<Deliverable D3.2: Analysis of horizontal targets for functional convergence >

Grant Agreement number: 317762

Project acronym: COMBO

Project title: COnvergence of fixed and Mobile BrOadband access/aggregation networks

Funding Scheme: Collaborative Project – Integrated Project

Date of latest version of the Deliverable: 30 April 2015

Delivery Date: 30 April 2015

Leader of the Deliverable: IMT-TB

File Name: COMBO_D3.2_v1.0.docx

Version: V1.0

Authorisation code: PU = *Public*

Project coordinator name, title and organisation: Jean-Charles Point, JCP-Connect

Tel: + 33 2 23 27 12 46

E-mail: pointjc@jcp-connect.com

Project website address: www.ict-combo.eu

Executive Summary of the Deliverable

The purpose of Work Package 3 “Fixed Mobile Convergent Architectures” is to propose, define and technically assess candidate architectures for future Fixed-Mobile Convergent (FMC) networks, both in terms of data plane and control plane. This document D3.2 specifies and develops key functional blocks for FMC, called Horizontal Targets, which are consistent sets of FMC generic functions allowing functional convergence and solving some key “horizontal” end-user related tasks in a 5G context such as universal authentication and interface selection control. It leverages on previous work of COMBO Work Package 3, and in particular on the identification by D3.1 of the main functional groups, which need significant effort to reach functional convergence.

D3.2 focuses only on the functional aspects of FMC. It develops high-level functional solutions for FMC, independently from the actual organization and implementation of future 5G networks. Horizontal targets are shown to be essential intermediate goals for allowing true FMC in 5G networks. Concrete indications are provided regarding the available protocols and methods that could be used to implement the high-level solutions for FMC. Some procedures, protocol message exchange or flow charts are thus made available here, but they do not pretend to be exhaustive since D3.2 is the first deliverable addressing COMBO solutions.

Two horizontal targets have been derived, based on the analysis of functional groups performed in D3.1:

- **HT1: Converged subscriber and session management:** This horizontal target addresses convergence of authentication and subscriber data management. It aims at avoiding the drawbacks of the proliferation of identities, user/subscriber profiles and authentication mechanisms in fixed and mobile networks.
- **HT2: Advanced interface selection and route control:** The motivation of this horizontal target is to provide the FMC network operator with means to dynamically control mobile traffic data paths, when several data paths are available (e.g. via fixed network or Wi-Fi access), while maintaining session continuity whenever necessary.

HT1 and HT2 proposed solutions are shown to solve most of the functional gaps existing between the current, non-converged, situation and a situation where an FMC operator can take advantage of a global control of its resources and a global management of its subscribers. Diverse and multiple initiatives already address the scopes of HT1 and HT2. Nevertheless, they do not allow actual merging of functionalities in a 5G context, do not meet some of the requirements from the mobile network, are implementation specific, or only define protocols without taking into consideration the architecture of the access/aggregation network.

Although a full merging of functionalities is highly desirable for classical communication or data services delivered to human users in an FMC context, such a merging may not be required, nor even desirable, for all services carried in the future 5G networks (e.g. M2M and IoT related services). The proposed solutions therefore

do not replace existing ones, but allow them to smoothly interact, and to provide a virtual merging of functions.

COMBO proposes and develops universal subscriber and user Authentication (uAUT) as a set of technical solutions for solving HT1. This functional block allows a user to authenticate once and have access to multiple networks and/or services. uAUT interfaces with the management plane and is part of the control plane. It leverages on the 3GPP's User Data Convergence (UDC) concept, namely splitting subscribers' data repository from the application logic specific to each access type. uAUT is a single functional block proposed by COMBO as a significant improvement of UDC concept. It would link several application logics (called "Front Ends" in the UDC framework) with a single global User Data Repository (UDR). The proposed solution is described in detail and is shown to complement the existing UDC framework with various original features regarding its implementation, its operation and its applicability.

So as to fulfil HT2 requirements, COMBO proposes a set of functional blocks that realises a "Universal Data Path Management" (uDPM). It allows redirecting (part) of mobile data traffic over the fixed/Wi-Fi data paths from the default LTE path, while maintaining session continuity (even during mobility) and enabling multiple data paths to be used simultaneously for a given user session. The uDPM solution includes alternatives for the current implementations of handover and mobility support, forwarding (and especially tunnelling), charging and route control. It also leverages on uAUT proposed for solving HT1.

uDPM is composed of four functional blocks which are designed, developed and described: the Decision Engine is part of the control plane and interacts with the management plane. It selects how the session is to be mapped on data paths. The Data Path Creation and Destruction functional block handles the control of path creation/destruction on the available interfaces; the Path Coordination and Control block ensures that concurrent data paths smoothly deliver the packets corresponding to the session, and that session continuity is guaranteed, even in case of UE mobility; both are part of the control plane. Finally, the Session Mapping Execution block applies the session mapping decision taken by the Decision Engine; it is part of the transfer plane.

While uAUT can be considered as a unique solution for HT1, uDPM is a set of partial solutions to specific problems; three such solutions are identified: "very tight coupling" between Wi-Fi and LTE access, which allows a user to seamlessly move from LTE to Wi-Fi, smooth SIPTO based content distribution, which allows a user to seamlessly stream video traffic thanks to Local Gateways (LGWs), and reactive content placement, which is based on the Content Distribution Service (CDS) management to react to user location in order to improve content placement.

The proposed uAUT and uDPM functional blocks will allow the support of advanced FMC use cases, namely "Unified FMC access for mobile devices" (UC1), "Converged content caching for unified service delivery" (UC2), "Universal access bundling for residential gateway" (UC4), and "Network sharing" (UC8). It is shown how uAUT and uDPM implement in a unified way the advanced functional features required by these FMC use cases.

Development of uAUT and uDPM functional blocks is a fundamental step towards attaining the functional convergence in future 5G networks. Functional developments described in this document are feeding implementation of various selected functional test cases, which are being developed in Work Package 6. In particular, the functional components of COMBO demonstