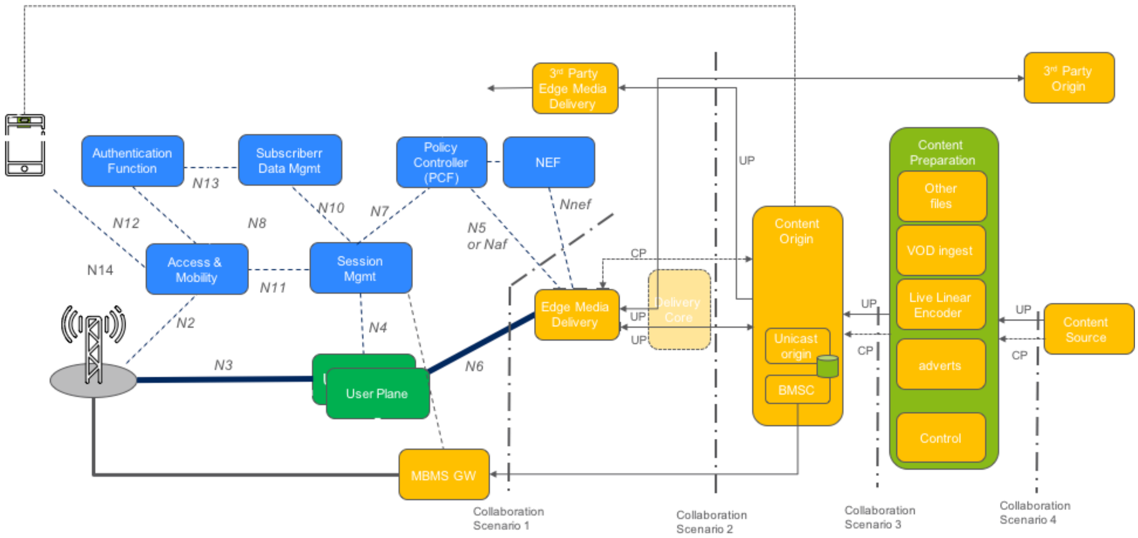


Figure 5 – Media on 5G System Architecture

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 study:

- On what grounds shall a UE be redirected to an appropriate edge cache, which cache is the appropriate cache and how long shall this redirection remain valid (when the UE moves)?
- The xMB interface, accessible over NEF, may need to support signalling of DNS entries of edge caches and other relevant information for media handling.
- An AF (e.g. XCF) may influence traffic routing in 5G networks. How are 3rd parties (e.g. CDNs) envisioned to use this functionality?
- When SAND (or MultiLink) is deployed then how can it be assured that the resources are also actually available/reserved for these clients?

5.1.1.3 Potential Mapping to 5G System

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

Media Streaming Function	5G System Function	5G System Interface to	Description

Streaming Server - Control Plane	AF (XCF)	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 (XUF)	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 (XCF/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.

Table 2 – Potential mapping of Media Streaming Functions to 5G System Functions

5.1.1.4 External Interface Aspects

To address the gaps identified in section 5.1.1.2 on OTT SAND deployments, it may be desirable to enable mechanisms that expose real-time network information from the operator network to an OTT DANE. As one potential solution to realise this goal, ETSI MEC APIs 009 - 011 can be considered. An OTT DANE may then use the real-time network information towards deriving the relevant SAND PER messages, as defined in 3GPP TS 26.247 [15].

5.1.1.5 API Aspects

Section 7 in the study documents a few use cases relevant for media distribution in the context of 5G that are relevant for Device API aspects. 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
- ISO BMFF API

Any work on MIME parameters extensions may be considered if progressed.

5.1.1.6 Conclusions from 5G Media Distribution Study

The following list summarises the 5G aspects and functions that are relevant for media distribution and discusses the missing enablers to offer full usability:

- AF: Application Functions (i.e. XCF and XUF) 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. will be realised as AFs. Network assistance, such as SAND and congestion marking, are also offered by AFs in collaboration with other 5G functions such as the PCF. It is however not clear how the AF is instantiated and controlled by the content provider in a secure way.
- NRF: Some of the media distribution functions require that the traffic is routed through the AF, which can for instance be achieved through appropriate DNS resolution. To enable this, appropriate DNS configuration is required. It should be possible for the content provider or the AF to influence the DNS resolution and traffic routing to process data locally.
- QoS Management: the 5G System offers a set of standardised 5QI profiles and allows for non-standardised 5QI definition. QoS profiles can be matched to traffic and special treatment will then be applied to the traffic. Media distribution services should be able to define and apply QoS profiles to their traffic based on agreements with the operator.
- NEF: media distribution services over 5G systems may benefit from plenty of new functionalities offered by the 5G system. However, this requires that the service provider is able to influence the way its traffic is handled. Appropriate NEF APIs are required to enable the service provider to adjust DNS resolution, traffic handling, traffic processing, QoS, etc.

In addition, 3GPP has an ambition to define a single end to end architecture that can support fixed and mobile access networks in Phase 2 of their 5G specifications [16]. Presently, the 3GPP System Architecture (SA) group has approved two work items that can help to achieve this goal:

- Study on the Wireless and Wireline Convergence (see section 5.1.2)
- Study on Access Traffic Steering, Switching and Splitting (see section 5.1.3)

5.1.2 Study on the Wireless and Wireline Convergence

The (3GPP Release 16) study on the Wireless and Wireline Convergence for the 5G system architecture [17], is looking to enhance the 5G core network specification to support wireline access networks and trusted non-3GPP access networks. It will include an investigation on whether existing 5G core network interfaces need to be modified to support wireline devices via wireline access networks. It will also consider aspects such as the impact on security, authentication, policy, Quality of Service, network slicing, mobility and session management. This study will take into account information provided by the Broadband Forum on wireless and wireline convergence (see section 5.2 about ongoing Broadband Forum work).

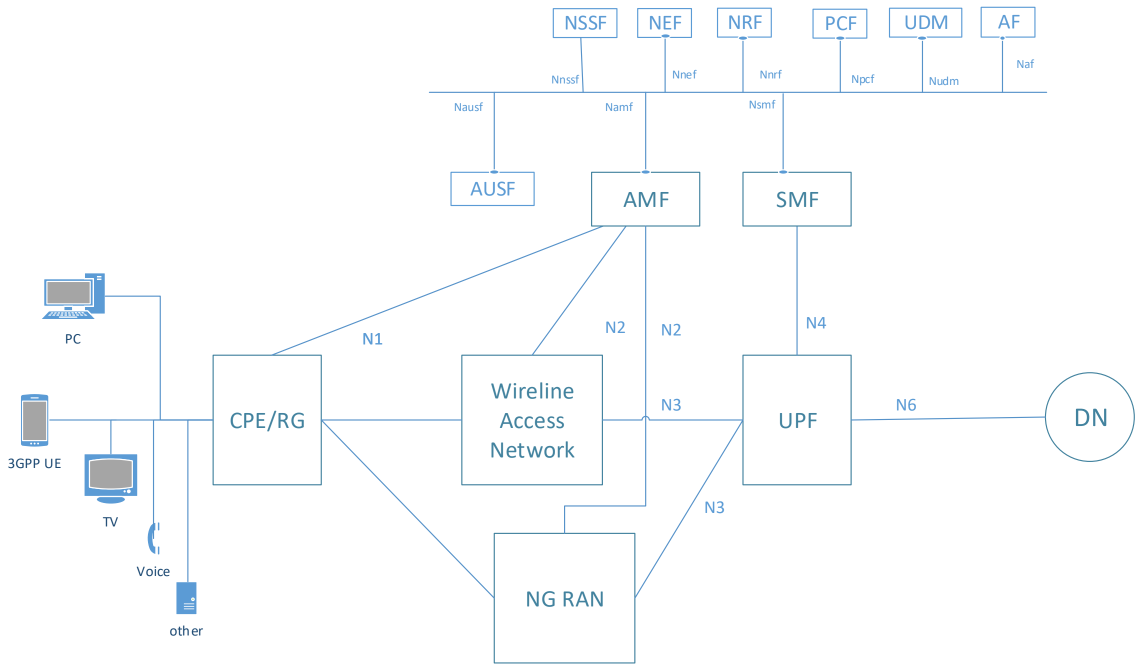


Figure 6 – The architecture for 5GS supporting Hybrid Access and devices connected behind CPE/RG not supporting N1

As shown in Figure 6, both the NG RAN and Wireline Access Network have a connection to the AMF, the UPF and the Residential Gateway. The Residential Gateway has N1 connectivity to the AMF and the devices in the home environment do not support N1.

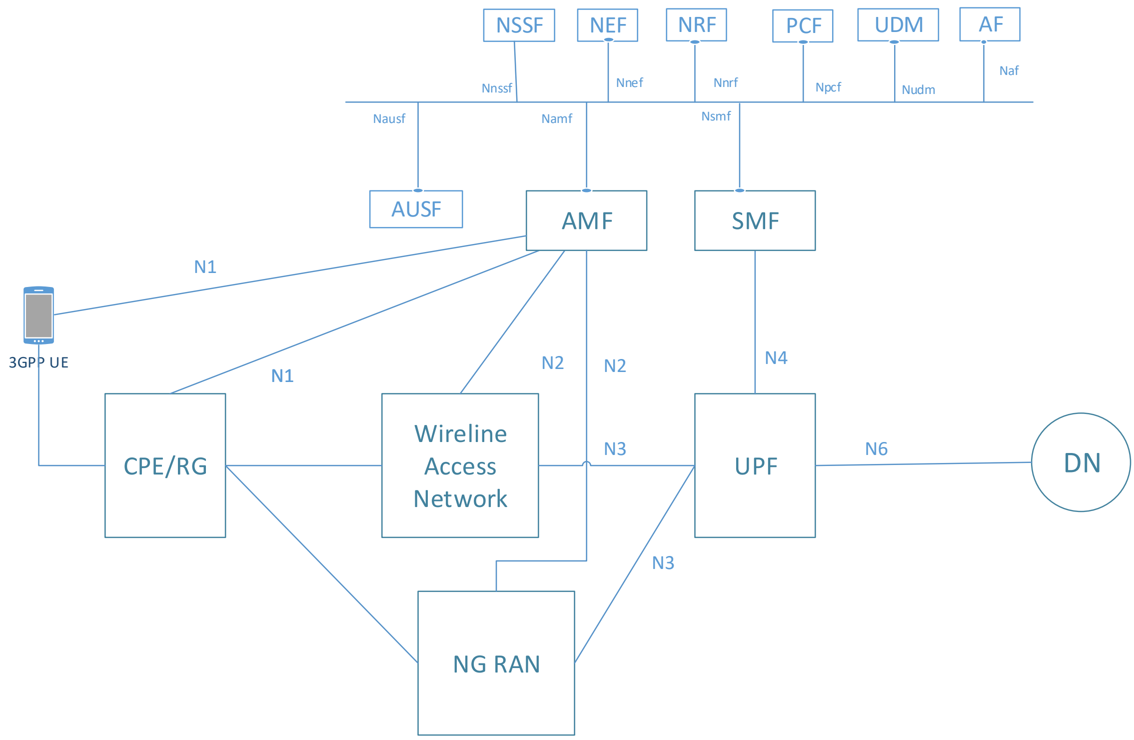


Figure 7 – The architecture for 5GS supporting Hybrid Access and devices connected behind CPE/RG supporting N1

As shown Figure 7, both the NG RAN and Wireline Access Network have a connection to the AMF, the UPF and the Residential Gateway. The Residential Gateway has N1 connectivity to the AMF and the devices in the home environment support N1.

One key capability required to deliver a seamless experience to the end user is the automated discovery and selection of the most appropriate access networks for a UE to connect at any given time. Existing and upcoming 3GPP mechanisms allow for access discovery and selection of 3GPP Radio Access Technologies (RAT). Additionally, as explained in section 3.2, ANDSF mechanisms support policy-driven static discovery and selection across 3GPP and non-3GPP access, with an option to include RAN-provided information in the decision. The authors of this paper expect 3GPP to review existing mechanisms and extend them to become more granular and dynamic (e.g. dependent on real-time network conditions).

Note: Since the start of the 5G-Xcast project work has progressed in further in Release 16 and the study now includes implementation details for IPTV support. The solution describes how to support the IPTV services without impacting the IPTV platform, including IPTV Authentication, IP allocation from IPTV network and the support of Unicast Packets transmission and Multicast Packets transmission without MBS support in 5GC and NG RAN.

The 3GPP solution is based on establishment of a PDU session dedicated to an IPTV service where the IP address is assigned by the IPTV platform. The UPF performs the control of user to the multicast group via IGMP, packet replication and enforces the Channel Access Control of Multicast Channels.

The solution provides two alternatives of Channel Access Control list Provision Procedure.

- Alternative one: the Channel Access Control list is provided by IPTV SMS (Service Management System) in the IPTV network to the UDM in order to be included in the user subscription data via the Channel Access Control list Provision Procedure.
- Alternative two: the Channel Access Control list is provided by IPTV SMS (Service Management System) in the IPTV network to the PCF to be included in 3GPP policy via the Channel Access Control list Provision Procedure.

5.1.3 Study on Access Traffic Steering, Switching and Splitting (ATSSS)

This section is devoted to the analysis of ITU requirements and 3GPP documents. Analysing ITU-T Y.3130 “Requirements of IMT-2020 fixed mobile convergence” we can see, that requirements of session management for fixed and mobile convergence include:

- An IMT-2020 FMC network is required to support traffic switching, splitting and steering between fixed access networks and mobile access networks on the network side.
- An IMT-2020 FMC network is required to support traffic switching, splitting and steering on the user equipment side.

These ITU requirements are mapped in 3GPP TR 23.793. Study on Access Traffic Steering, Switching and Splitting support in the 5G system architecture is described in [18].

All solutions considered within TR 23.793 are based on and aligned with the 3GPP 5GS Phase-1 normative work including work on policy management, as mentioned above documents in TS 23.501, TS 23.502 and TS 23.503.
In particular, the document TR 23.793 considers solutions that specify the following:

- How the 5GC and the 5G UE can support multi-access traffic steering between 3GPP and non-3GPP accesses.
- How the 5G Core network and the 5G UE can support multi-access traffic switching between 3GPP and non-3GPP accesses.
- How the 5G Core network and the 5G UE can support multi-access traffic splitting.

The scope of TR 23.793 excludes the following aspects:

- Changes to the charging framework are not considered. However, it may be considered what information needs to be provided to the charging framework in order to charge traffic that is switched and/or split between 3GPP and non-3GPP accesses.
- ATSSS procedures that may be applied in the NG-RAN are not considered. The study is restricted only to ATSSS procedures applied in the 5G core network.
- 5GS enhancements to support trusted non-3GPP access networks are not considered.
- 5GS enhancements to support wireline access networks are not considered.

The study in this document is organised into two phases.

- Firstly ATSSS solutions enabling traffic selection, switching and splitting between NG-RAN and untrusted non-3GPP access networks will be considered.
- Afterwards, once the 5GS architecture will be further enhanced to also support trusted non-3GPP access networks the study will also consider ATSSS solutions enabling traffic selection, switching and splitting between NG-RAN and trusted non-3GPP access networks.

5.1.3.1 ATSSS solutions

There are three main solutions for ATSSS, which are under the study in 3GPP.

5.1.3.1.1 Solution 1: Proposed architecture framework for ATSSS

In terms of architecture requirements for this solution, the proposed ATSSS architecture framework:

1. Shall support untrusted non-3GPP access, NG-RAN and E-UTRAN. GERAN and UTRAN are not supported. As part of untrusted non-3GPP access, WLAN access shall be supported.
2. Shall support simultaneous connections of an UE via multiple access technologies.
3. Shall support separation of control and user plane functions.
4. Shall support IP (IPv4 & IPv6) and Ethernet PDU session types

5. Shall reuse and align as much as possible with the Release 15 architecture, reference points and functional capabilities in 5GS as stated in TS 23.501 and the procedures as stated in TS 23.502.
6. Shall support roaming. As part of roaming, the architecture shall support both local breakout and home routing scenarios.
7. May support traffic usage reports from UE and ATSSS user plane entity to the ATSSS control plane of the core network to enable ATSSS control plane to get an end-to-end view of traffic performance over multiple access networks in order to dynamically manage ATSSS operations over these accesses.

Figure 8 depicts the initial high level ATSSS architecture for the non-roaming case.

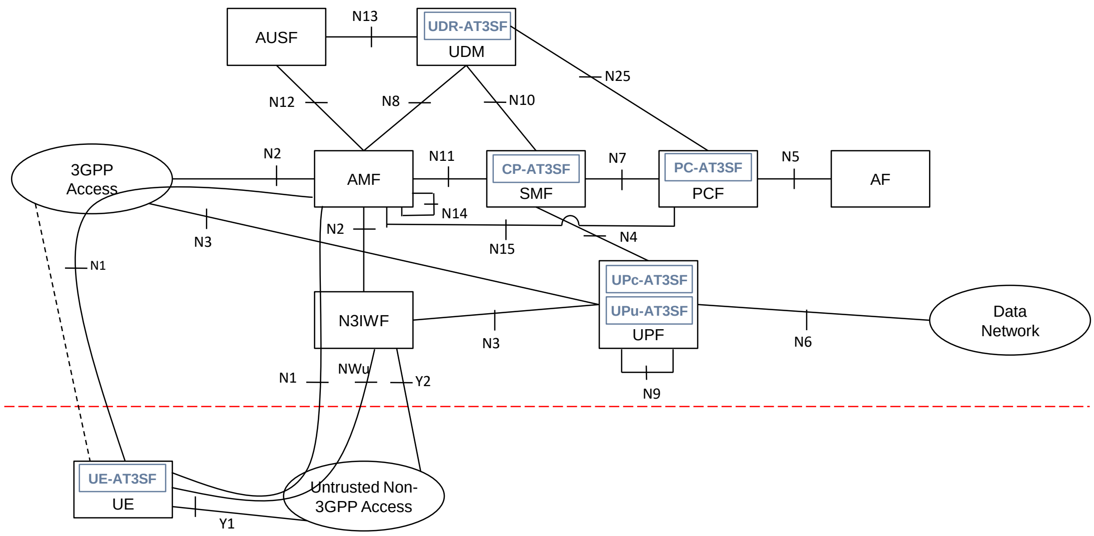


Figure 8 – Initial high level view of non-roaming ATSSS architecture

Figure 9 depicts the ATSSS roaming architecture in the case of LBO with N3IWF in the same PLMN as 3GPP access (assumed AF resides in VPLMN).

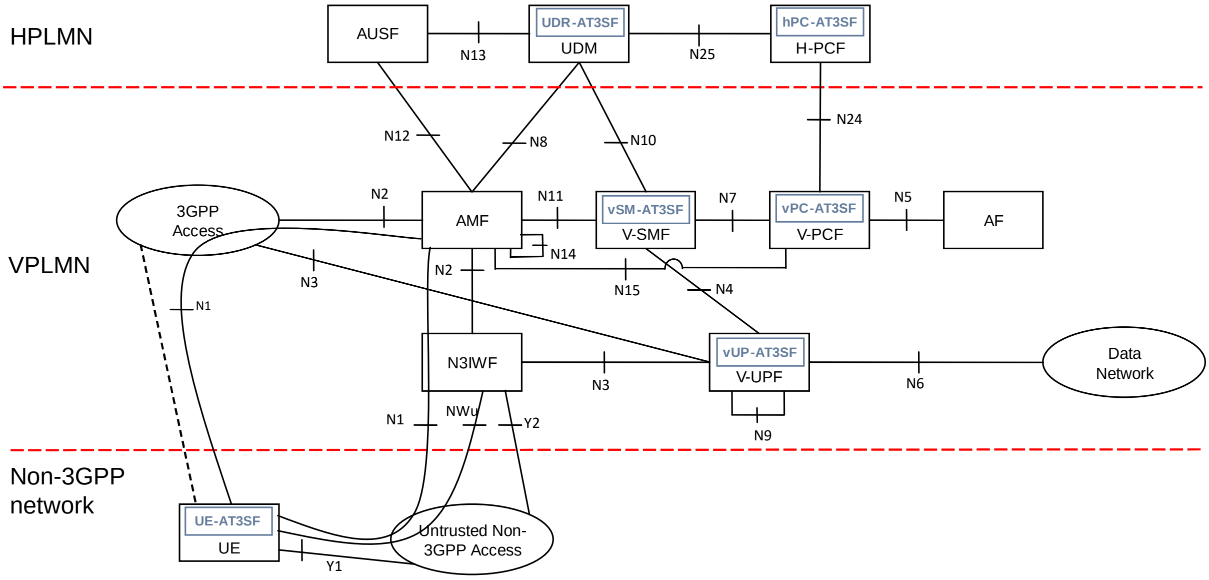


Figure 9 – Roaming ATSSS architecture – LBO scenario with N3IWF in same the PLMN as 3GPP access (assuming AF resides in VPLMN)

Figure 10 depicts the ATSSS roaming architecture in the case of HR with N3IWF in the same PLMN as 3GPP access.

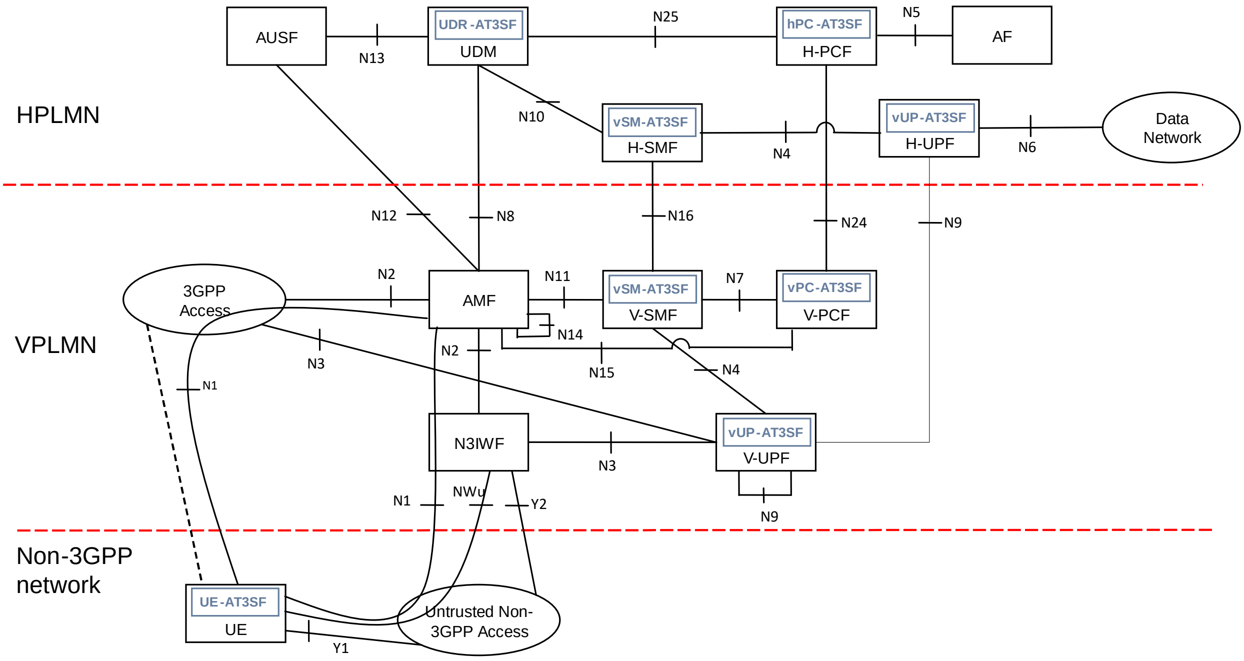


Figure 10 – Roaming ATSSS architecture - HR scenario with N3IWF in same the PLMN as 3GPP access

The following functional elements of the architecture shown in Figure 10 are described in TS 23.501. Procedures are for further study and the use of the access metrics at the UPc-AT3SF is also for further study.

5.1.3.1.2 Solution 2: Support of Multi-Access PDU Sessions

A Multi-Access PDU (MA-PDU) session is created by bundling together two separate PDU sessions, which are established over different accesses.

An MA-PDU session is schematically illustrated in Figure 11. It is composed of two PDU sessions, referred to as "child PDU sessions"; one established over 3GPP access and the other established over untrusted non-3GPP access (e.g. a WLAN AN).

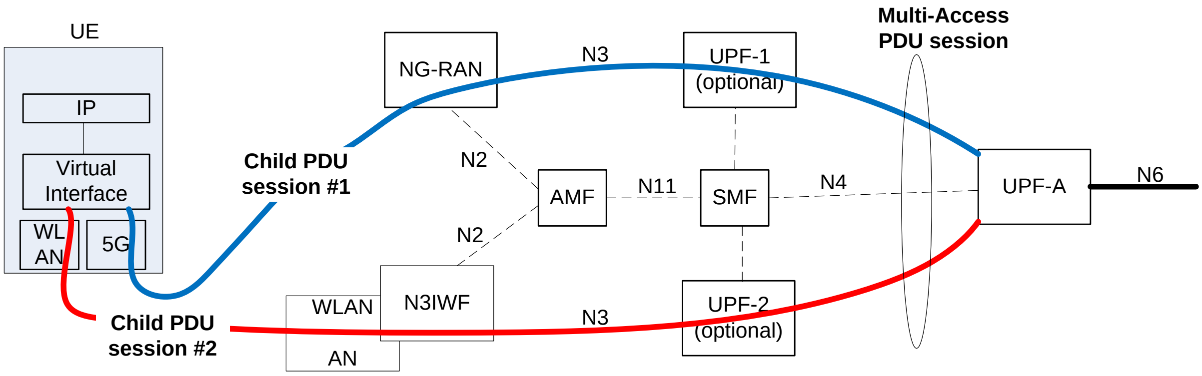


Figure 11 – Illustration of a Multi-Access PDU session with two child PDU sessions

An MA-PDU session realises a multi-path data link between the UE and an UPF-A, as shown in Figure 11. It operates below the IP layer and it is transparent to the IP layer and all layers above the IP layer.

An MA-PDU session can be established with one of the following procedures:

(i) Established with two separate PDU session establishment procedures; one of each child PDU session. This is called "separate establishment".

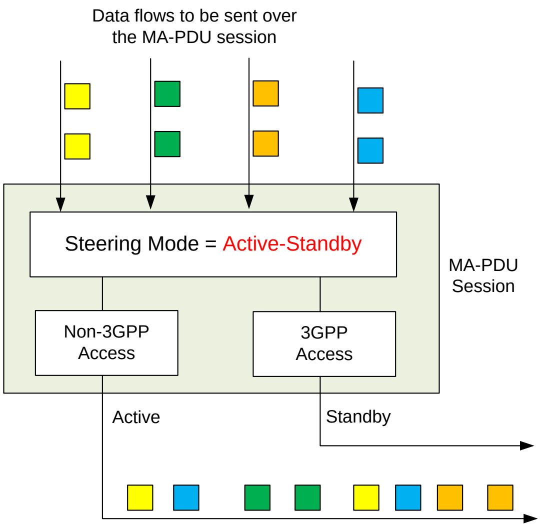
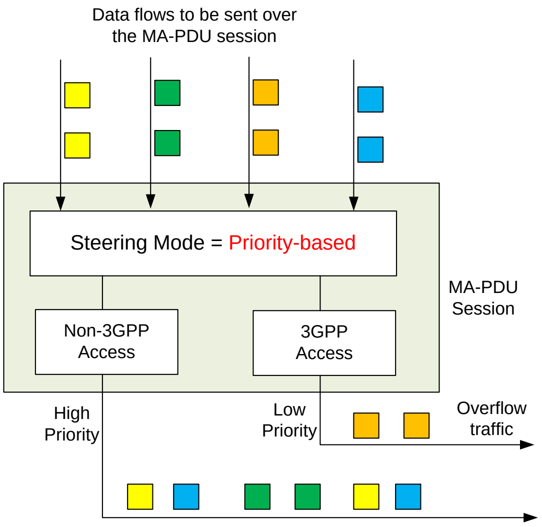
(ii) Established with a single MA-PDU session establishment procedure, where the two child PDU sessions are established in parallel. This is called "combined establishment".

The child PDU sessions established with any of the above procedures have the same IP address.

After a MA-PDU session is established, SM signalling related to this MA-PDU session (e.g. for modification or removal of a child PDU session, etc.) can be conducted over any of the available accesses of the MA-PDU session.

It is expected that the UE discovers the support of ATSSS for a PDU session upon attempting to establish the MA-PDU session. Additional changes to the existing "UE Requested PDU session establishment" procedure are for further study.

It has to be noticed that the advantages of the combined establishment versus the separate establishment should be further analysed, considering also failure scenarios and further details of the "UE Requested MA-PDU session establishment" procedure (including PCC, QoS, handover and user-plane aspects) are for further study.

	Steering Mode	Comments
#1		**Active-Standby steering:** All (or some of) the traffic of the MA-PDU session is sent to one access only, which is called the "Active" access. The other access serves as a "standby" access and takes traffic only when the active access becomes unavailable. When the active access becomes available again, the traffic is transferred to the active access. The active access can be defined when the MA-PDU session is established and can either (a) remain the same during the lifetime of the MA-PDU session or (b) can change during the lifetime of the MA-PDU session. - Benefit: The MA-PDU session can provide enhanced continuity. Traffic can be switched from the active access to the standby access and vice versa with minimum or no signalling. - Complexity: Small.
#2		**Priority-based steering**: The two accesses are assigned a priority, e.g. during the establishment of the MA-PDU session. All traffic (or some) of the MA-PDU session is sent to the high priority access. When congestion arises on the high priority access, new data flows (the "overflow" traffic) are sent to the low priority access. In addition, when the high priority access becomes unavailable, all traffic is switched to the low priority access. If needed, it could be possible to change the priorities of the accesses during the lifetime of the MA-PDU session. - Benefit: The MA-PDU session can provide enhanced continuity and increased bandwidth (because both accesses can be used simultaneously). - Complexity: Requires the UE and/or the network to determine when congestion arises on the high priority access. This may be a complex task.

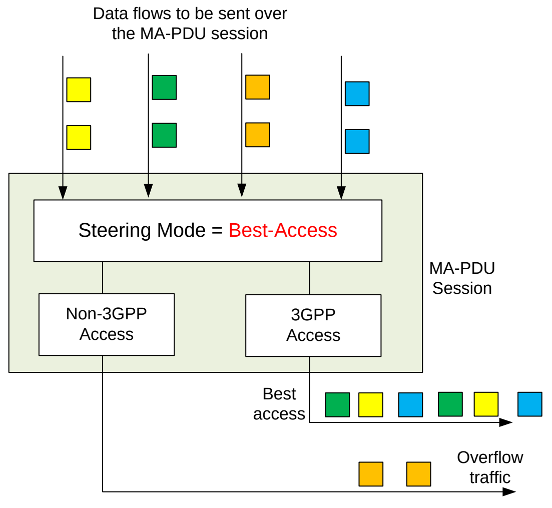
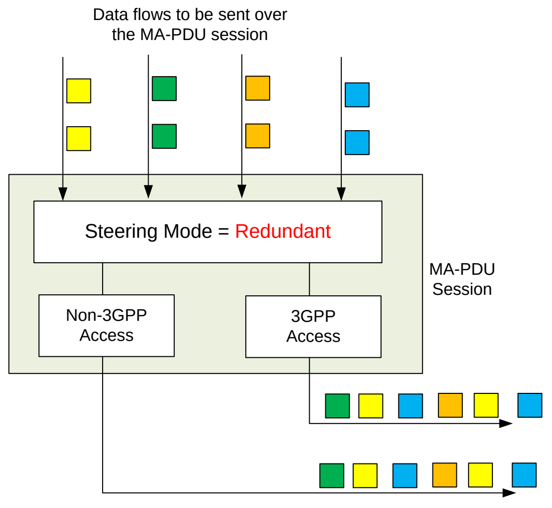
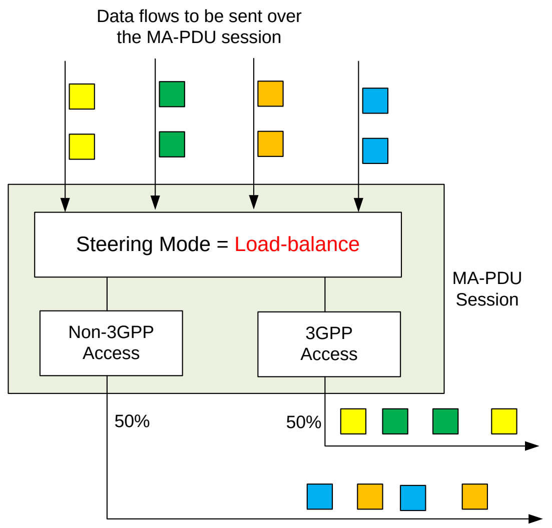
#3		**Best-Access steering**: The same as the Priority-based steering but with the following difference: The high priority access is the one that can provide the best performance, e.g. the one with the smallest RTT. In this case, the high priority access is not pre-defined (as in Priority-based steering) but it is estimated in real-time and can change dynamically. - Benefit: The MA-PDU session can provide enhanced continuity, increased bandwidth and better performance (compared to Priority-based). - Complexity: Requires the UE and/or the network to estimate the best access and to determine when congestion arises on the best access. This may be a complex task. NOTE: The best access could be estimated by the network by using Access Measurements as those considered in Solution 1 (see TR 23.793, clause 6.1.2.1) or by using a Path Performance measurement function as the one considered in Solution 3 (see TR 23.793, clause 6.1.2.1).
#4		Redundant steering: All (or some) data flows are transmitted on both accesses in order to increase reliability. - Benefit: The MA-PDU session can provide very high data reliability (i.e. very small packet error rate) and session continuity. - Complexity: High - The receiving side should be able to detect and discard duplicate packets. For this purpose, the sending side could append sequence numbers to the transmitted packets. In addition, to provide in-sequence delivery, the receiving side should be able to buffer and re-order the received packets.
#5		**Load-balance steering**: Each access receives a percentage of the data flows transmitted via the MA-PDU session. Each access is assigned a weight factor (e.g. 50%) and receives a percentage of the MA-PDU session traffic corresponding to this factor. In a 50/50 load-balancing, the overall traffic of the MA-PDU session is equally split across the two accesses. In an 80/20 load-balancing, about 80% of the overall traffic is sent on one access and 20% on the other access. - Benefits: The MA-PDU session can provide bandwidth aggregation with a certain load balancing ratio. - Complexity: Relatively small - There is no need to assess the transmission performance on every access.

Figure 12 – Steering modes for traffic distribution in a MA-PDU session

5.1.3.1.3 Solution 3: TFCP (Traffic Flow Control Protocol) based architecture framework for ATSSS

In terms of the architecture requirements for this solution, the proposed ATSSS architecture framework is shown below:

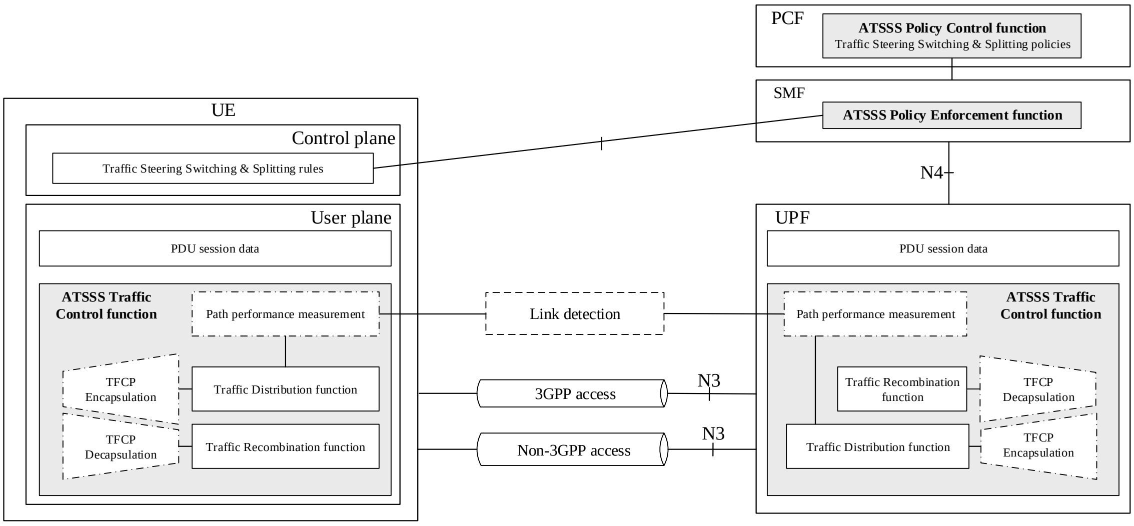


Figure 13 – ATSSS architecture

The ATSSS Policy Control function in PCF defines the following policies according to the application-specific information, the UE subscription data, user preference, local policy or any combination of them.

If any of the above example aspects needs to be standardised, it is for further study and whether path performance measurement impacts the traffic distribution are for further study too.

In the 5G-Xcast core architecture ATSSS is implemented by a multilink solution. Procedures for ATSSS with multilink are described in D.4.3. Multilink almost works in accordance with ATSSS architecture (Solution 3).

The main steps in the multilink ATSSS at the IP layer are:

1. Evaluate in real time the changes in application-level performance of each link (e.g. “goodput”, latency, jittery behaviour) in each of the relevant directions (e.g. uplink, downlink).
2. Evaluate the total available “goodput” at each point in time.
3. Based on the evaluation and content delivery strategy, one of the following alternatives will be provided:
   a) To select an access network for a new data flow and send the traffic of this data flow over the selected access network.
   b) To move all traffic of an ongoing data flow from one access network to another access network in a way that maintains the continuity of the data flow.
   c) To split the ongoing stream of content-to-be-delivered to all available links according to the performance of each link (i.e. not “overload” any

of these links). Note that there are scenarios where some content may not be delivered to all terminals.
4. Buffer is necessary at the UE side to accommodate out-of-order packet arrival, missing packets etc.
5. If content was split, to combine it from multiple links into one common stream on the UE's side.
6. In some cases (e.g. live video), it is possible to add an integrated video encoding process which outputs a live encoded or just-in-time transcoded video stream that adaptively matches the momentary performance of the multiplicity of virtually bonded links. However, this is not one of the primary scenarios in the 5G-Xcast project.

Multi-link ATSSS is currently considered only for unicast streams, bringing together distinct unicast connections to support a stream, but it is suitable for multicast and broadcast in some use cases, especially for object-based broadcasting, seamless transitions between different coverage areas etc. The performance of each available link is measured between NR and Multilink middleware by exchanging the information. Therefore, simpler (in the sense that it has no feedback measurement mechanism) multicast or broadcasts over a common set of links lacking measurement information will not work in many cases.

5.2 Ongoing Work in Broadband Forum

In parallel to the work in 3GPP and in liaison with it, the Broadband Forum (BBF) is also carrying out a study on 5G Fixed Mobile Convergence in the Wireless and Wireline Convergence Work Area in order to explore how a Wireline Access Network, defined by the BBF can be integrated into a 5G core, defined by 3GPP. One key objective is to provide the 3GPP with the necessary requirements and architectural enhancements that need to be incorporated into their architecture.

To that purpose, the BBF has so far defined two approaches to support fixed network access into the 5G architecture:

- ***Integration*** – the 5G core network delivers all functions required to support a wireline access network, including user profile information, authentication and session management. This option requires modification to the Residential Gateway as well as some adaptation of the Wireline Access Network in order to support integration with the 5G core. The adaptation of the Wireline Access Network is expected to be achieved by either the deployment of a 5G Access Gateway Function (AGF) alongside existing Wireline Access Node or by full integration of the 5G AGF within existing Access Nodes. How this would be achieved is currently under discussion.

- ***Interworking*** – the Wireline Core Network continues to provide most of the existing functions (e.g. subscriber management or session management) but enables some convergence with the 5G core through the Fixed Mobile Interworking Function (FMIF). This solution provides an interim option for broadband network operators not wanting to migrate all of their subscribers to the 5G core, or modify their existing Residential Gateways and/or Wireline Access Network nodes, whilst allowing them to deliver some convergence capabilities.

As mentioned previously, this work item is still being progressed in Wireless and Wireline Convergence Work Area in the BBF and the output will be fed into 3GPP standards to be incorporated into phase 2 of their 5G technical specifications.

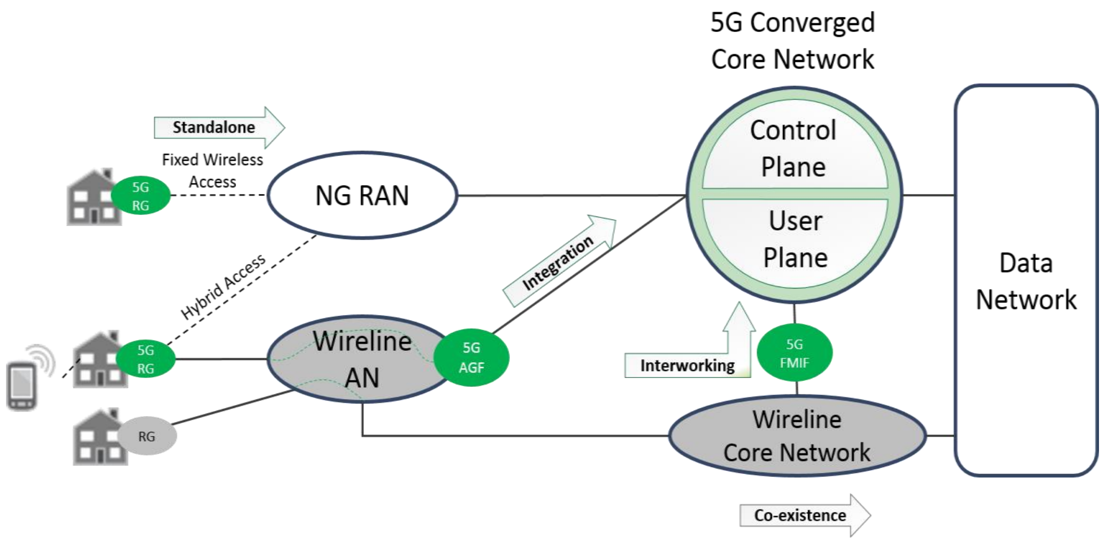


Figure 14 – Broadband Forum 5G Fixed and Mobile Convergence Study [2]

To explore the BBF converged architecture in more detail, a recent liaison statement was shared from BBF to 3GPP outlining their proposed architecture for convergence:

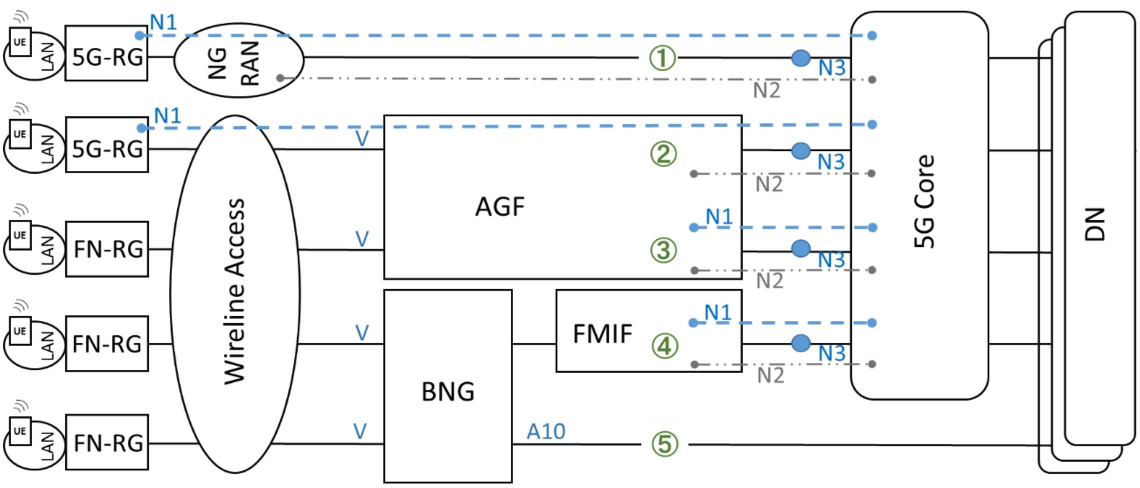


Figure 15 – Broadband Forum Convergence Architecture

Architectural Options:

[2] Figure reproduced with permission from the Broadband Forum.

(1) Fixed Wireless Access (5G-RG) – The 5G-RG is connected over the NG-RAN.

(2) Integration in Direct Mode (5G-RG) – The RG is connected over the wireline access network. An Access Gateway Function (AGF) mediates between the wireline access network (aggregated at layer 2) and the 5G core network, based on N2 and N3 interfaces. The 5G-RG is able to register directly with the 5G core network based on N1 interface. For this reason, the AGF is said to integrate the session in "direct mode".

(3) Integration in Adaptive Mode (FN-RG) – Similar to (2), the RG is connected over the wireline access network and the AGF mediates layer 2 traffic with the 5G core network based on N2 and N3 interfaces. However, FN-RG does not support N1, so the AGF acts as end point of N1 on behalf of the FN-RG. The AGF is said to integrate the session in "adaptive mode".

(4) Interworking (FN-RG) – The session is managed by a BNG. Services that are based on 5G core network are passed to the 5G core via a Fixed Mobile Interworking Function (FMIF). The FMIF supports N2 and N3 interfaces to the 5G core network. Since the FN-RG does not support N1, the FMIF acts as end point of N1 on behalf of the FN-RG.

(5) Coexistence (FN-RG) - This is not a converged session model, as these sessions are not part of the 5G core network. However, coexistence is required to allow services that are not supported by the 5G core network to be available in a converged service provider network. Coexisting subscriber sessions are managed only by the BNG. When the wireline access network is shared between converged sessions (see models (2), (3), (4)) and coexisting sessions, the AGF / FMIF may need to be aware of the coexisting sessions, in order to offer an accurate traffic management in the wireline access network.

5.3 Network Slicing

Addressing the requirements of 5G services (e.g. Ultra-Low Latency, Ultra High Bandwidth, Massive IOT, etc.) on a single converged network presents a huge challenge for operators. Network slicing enables operators to create multiple virtual networks dedicated to different services/service types in order to address those requirements.

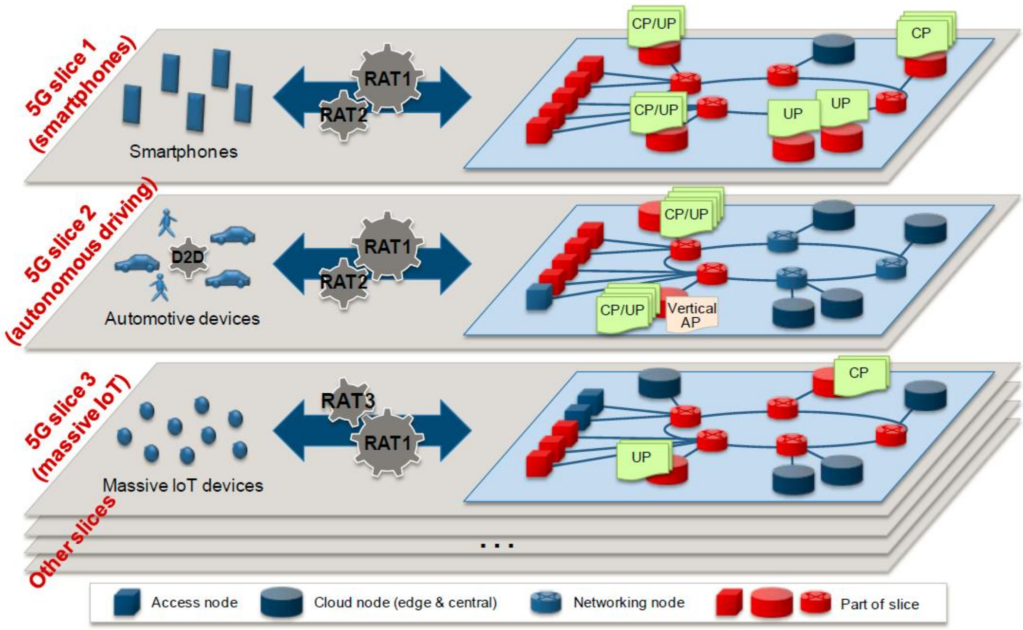


Figure 16 – 5G network slices implemented on the same infrastructure [1]

There are a number of key drivers for Network Slicing:

- ***Rapid deployment of new services***
  - Need to deploy new services with no disruption to existing ones.
  - Agility is required to compete and meet market demand.
  - NFV and SDN enable orchestration across slices.
  - End goal is fully automated deployment of new network slices.
- ***Support for different operational models***
  - Different SLAs (e.g. security, reliability etc.) may require isolation between different slices.
  - Slicing can be done on a per service-type or even for individual customers.
- ***Conflicting functional requirements***
  - Some functional requirements may be mutually exclusive, e.g. high data throughput versus low latency or highly-mobile versus fixed access.
  - Optimisation of each slice for the specific functionality required (e.g. slice supporting UEs with no or low mobility).
  - There may be alternative approaches to meet this particular goal (e.g. flexible anchor points, early detection of mobile devices etc.)

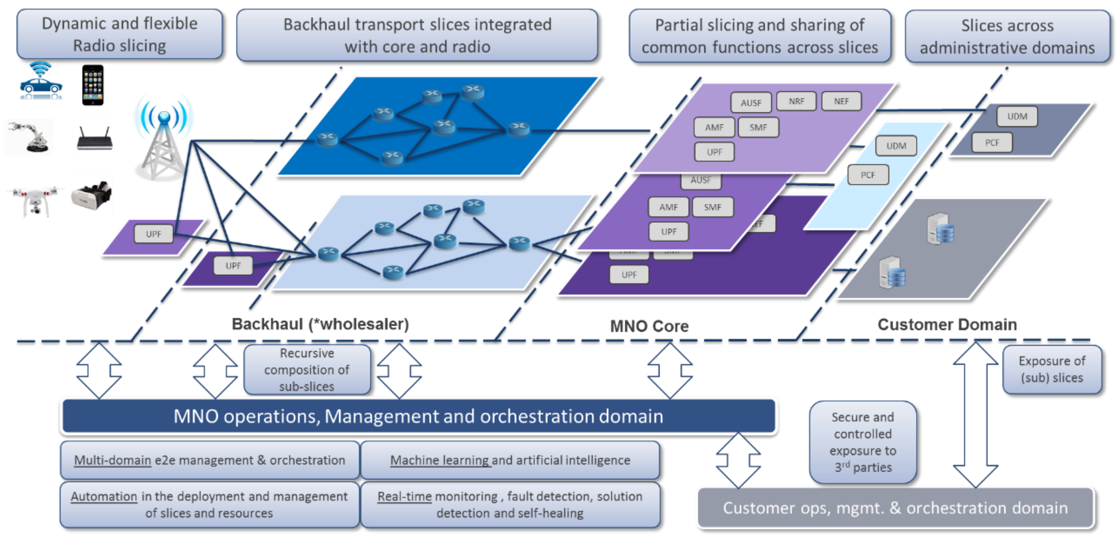


Figure 17 – Complexity of End to End Network Slicing

As shown in Figure 17, end to end network slicing may incorporate the following:

- Dynamic and flexible radio slicing.
- Backhaul transport slices integrated with core and radio.
- Partial slicing and sharing of common functions across slices in the core network.
- Slices across different administrative domains (e.g. for roaming).

This presents a number of challenges for management and orchestration in regard to end-to-end slicing:

- Multi-domain end-to-end management and orchestration.
- Automation in the deployment and management of slices and resources.
- Machine learning and artificial intelligence.
- Real-time monitoring, fault detection, solution detection and self-healing.

- Interworking with customer operations, management and orchestration domains.
- Secure and controlled exposure to 3rd parties.

This end picture is very complex and further work is required in industry in order to solve some key architectural issues.

5.3.1 Network Slicing in 3GPP

5.3.1.1 3GPP Slicing Summary

Network slicing is going to be supported by the first release of 3GPP 5G system (i.e. Rel-15). The main principles of network slicing in 3GPP can be summarised as follows. A network slice comprises of core network control and user plane functions and an access network such as the next generation radio access network. Network slices may be used to offer network function optimizations and support of features required by a specific use case. A network slice is identified by Single Network Slice Selection Assistance information (S-NSSAI). S-NSSAI consists of Slice/Service Type (SST) and Slice Differentiator (SD). There are standardised values of S-NSSAI for eMMB, URLLC and IoT services. These standardised values do not use SD. For each network slice, an operator can deploy multiple network slice instances delivering the same services to different groups of UEs. And each UE can be served by up to eight network slice instances. The AMF is the common network function to the network slice instances serving the UE. A PDU session, which abstracts an association between UE and a data network that provides a connectivity service to the UE [5], belongs to only one specific network slice instance in the network. This means that different network slice instances do not share a PDU session, but they share an AMF as described above. The set of network slices and network slice instances is initially selected during the UE's registration with the network and the set can be modified while UE is registered.

The first questions related to 5G-Xcast and network slicing in the 3GPP solution of converged network is whether multicast and/or broadcast is provided as a service via a dedicated network slice or whether multicast and/or broadcast are offered as part of PDU connectivity service together with unicast. The answer to this question may differ from use case to use case. We discuss network slicing further in the following sections in the context of the 5G-Xcast uses cases: media and entertainment, public warning, automotive and IoT.

5.3.1.2 Media and Entertainment

It is understood that a unicast transport is always available in all media and entertainment use cases while a multicast transport is used as an optimisation for efficient system operation. The unicast transport is enabled by the connectivity service of PDU session. More specifically, if we assume that the connectivity is provided using a PDU session belonging to an eMBB network slice instance, then the multicast transport can be provided within the same network slice instance or via a different network slice instance dedicated only to the multicast transport.

To use a dedicated network slice instance for the multicast transport, the UE would be required to request a registration with the network for two network slices using at least two S-NSSAI. A common scenario would be that the network slice instances would connect to the same data network. While the approach of using the dedicated network slice instance to the multicast transport would provide the separation of unicast and multicast resources, it is not clear what the benefits are considering that this approach requires two network slice instances per data network to provide the unicast and multicast transport capabilities, or that it would even make some services impossible. It should be noted that at least in the first release of 3GPP 5G system (i.e. Release 15) the number of network slice instances to which the UE can be connected concurrently is limited to eight. This approach would effectively limit the maximum number of (local) data networks to which UE can be connected with unicast and multicast transport support to four.

The alternative is to enable the multicast transport within the eMBB network slice instance and for the PDU session which UE already uses for the unicast transport. This means that the network could modify the network slice instance based on the need of multicast transport. The network slice modification would be transparent to UE as the UE remains connected to the same network slice instance. If the multicast resources are managed using the multicast context concept [5], then the multicast network resources (e.g. network functions) within the network slice instance can be shared among PDU sessions interested in the same multicast content [19]. For completeness, we should also consider the situation when the network is operated so that different network slice instances to the same data network at one location (tracking area) are used to serve different groups of UEs. The network operator can deliver a different committed service deploying the network slices this way (e.g. the network slices use the same S-NSSAI but a different SD) [11]. This deployment could lead to a network configuration when the same data are delivered to the access network via the different network slices. Because the different network slices are going to be commonly used for differentiation, it could be also assumed that the access network treats the traffic in each of these network slice instances differently applying QoS differentiation and specific RRM policies [20] even though the data are the same. The network slicing principle of 3GPP, which says that a PDU session belongs to only one network slice instance (i.e., the resources associated with PDU session are not shared between the network slice instances), also applies in this scenario. We can conclude that enabling the multicast transport within the same network slice instance along with the default unicast transport represents a better solution in comparison to the use of a dedicated network slice for the multicast transport. The conclusion is in line with [21] that presents multicast as an example of additional system features of network slice instance. Unicast and multicast resources can still be separated within a network slice instance to some extend by adopting the concept of network slice subnets. A network slice subnet instance represents a group of network function instances that form part or complete constituents of a network slice instance [19]. The network slice subnet can represent a group of network function instances providing multicast transport functionality in a network slice instance.

The discussion focused on the provisioning of multicast so far. The architectures and the procedures designed by the 5G-Xcast project support also point-to-multipoint services. The point-to-multipoint services offer the possibility to broadcast content in a geographical area (see [5]) possibly to receive-only UEs. The support of receive-only UEs seems to be the key requirement impacting the system design – not only from the network slicing perspective. In the context of receive-only UEs, the point-to-multipoint services are unidirectional services that are not associated with any bidirectional services. An example of such service is terrestrial broadcast.

The point-to-multipoint service provisioning with a dedicated network slice instance(s) separated from other services represents a very attractive solution from the content service provider and network operator perspective. In the 5G-Xcast architectures, when the XCF is a part of a network slice subnet instance, that network slice subnet instance may or may not be shared by more than one network slice instance. In both cases, the XCF could store network slice information (NSSAI) for content service provider. Network slice information could be specific to xMB service or xMB session. The XCF could use the stored network slice information for selecting the network resources of requested network slice as described in [5]. Alternatively, xMB interface can be enhanced to include network slice information provisioning by content service provider.

When it comes to NG RAN, few facts need to be considered. At the time of writing this document, the agreements in 3GPP are that NG RAN is not going to provide any information about supported network slices to idle UEs and that a UE may provide network slice assistance information (list of S-NSSAIs) to NG RAN during RRC connection establishment. This means that an idle UE, which is equivalent to receive-only UE in this context, does not have any information about the availability of network slice by a cell, which could be used to determine if it can receive point-to-multipoint services provisioned by a network slice instance in the cell. Furthermore, if we assume that a network slice instance may be used to provide more than one point-to-multipoint service in PLMN and the point-to-multipoint services may be provided in different, possibly overlapping (geographical) areas then network slicing information is not suitable for informing UEs in RRC idle state and receive-only UEs about point-to-multipoint service availability and other concepts should be considered. For example, an UE could be provisioned with a list of cells and frequencies where a service is available. It may become impractical to use a list of cells if a service is available in large area comprising large number of cells. In such case, the list can be given an alias that then needs to be broadcast by the cells.

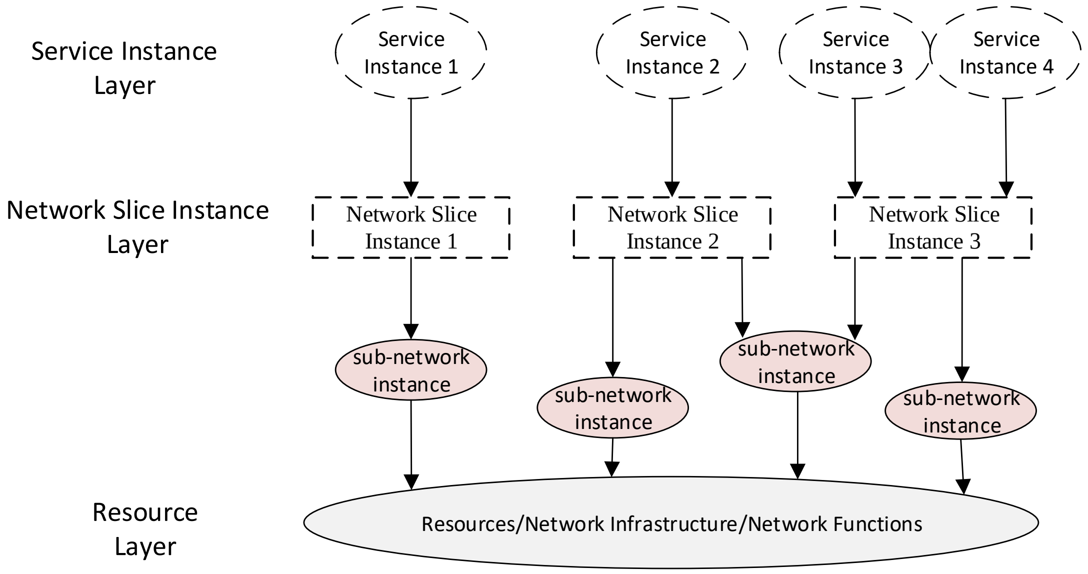


Figure 18 – Network slicing concept with sub-network instances

5.3.1.3 Public Warning Systems

Public warning is offered to the general public whilst aiming for the largest addressable audience. The audience may subscribe to different network slices depending on the customer needs and operator policies.

A dedicated network slice for delivering public warning message might be feasible however for the same reasons as mentioned in 5.3.1.2, no benefits are foreseen, whilst it does appear to complicate matters. To deliver the public warning over a dedicated network slice, the UE will need to be either continuously connected to such slice, or connect when needed. Currently there is no procedure defined for the network to request to UE to connect to a slice, whilst such a procedure may delay receipt of the public warning message. Having the UE continuously connected to a network slice that is only used rarely may be an uneconomical usage of network slices since the amount of concurrent connected slice is limited to eight.

The public warning messages could be delivered using multicast in the eMMB slice or any other relevant network slice serving human subscribers or devices with human to machine interfaces that can suitably handle the messages. For example, the public warning messages could be delivered to vehicles connected to an automotive network slice and displayed by in-car system. The public warning service could be always available within the network slice or the network slice is modified to make the service become available only when needed. The later maybe beneficial resource-wise but does require the network slice modification to be performed reliably and without undue delays.

Network slices dedicated to serving Automotive and IoT could also be equipped with public warning provided the UE connected to the slices can suitably handle the messages.

In case a UE connected to multiple network slices concurrently, the UE is receiving duplicate public warning messages over the different slices, unique message identifiers should be used to ensure the same warning message is only presented once to the end-user.

5.3.1.4 Automotive and IoT

The automotive use cases are in many aspects similar to the media and entertainment use case. [20] describes a mapping data application in which maps represent large data volumes. The application could benefit from point-to-multipoint delivery in case there are multiple vehicles downloading the same data. We need to consider that provisioning of information from a vehicle to other vehicles and an automotive system most likely residing in a distributed cloud may be equally important in the context of up-to-date mapping. A system can be foreseen that uses advanced sensors such as LIDAR mounted on vehicles to sense the environment and compares the sensory output with the stored mapping data. In case of the stored data do not match with the sensor data, the system sends an update which then triggers other cars to download updated mapping data. From this perspective, the system could operate in a similar matter as described for M&E use case when unicast is always available and multicast is used as an optimization. This leads to the same conclusion regarding network slicing as for the M&E use case that multicast should be provided within the same slice rather than a dedicated network slice from UE perspective.

[20] describes the alerting and signage information use cases, which seems to be purely unidirectional services. Alerts or digital route signs have typically a relation to a geographical area. These use cases can be supported by the 5G-Xcast architecture using the point-to-multipoint services as described in section 5.3.1.2 and the same conclusions apply.

In the massive software, firmware updates and configurations for IoT devices use case, it should be remembered that the core functions of any IoT system is to read values from an IoT device or to control an IoT device sending control information to an IoT device, which require unicast communication. While it could be possible to provide massive software, firmware updates and configuration via a dedicated network slice instance offering multicast/broadcast functionality, it seems not feasible and practical to require an IoT UE to connect to the dedicated network slice instance for the reasons described for the multicast transport in section 5.3.1.2.

5.4 Caching Technologies

5.4.1 Overview

Content caching is a way to optimise content delivery. This is typically done by storing assets in various network places, closer to the end user in order to optimise both QoS and network resources.

Caching can be triggered from various criteria:

- Session counting
- Localised content
- Content popularity
- …

5.4.1.1 Why caching works

Studies found that content delivery follows *Pareto*'s law: That is, in a standard system, 20% of content represents 80% of viewings. Previous figures can vary according to video services offered and users, but the principle remains. Such trends have been noticed in networks ranging from cellular, to user generated content, to IPTV and VoD. As an illustration, for both YouTube and Daum (service in Korea), 10% of the most popular videos account for nearly 80% of views, while the remaining 90% account for total 20% of views as depicted in the figure below [22]:

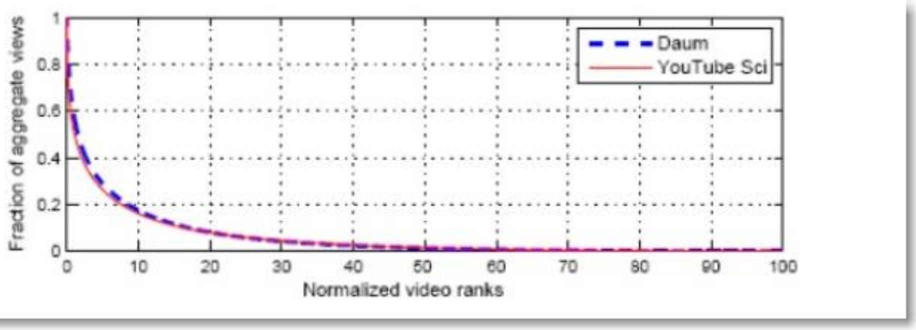


Figure 19 – Media on 5G System Architecture

Content popularity tends to follow a heavy tail distribution which means that the majority of requests occur for a relatively small fraction of the content. Observations of this characteristic have been noted in data logs for various CDNs.

This can be quantified using *Zipf*'s law. Specifically, if we ordered the files from most to least popular at a given point in time, then the relationship governing the frequency at which the file of rank *i* will appear is given as follows [23]:

$$f(i) \propto \left(\frac{1}{i}\right)^{\alpha}$$

<u>The probability of a request occurring for file *i* is inversely proportional to its rank, with a shaping parameter *α*.</u>

A larger *α* implies that more requests occur for a smaller fraction of the content, making the network more amenable to a caching solution. Values of *α* have been cited between 0.5 and 0.8 [23].

We can divide caching in two main algorithms:

- Generic caching based on session counting.
- Predictive caching (prefetching of content before they are requested).

Next sections will briefly present these two algorithms.

5.4.1.2 Caching based on content popularity

This is the most common implementation of caching. When a content is requested on a streamer, this content is automatically fetched from the origin and cached locally so it can be served in the future without using the origin server. This drastically increases the QoS (and QoE) and optimises the network resources.
This however has some drawbacks:

- It has the side effect of decreasing the storage lifetime in the Edge due to the huge number of write cycles implied.
- It increases the storage size on the streamers though dramatically increasing the cost of the edge servers.

To deal with these issues, some additional intelligence may be implemented. The two sections below describe two possibilities.

5.4.1.2.1 Cache tiering

To mitigate the cost increase related to storage, some servers use a “cache tiering” feature. Basically, servers have different types of storage from the less performant and cheapest drives to the high performing and most expensive media. A threshold, internal to the server, is configured to determine how many concurrent streams are required to move content from one media to another.
Typically, three different levels of storage can be managed:

- SAS/SATA hard drives (can be internal, DAS or SAN) (less popular content).
- First cache level: SSD drives (medium popular content).
- Second cache level: RAM (most popular content).

5.4.1.2.2 CDN Intelligent Edge Caching

CDNs implement caching advanced algorithms such as Intelligent Edge Caching. This can be done thanks to their topology knowledge (management of multiple POPs and servers). With this feature, it becomes possible to state that long tail contents are streamed from a central location (servers with huge storage capacity) and that only popular content are cached in edge servers and streamed from there.

Streamers are elected by the CDN based on the configured topology:

- Central server: if the content threshold has not been reached yet, the central server is elected to deliver the session.
- Edge servers and cache affinity: Edge servers will be elected once the threshold has been reached, furthermore if a server has already been elected for a content then future sessions on this content will be redirected to this server in priority.

This way of distributing the sessions circumvents the two issues described in previous section. The trickiest part being the determination of the threshold used to decide when an asset should be streamed from the edge.

5.4.1.3 Predictive caching

For Live channels or content which high popularity level can be foreseen in advance (localised content, popular series' new episodes …), a pre-provisioning might be possible: the caching of content can be forced before it is requested by a customer.

In that case, the streamer is fed with a list containing the high popularity assets. The streamer scans the list and decides to prefetch the top content based on its remaining storage.

5.4.2 Web Caching

Web caching refers to caching requests and responses of HTTP communication. While caching is an entirely optional feature of HTTP/1.1, the caching generally improves the system performance by reducing latency and network traffic, therefore caching is widely used in today's systems. A web cache is located between a client and a web server in various forms which include the following

- Client cache: is a private cache in a client application such as a web browser dedicated to a single user. In case of cache hit, no communication between the client application and the network occurs. This type of cache is also referred as in-device cache.
- Proxy cache: is a shared cache deployed in the communication path between a client and an origin server. A configuration in the client or an interception relying on the routing in underlying network are used to route requests through the proxy caches. The latter approach is also known as transparent network caching. Network operator may deploy multiple proxy caches in more than one location forming a cache hierarchy. The interception cache does not work with transport layer encryption.
- Gateway cache: is a cache that is typically deployed as a part of platform or a web server. Requests can be routed to gateway caches using various methods including, for example, DNS methods for load balancing, which make the gateway caches appear like the origin server to clients.

Requirements on HTTP/1.1 cache are specified in [24]. Although, the HTTP standard defines GET, HEAD and POST methods to be cacheable [25], i.e. a cache is allowed to store the responses to these functions for future reuse, only caching of GET responses is commonly implemented (e.g. see Google Cloud CDN documentation [26]).

A cache-control mechanism of HTTP/1.1 can be used to control what requests and responses may be stored by a cache for future use. The cache-control mechanism uses "Cache-control" HTTP header to specify how a cache shall treat a response. The control information includes parameters such as maximum age, prohibiting response storage (no-cache), prohibiting storage of request and response (no-store), etc. [24].

In the light of current trend of using transport layer encryption, 5G-Xcast solutions should focus on the client cache and gateway cache models.

5.4.3 Web Caching with Edge Cloud

Efficient caching requires caches to be placed in strategic locations so that the cache hit ratio is high. The potential locations are distributed data centers with compute and storage capabilities accessible through virtualisation solutions throughout mobile and converged networks. Multi-access Edge Computing (MEC) platform aims at harmonising the industry in the effort to make the compute and storage resources widely available to virtually anyone. MEC's goal is to provide a standardised platform with well-defined set of interfaces allowing deployment of edge cloud services at scale, which may benefit from information about network information.

Cloud CDN solution offerings from large industry players such as Amazon, Google, Akamai and others already incorporate possibilities to use network edge locations for caching. MEC will provide compute and storage resources even closer to the user. Figure 14 illustrates an example how cloud CDN solutions could utilise MEC platform for optimal content placement and gain access to network information for enhancing user's experience even further. This example does not describe any particular implementation but it is loosely based on Amazon CloudFront solution.

We start with the assumption that generic hardware with compute, store and networking capabilities is deployed in various aggregation sites in a network. The access to these capabilities is offered through the MEC platform. Different deployment models may exist. For example, network operator deploys the MEC platform at network operator's cost with the intention to offer services to over-the-top service providers. A CDN service provider is used as an example of such OTT service provider. The CDN service provider may benefit from shorter latencies but he also gains an access to MEC platform services such as bandwidth management services, which can help the CDN operator to improve customer's experience even further. Alternatively, there may be already a demand for the MEC platform from the CDN service provider who may be interested in improving consumer's experience. In this case, the network operator and the OTT service provider may share the capital expenses.

The CDN operator may access the MEC platform's management and orchestration (2) via an interface (1) and integrate MEC platform with own management system. Then the CDN operator can provide application packages to the MEC platform, and to request the MEC platform to instantiate the applications and to apply needed configurations (e.g. DNS). In our example, the metro aggregation site could run CDN edge instance (similar to AWS edge location) with DNS configuration (3). The system can use a dedicated domain for edge instances or a single domain. The dedicated domain can be used to control the routing of requests (e.g. *.cloudfront.net), which is then also used in application design to control what objects can be cached at the edge. In addition to the edge instances, the CDN architecture could also include intermediary instances (similar to AWS regional edge caches but perhaps with finer granularity than currently used [27]) (4). When a user's client application requests a resource from CDN (5), DNS handling at MEC platform directs the request to the edge instance. The edge instance checks whether the resource is locally available. If the resource is not available, then the edge instance forwards the request to the intermediary instance (6) and so on. Ultimately, the request may be forwarded all the way to the origin location of the resource in cloud (e.g. AWS S3 bucket or origin HTTP server). The response may be cached in the intermediary instances or edge instances of CDN according to CDN's internal decision. It should be noted that the requests could be routed to the edge instances by other means than DNS.

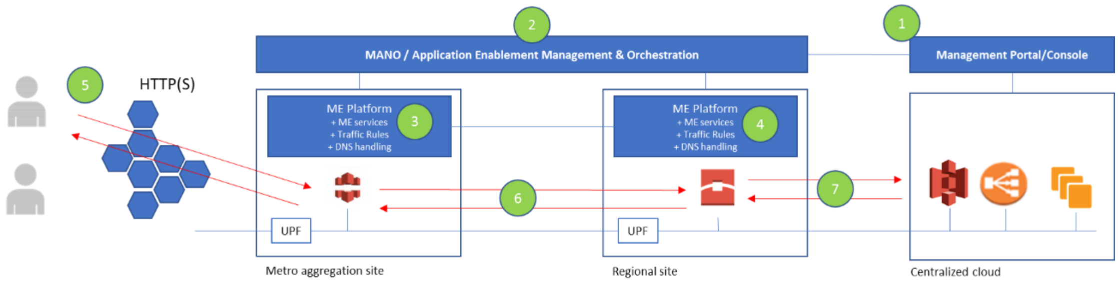


Figure 20 – Example edge cloud CDN with MEC platform

5.5 Seamless Content Delivery

Seamless content delivery aims at offering a transparent experience perceived by the end user when an important transition event occurs at the UE or in the network. In the context of converged network, the event could happen when the end user can move in/out of the coverage of a given service or the network decides to switch the delivery mode (e.g. transition from unicast to multicast/broadcast and vice versa).

The seamless delivery can be achieved with the help of multiple factors from the UE, the network, the content provider, etc. From the content provider perspective, the delivery protocol and/or the protocol's metadata should facilitate the transition. Adaptive streaming such as DASH, HLS could be an example where both mechanism and its metadata can be adapted to seamlessly delivery the content to the end user. 3GPP define in 5G different modes for session and service continuity (SSC) to address the various continuity requirements of different applications/services for the UE [10]. However, these SSC modes normally are linked with a user profile and are applied when the hangover occurs. More importantly, in the context of convergence where both fixed broadband and mobile networks are involved especially these networks may not belong to the same operator, the UE should be proactive by employing an intelligence application logic implemented inside the UE that handles and hides the complexity when the transition event occurs.

At this present time, the seamless content delivery is possible in mobile network with eMBMS capabilities thanks to service continuity [28] when the UE moves in/out of eMBMS coverage. In the fixed broadband network, multicast ABR is being specified by the DVB TM-IPI working group. The challenge in converged network is to provide a seamless and transparent experience perceived by the end users regardless of whether they are using a fixed/broadband and/or mobile connection and regardless of transitions between unicast and multicast/broadcast. More specifically, the challenge is to drive to the convergence between broadcast technology in mobile network and multicast technology in fixed broadband network.

6 Converged Architecture Options

This section describes a number of converged architecture solutions that have been considered in 5G-Xcast. Deliverable 4.1 [4] outlined a number of initial architectural proposals which have been enhanced in this section to include a number of convergence implementation options on top off the original proposals. Deliverable D4.1 focusses on the following two possible 5G Core Network architectures:

- “Option 1”, i.e. a solution deemed as 5G architecture friendly as possible.
- “Option 2”, i.e. the solution deemed with minimal changes to the eMBMS architecture and specification described for LTE.

Deliverable 4.1 enhanced the current functionalities of BM-SC and MBMS GW as they exist in LTE, by splitting them into a control plane part (XCF) and a user plane part (XUF).

In the subsequent solutions the network functions specific to convergence are shown in the figures below. The alternative approaches have different points of convergence depending on the architectural solution. The architectural solutions are listed as follows:

1) D4.1 Option 1 – Trusted with UE and UPF as the convergence points

2) D4.1 Option 1 – Trusted with 5G/RG and UPF as the convergence points

3) D4.1 Option 1 – Overlay

4) D4.1 Option 2 – Trusted with UE and UPF as the convergence points

(Note: ‘Trusted’ and ‘Untrusted’ terminology from 3GPP is explained in section 3.2)

The architecture shown in Figure 21 is taken directly from 5G-Xcast Deliverable 4.1 [4] and is used as a starting point for all subsequent proposed converged architectural solutions.

In this architecture the XUF interfaces with the content provider via xMB-U interface and with the UPF via N6 reference point. The XUF provides multicast and broadcast specific functionalities for the user plane. Also with the participation of the UPF it provides Multilink User Plane functionality, described in Deliverable 4.1 [4]. The control plane functionalities are provided by the XCF, which interacts with other NFs for session control and management. For more information about this architecture, please refer to Deliverable 4.1 [4].

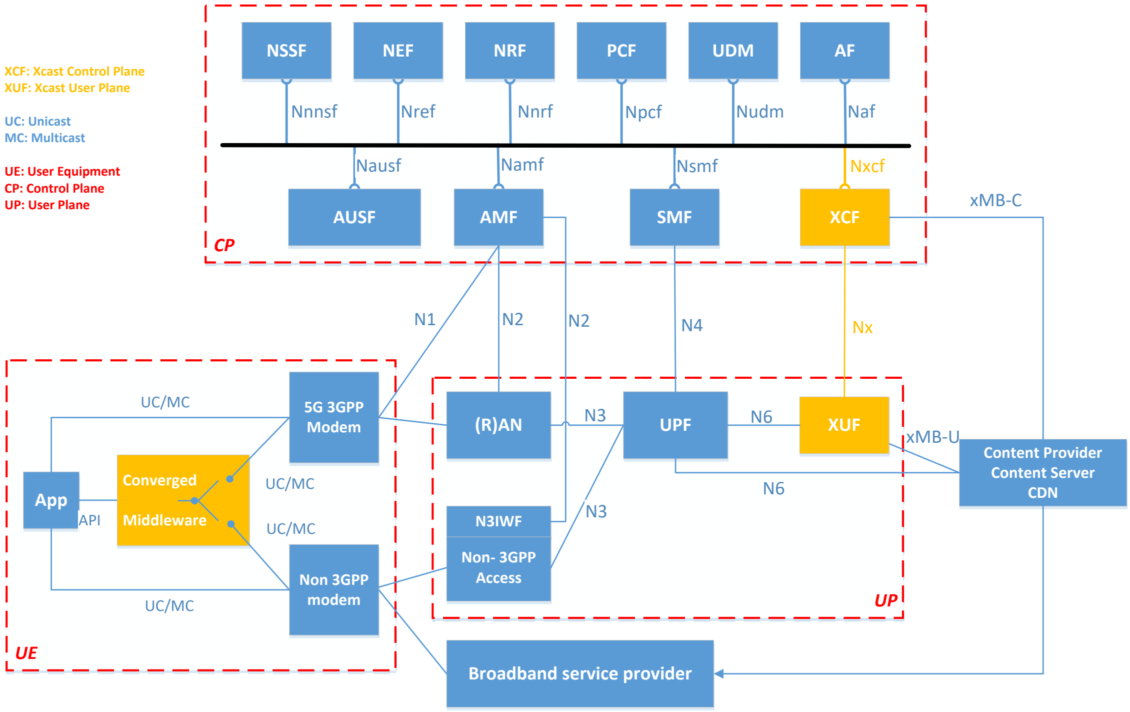


Figure 21 – D4.1 Option 1 - Untrusted

6.1.1 D4.1 Option 1 – Trusted with UE and UPF as the convergence points

This option relates to the Mobile Hybrid Access deployment scenario outlined in section 2.2. This architectural solution is an integrated converged core network approach in which the UE and UPF act as the points of convergence. This solution has a number of benefits including: good access to core information and policy, common authentication and policy enforcement across networks, potential for lowest cost of ownership and deployment for Operators.

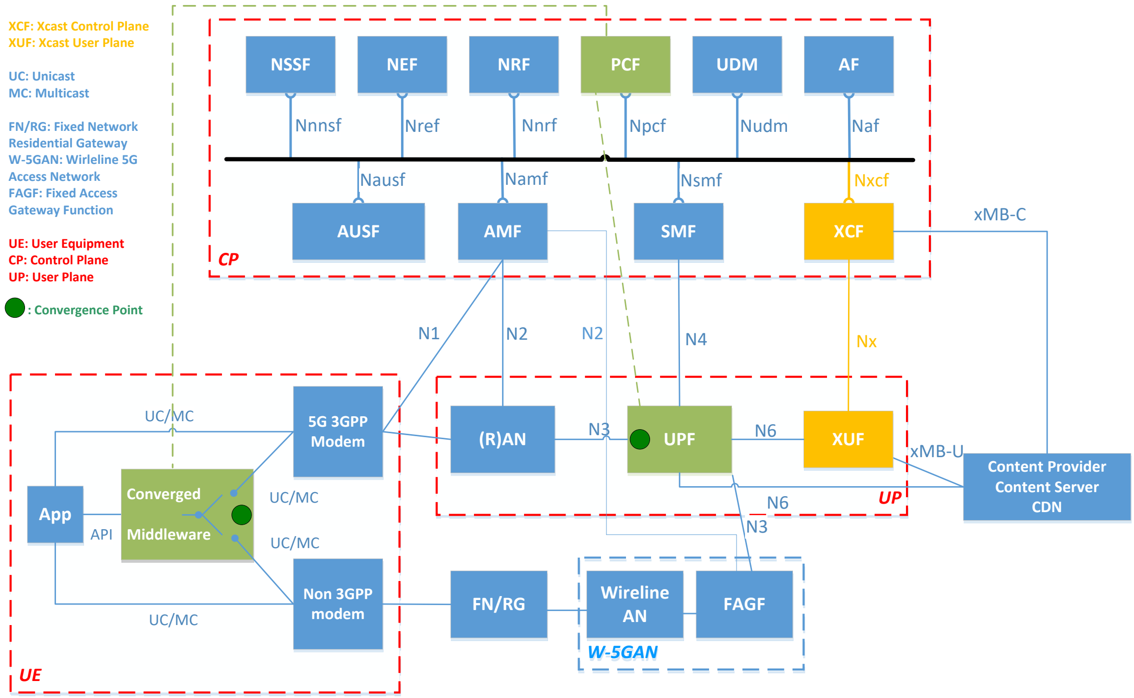


Figure 22 – D4.1 Option 1 – Trusted with UE and UPF as the convergence points

The most relevant aspects for the converged architecture shown in Figure 22 are listed hereafter:

- User Equipment (UE) – In order for this architectural solution to be attainable the UE must have converged middleware support (note: in this use case the converged middleware is Multilink Middleware). The UE has simultaneous connectivity with both the 5G RAN and the Fixed Access Network. The latter connection is achieved by the UE through a Wi-Fi connection to the Fixed Network Residential Gateway which in turn is connected to the wireline access network. The UE also has a logical connection to the PCF.
- Fixed Access Gateway Function (FAGF) - A new entity to be defined by 3GPP acts as an intermediary function between the Fixed Network and Mobile Core. The FAGF has an N3 interface to the UPF.
- User Plane Function (UPF) – Is responsible for handling all the user plane traffic between the converged core network and the UE. It supports simultaneous connectivity with both the 5G RAN and the Fixed Access Network via its N3 interfaces to the FAGF and 5G RAN. The UPF has N6 interfaces to both the XUF and CDN respectively. The UPF also has an N4 connection to the SMF and a logical connection to the PCF.
- Policy Control Function (PCF) – has logical connections to both the UE and UPF. The PCF is responsible for instructing the UE and UPF on the functional behavior to perform for any given multilink session, based on Operator policy.

6.1.2 D4.1 Option 1 – Trusted with 5G/RG and UPF as the convergence points

This option relates to the Fixed Hybrid Access deployment scenario outlined in section 2.1. The architectural solution is an integrated converged core network approach in which the 5G Residential Gateway and UPF act as the points of convergence. This solution has a similar set of benefits as those 6.1.1.

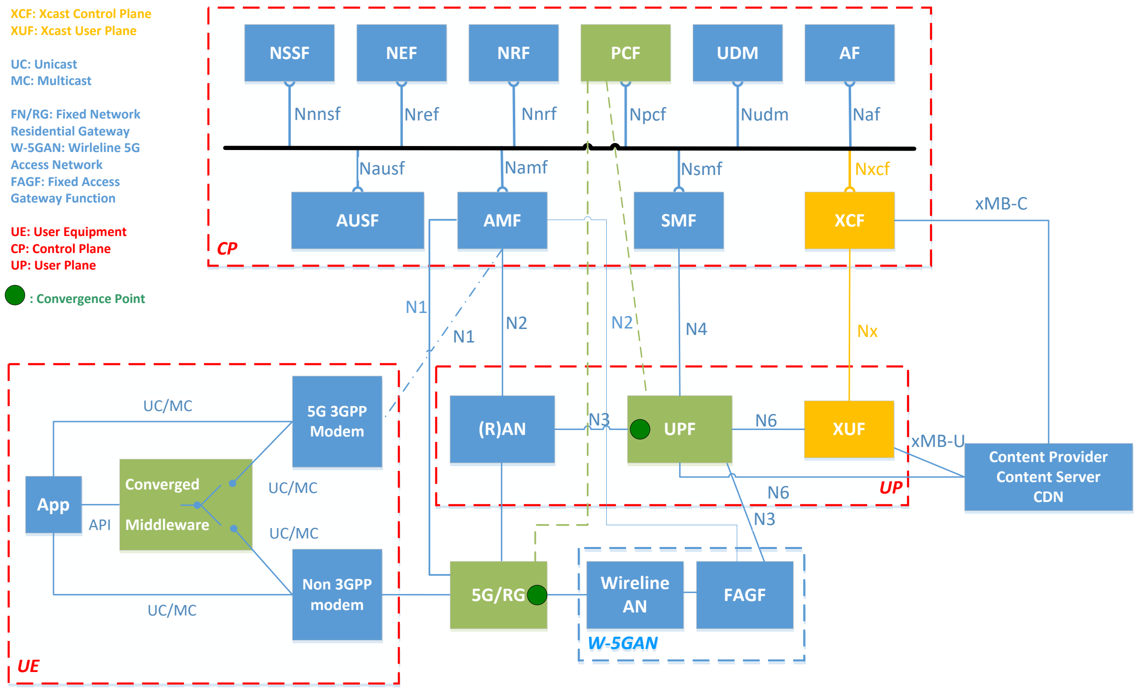


Figure 23 – D4.1 Option 1 – Trusted with 5G/RG and UPF as the convergence points

The most relevant aspects for the converged architecture shown in Figure 23 are listed hereafter: There is no direct reference point anymore between UE and (R)AN and a direct reference point between (R)AN and 5G/RG is introduced. In this option, because the 5G/RG acts as the point of convergence in the end user domain, any device (5G or non-5G) connecting to the 5G/RG reaps the benefits from convergence.

- 5G Residential Gateway (5G/RG) - The 5G/RG has simultaneous connectivity with both the 5G RAN and the Fixed Access Network. The 5G/RG terminates any multilink sessions for the devices connected to it via the LAN. The 5G/RG has an N1 interface to the AMF. It also has a logical connection to the PCF.
- User Equipment (UE) – The UE is intended to be connected to the converged core network only via a Wi-Fi connection to the 5G/RG which in turn is connected to both the Fixed Access Network and the (R)AN. Since the UE is intended to be 5G enabled, it also has its own N1 connection to the AMF and it can be individually authenticated to the converged core network, separately to the 5G/RG.
- User Plane Function (UPF) – Is responsible for handling all the user plane traffic between the core network and the 5G/RG. It supports simultaneous connectivity with both the 5G RAN and the Fixed Access Network via its N3 interfaces to the FAGF and 5G RAN. It has N6 interfaces to both the XUF and CDN respectively. It also has an N4 connection to the SMF and a logical connection to the PCF.
- Policy Control Function (PCF) – has logical connections to both the 5G/RG and UPF. The PCF is responsible for instructing the 5G/RG and UPF on the functional behavior to perform for any given multilink session, based on Operator policy.

6.1.3 D4.1 Option 1 – Overlay

This converged architectural solution shown in Figure 24 is an overlay approach in which there is no or very little integration to the core network. This solution has a number of benefits including: simple to deploy as requires little change to RAN and core networks, easiest to upgrade capability to support new services, etc. compared to previous solutions. However, this solution has no or limited access to core network KPIs and no or limited integrated with core network policy.

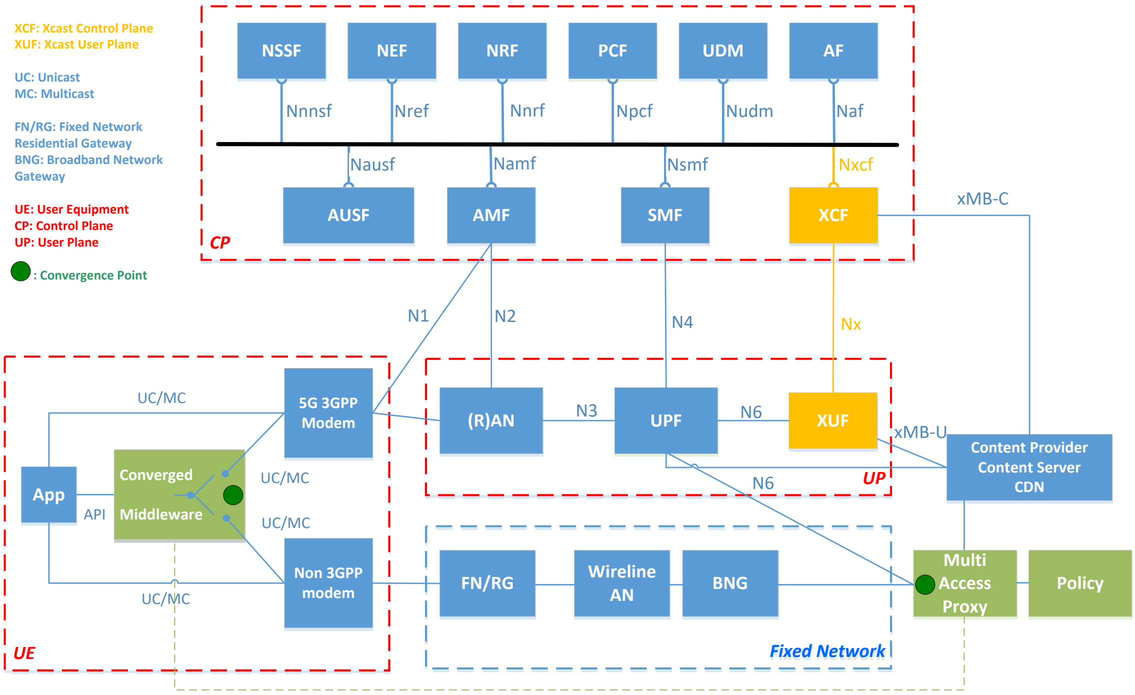


Figure 24 – D4.1 Option 1 – Overlay

The most relevant aspects for the converged architecture shown in Figure 24 are listed hereafter:

- User Equipment (UE) - In order for this architectural solution to be attainable the UE must have converged middleware support in order for this architectural solution to be attainable (note: in this use case the converged middleware is Multilink Middleware). The UE has simultaneous connectivity with both the 5G RAN and the Fixed Access Network. The UE has a Wi-Fi connection to the Fixed Network Residential Gateway which in turn is connected to the wireline access network. In this solution the UE does not have a logical connection to the PCF.
- Multi Access Proxy – can have connectivity across multiple access networks. In this use case it has connectivity to the UPF in the core network and to the BNG in the fixed network.
- Policy – It is responsible for instructing the UE and Multi Access Proxy on the functional behavior to perform for any given multilink session, based on defined policies.

6.1.4 D4.1 Option 2 – Trusted with UE and UPF as the convergence points

In Deliverable 4.1 Option 2 leverages the LTE eMBMS architecture for offering point-to-multipoint services. This solution allows for the synchronisation for multi-cell transmission at the XUF. Option 2 aims at minimising the changes to the functionalities developed in LTE eMBMS. Option 2 can also offer the transparent multicast transport through the same enhancements over the unicast architecture and Multilink functionality as in the case of Option 1.

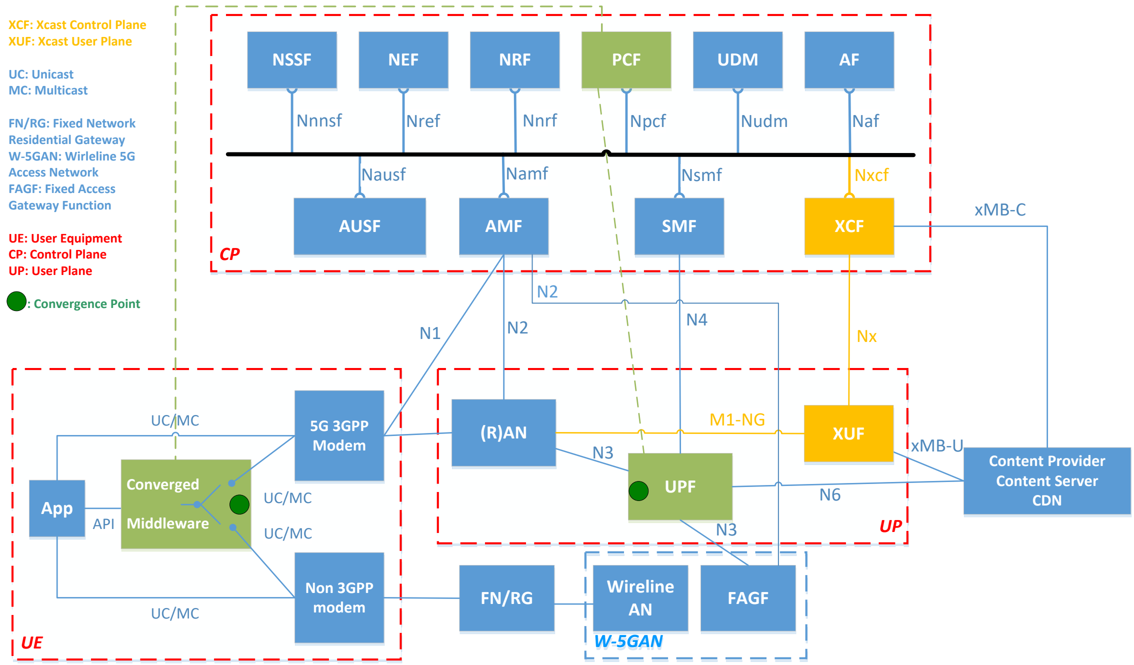


Figure 25 – D4.1 Option 2 – Trusted with UE as convergence point

The most relevant aspects for the converged architecture shown in Figure 25 are listed hereafter:

- User Equipment (UE) - In order for this architectural solution to be attainable, the UE must have converged middleware support (note: in this use case the converged middleware is Multilink Middleware). The UE has simultaneous connectivity with both the 5G RAN and the Fixed Access Network. The latter connection is achieved by the UE through a Wi-Fi connection to the Fixed Network Residential Gateway which in turn is connected to the wireline access network. The UE also has a logical connection to the PCF.
- Fixed Access Gateway Function (FAGF) - A new entity defined by 3GPP acts as an intermediary function between the Fixed Network and Mobile Core. The FAGF has an N3 interface to the UPF.
- User Plane Function (UPF) – Is responsible for handling all the user plane traffic between the converged core network and the UE. The UPF supports simultaneous connectivity with both the 5G RAN and the Fixed Access Network via its N3 interfaces to the FAGF and 5G RAN. It has an N6 interface to the CDN. It also has an N4 connection to the SMF and a logical connection to the PCF. Differently from section 6.1.1, in this converged architecture XUF connects directly to the (R)AN (over M1-NG reference point) rather than through the UPF.
- Policy Control Function (PCF) – has logical connections to both the UE and UPF. The PCF is responsible for instructing the UE and UPF on the functional behavior to perform for any given multilink session, based on Operator policy.

7 Considerations on the security aspects

A 3GPP (Release 16) Study Item on the *Wireless and Wireline Convergence for the 5G system architecture* is currently **in progress** within 3GPP SA WG2 (see [36], [37]). A "companion" 3GPP (Release 16) Study Item on the *Security of the Wireless and Wireline Convergence for the 5G system architecture* is also **in progress** within 3GPP SA WG3 (see [38],[39]). 3GPP **normative** guidance cannot be provided within a Study Item, but it can be provided within appropriate Work Item(s) to be readily started once the study phase concludes that 3GPP normative guidance is needed or useful.

With reference to the n.4 converged architecture Options described in Section **¡Error! No se encuentra el origen de la referencia.**, the first n.3 are based on the initial architectural proposal introduced within Deliverable 4.1 [4] as "Alternative 1", whereas the fourth is based on the initial architectural proposal "Alternative 2".

General considerations on security aspects both for "Alternative 1" and "Alternative 2" are provided in a dedicated section within Deliverable D4.1 [4] and are not repeated here.

Further considerations on security aspects are provided hereafter for the reference points and the new elements (e.g. 5G/RG, FN/RG) being introduced by each converged architecture Option.

7.1 Considerations on security aspects for the converged architecture Option #1

As shown in Figure 22, in this architecture Option the points of convergence are the UPF and the Converged Middleware (within the UE), being able to receive data both from "5G 3GPP" and from "Non 3GPP" modems. From the UPF, beside the "3GPP" path (going to the UE through the (R)AN) and the reference point N3) there is also a "Non 3GPP" path involving a W-5GAN and a Fixed Network Residential Gateway (FN/RG).

Transport security for the interfaces between W-5GAN and 5GC needs to be provided. More in detail, confidentiality protection, integrity protection and replay-protection shall be supported over N2 (i.e. between W-5GAN and AMF) and N3 (i.e. between W-5GAN and the UPF). Security mechanisms for N2 and N3 are provided within [35] (see Clause 9.2 and Clause 9.3, respectively), however *Wireless and Wireline Convergence for the 5G system architecture* topic is expected to imply changes on those reference points and the security implications are going to be studied accordingly within [39].

Also the security along the "non 3GPP" path, from the UPF to the UE through the W-5GAN and the FN/RG is currently under study within 3GPP ([39]). For instance:

- authentication and authorization aspects for the FN/RG need to be studied;
- a 5G capable UE behind a FN/RG shall be able to authenticate to the 5GC (and any deviation from usual 5G authentication is expected to need strong motivation);
- User plane data of a 5G-capable UE behind a FN/RG shall be forwarded via the RG in a manner that preserves data confidentiality, integrity and replay-protection. For a 5G-capable UE connecting to the 5G core network via a FN-RG the 5GS shall provide a level of user plane security that is at least equal to the one provided to a UE connected via NG RAN.
- The user plane of a FN/RG to the 5GC shall provide a level of user plane security that is at least equal to the one provided to a UE connected via a gNB.

7.2 Considerations on security aspects for the converged architecture Option #2

As shown in Figure 23Figure 22, in this architecture Option the points of convergence are the UPF and the 5G Residential Gateway (5G/RG). The UE may get the content through a “Non 3GPP” path involving a W-5GAN and a 5G/RG and also through a (modified) “3GPP” path where the direct reference point between UE and (R)AN is removed and “diverted” through the 5G/RG.

Confidentiality protection, integrity protection and replay-protection need to be supported on the interfaces between W-5GAN and 5GC (i.e. over N2 and N3) and between 5G/RG and AMF (i.e. over N1). Security mechanisms for N1, N2 and N3 are provided within [35] however *Wireless and Wireline Convergence for the 5G system architecture* topic is expected to imply changes on those reference points and the security implications are going to be studied accordingly within [39].

Also the security along the “Non 3GPP” path, from the UPF to the UE through the W-5GAN and the 5G/RG is currently under study within 3GPP ([39]). For instance:

- authentication for the 5G/RG needs to be studied;
- a 5G capable UE behind a 5G/RG shall be able to authenticate to the 5GC (and any deviation from usual 5G authentication is expected to need strong motivation);
- User plane data of a 5G-capable UE behind a 5G/RG shall be forwarded via the RG in a manner that preserves data confidentiality, integrity and replay-protection. For a 5G-capable UE connecting to the 5G core network via a 5G-RG the 5GS shall provide a level of user plane security that is at least equal to the level provided to a UE connected via NG RAN.
- The user plane of a 5G/RG to the 5GC shall provide a level of user plane security that is at least equal to the level provided to a UE connected via a gNB.

7.3 Considerations on security aspects for the converged architecture Option #3

As shown in Figure 24Figure 22, in this converged architecture Option there is no or very little integration to the core network. The points of convergence are the Converged Middleware (within the UE) and the Multi Access Proxy having connectivity across multiple access networks (e.g. to the UPF in the 5GC and to the BNG in the fixed network). Since the UE may get the content both through a “3GPP” and through a “Non 3GPP” paths that are basically independent, for this converged architecture Option also the considerations on security aspects looks quite independent. For the “3GPP” path the considerations on security aspects provided for the “Alternative 1” within Deliverable D4.1 [4] apply.

7.4 Considerations on security aspects for the converged architecture Option #4

This converged architecture Option is based on the initial architectural proposal introduced within Deliverable 4.1 [4] as “Alternative 2” that aims at minimising the changes to the functionalities developed in LTE eMBMS.

With reference to Figure 25Figure 24Figure 22, as for the Option #1 (see section 6.1.1), in this converged architecture Option the points of convergence are the UPF and the Converged Middleware (within the UE).

The UE receives data both from a “3GPP” path and from a “Non 3GPP” path involving a W-5GAN and a Fixed Network Residential Gateway (as in the converged architecture Option #1).

Regarding the considerations on security aspects, those provided for the “Alternative 2” within [4] and the ones specifically provided for the converged architecture Option #1 (see Section 7.1) apply.

8 Converged Broadcast Network

Today's broadcast networks, i.e. Digital Terrestrial Television (DTT) networks, possess only a downlink to distribute their content in point-to-area mode. Usually these systems are operated in spectrum bands allocated to a 'broadcast service' and do not provide a corresponding uplink. In 3GPP, this type of broadcast was first introduced with 3GPP Rel-14 and it is called "Receive-Only-Mode (ROM) broadcast", throughout the 5G-Xcast documentation this mode is also referred to as "terrestrial broadcast". For "ROM broadcast", the UE need not register/attach with the network. Before 3GPP Rel-14 "ROM-broadcast" was not supported by 3GPP.

These broadcast networks are typically operated by a Broadcast Network Operator (BNO) which is either part of a Broadcast Service Provider (BSP) or contracted by them to provide services complying to a given Service Level Agreement (SLA). Typically these are operated independently of a Mobile Infrastructure Operator (MIO) although this is not always the case.

The usage of the term "broadcast" within the mobile industry originates from mobile systems that are always operated in spectrum allocated to mobile services, i.e. that have both uplink and downlink, and where the UE is registered with the network. This means that the UE is always capable of unicast communication with the network even if this is not required for reception of actual data carried over point-to-multipoint broadcast or multicast bearers.

The 5G-Xcast solution intends to address both broadcast networks and mobile systems (as defined above) through a converged approach, i.e. making use of as many common components as possible.

In the first instance the project will address mobile systems, focussing on multicast/unicast operation while considering how broadcast could also be enabled on this. The intention is to take into account the requirements of broadcast networks so that an extension of these later isn't impeded. The aim is to enable the synchronisation of cells (Single Frequency Network operation) within areas including across large areas, e.g. nationwide, as well as small areas, e.g. encompassing a stadium/venue, that can also vary in size.

Secondly the project will look to extend this architecture to operate on broadcast networks with the aim of minimising the differences. This will include support for 100% broadcast carriers and receive-only devices ("ROM broadcast") on these stand-alone networks while going beyond what has been specified in 3GPP Rel-14 under LTE by being based on the 5G core network architecture being developed within 5G-Xcast WP4.

For more information about broadcast network aspects please refer to "5G-Xcast Deliverable D4.3 Session Control and Management" [5].

9 Conclusions

The key industry drivers for full network convergence between fixed and mobile networks were outlined in this document. Full network convergence is defined in the introductory section as the “full integration of fixed and mobile access networks into a single end-to-end architecture, supporting a single set of identities, common authentication, seamless handover and simultaneous access across any access technologies”. Improved seamless user experience and efficiency are the main benefits of full network convergence.

The document describes several of the existing mechanisms that achieve some form of (partial) network convergence, together with their main capabilities and examples of use cases where they can be applied.

The document explains some of the ongoing efforts in standards bodies to fulfil the convergence vision. The ambition of the 5G-Xcast project is to have a very flexible and modular architecture that allows operators to simultaneously deliver fixed-only, mobile-only or combined fixed and mobile converged services in the most efficient manner, by using the same functional architecture and deploying only the functions required for each particular service type. A converged 5G system could enable operators to migrate their fixed and mobile customers onto a common platform if by doing so they are able to deliver both customer experience improvements and operational efficiencies. Key technologies were discussed within this document with ATSSS and Multilink being vital to the delivery of converged services.

Deliverable D4.1 “Mobile Core Network” [4] focusses on the following two possible 5G Core Network architectures:

- “Option 1”, i.e. a solution deemed as 5G architecture friendly as possible
- “Option 2”, i.e. the solution deemed with minimal changes to the eMBMS architecture and specification described for LTE.

In order to identify valid architectural solutions to enable a 5G *converged* Core Network, both D4.1 Options listed above were considered within this Deliverable. More precisely, a trusted architecture in which the UE and the UPF act as the points of convergence was detailed for the D4.1 “Option 2” (see section 6.1.4) and the following solutions were detailed for the D4.1. “Option 1”:

- A trusted architecture in which the UE and the UPF act as the points of convergence (see section 6.1.1).
- A trusted architecture in which the 5G Residential Gateway and the UPF act as the points of convergence (see section 6.1.2).
- An overlay solution which is less integrated with the core network than the previous two versions, which uses a Multi Access Proxy outside of the core and the UE as the points of convergence (see section 6.1.3).

The converged architectures were analysed in terms of whether Multilink could be supported for all traffic types, i.e. unicast, multicast and broadcast. Multilink aggregation is currently implemented only for unicast streams, bringing together distinct unicast connections to enable a better quality for a stream. In the current 5G-Xcast architecture (for each alternative) the performance of each available link is measured between Multilink User plane function and Multilink middleware by exchanging the specific service information. One of the challenges in the 5G-Xcast project is to see how a sophisticated non-naive broadcast or multicast strategy over multilink can be achieved. In this case, 5G-Xcast will benefit not only from seamless transitions between broadcast and multicast to unicast (and vice versa), and/or seamless unicast experience side by side with broadcasts to others, but also higher-quality broadcast will be enabled by using multiple connections. The other challenge of 5G-Xcast is related to the specific usage of multi-connectivity in wireless.

References

[1] "Definition of Use Cases, Requirements and KPIs," D. Ratkaj and A. Murphy, Eds., Deliverable D2.1, 5G-PPP 5G-Xcast project, Oct. 2017.

[2] "NGMN 5G White Paper", NGMN, 17/02/15, https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf

[3] https://www.mcg.upv.es/en/bmsb2018/speakers/

[4] "Mobile Core Network," T. Tran, Ed., Deliverable 4.1, 5G-PPP 5G-Xcast project, June 2018. http://5g-xcast.eu/documents/

[5] "Session Control and Management," B. Altman, Ed., Deliverable 4.3, 5G-PPP 5G-Xcast project. http://5g-xcast.eu/documents/

[6] "3GPP TS 36.300 – EUTRA/EUTRAN Overall Description (Release 14)"

[7] "3GPP TS 36.331 – Radio Resource Control (RRC) Protocol specification (Release 14)"

[8] "3GPP TS 25.331 – Radio Resource Control (RRC) Protocol specification (Release 15)"

[9] "3GPP TS 23.402 – Architecture enhancements for non-3GPP access (Release 15)"

[10] "3GPP TS 23.501 - System Architecture for the 5G System; Stage 2 (Release 15)"

[11] "3GPP TS 23.502 - Procedures for the 5G System; Stage 2 (Release 15)"

[12] "3GPP TR 22.891 - Feasibility Study on New Services and Markets Technology Enablers; Stage 1 (Release 14)"

[13] "3GPP TR 22.863 - Feasibility Study on New Services and Markets Technology Enablers - Enhanced Mobile Broadband; Stage 1 (Release 14)"

[14] "3GPP TS 22.261 - Technical Specification Group Services and System Aspects; Service requirements for the 5G system; Stage 1 (Release 16)"

[15] "3GPP TS 26.247 - Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH) (Release 15)"

[16] "3GPP 5G Roadmap", http://www.3gpp.org/images/articleimages/ongoing_releases_900px.JPG

[17] "3GPP TR 23.716 - Study on the Wireless and Wireline Convergence for the 5G system architecture (Release 15)"

[18] "3GPP TR 23.793 - Study on Access Traffic Steering, Switching and Splitting support; In the 5G system architecture (Release 15)"

[19] "3GPP TS 28.530 – Management and orchestration of 5G networks; Concepts, use cases and requirements (Release-15)"

[20] "3GPP TR 38.300 - NR; NR and NG-RAN Overall Description; Stage 2 (Release 15)"

[21] "3GPP TS 28.531 – Management and orchestration of 5G networks; Provisioning (Release-15)"

[22] "Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems", Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon, IEEE/ACM Transactions on Networking, Vol. 17, No. 5, October 2009.

[23] "An Intelligent Satellite Multicast and Caching Overlay for CDNs to Improve Performance in Video Applications", Chris Brinton, Ehsan Aryafar, Steve Corda (SES), Stan Russo (SES), Ramiro Reinoso (SES), Mung Chiang, AIAA ICSSC 2013

[24] RFC 7234, "Hypertext Transfer Protocol (HTTP/1.1): Caching", June 2014.

[25] RFC 7231, "Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content", June 2014.

[26] Google Cloud CDN documentation, URL: https://cloud.google.com/cdn/docs/

[27] Amazon CloudFront, URL: https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/Introduction.html

[28] "3GPP TS 26.346 - Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs (Release 15)"

[29] "Nokia White Paper 'Two Worlds Collide - The inevitable, imminent convergence of mobile transport and fixed access networks'", Nokia, 2016.

[30] "Content Delivery Vision," N. Nouvel, Ed., Deliverable D5.1, 5G-PPP 5G-Xcast project, Nov. 2017.

[31] "Key Technologies for the Content Distribution Network," T. Stevens, Ed., Deliverable D5.2, 5G-PPP 5G-Xcast project, Oct. 2017.

[32] "Point-to-Multipoint Communication Enablers for the Fifth-Generation of Wireless Systems", D. Gomez-Barquero, D. Navratil, S. Appleby and M. Stagg, IEEE Communications Standards Magazine, vol. 2, no. 1, pp. 53-59, March 2018.

[33] "Optimization of Tune-in and End-to-end Delay in DASH Broadcast over ROUTE," N. Bhat H.M and W. Zia, Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Valencia, Spain, 2018.

[34] "Evolution toward fixed and mobile network convergence," M. Cuevas, S. Johnson and J. Hart, ITP The Journal Vol 12 Part 2, 2018.

[35] "3GPP TS 33.501 - Security architecture and procedures for 5G system (Release 15)"

[36] 3GPP Study Item Description SP-170115 on "Study on the Wireless and Wireline Convergence for the 5G system architecture" (Release 16)

[37] http://www.3gpp.org/ftp//Specs/archive/23_series/23.716/; 3GPP TR 23.716: "Study on the Wireless and Wireline Convergence for the 5G system architecture (Release 16)"

[38] 3GPP Study Item Description SP-180438 on "Study on the security of the Wireless and Wireline Convergence for the 5G system architecture" (Release 16)

[39] http://www.3gpp.org/ftp//Specs/archive/33_series/33.807/; 3GPP TR 33.807: "Study on the security of the Wireless and Wireline Convergence for the 5G system architecture (Release 16)"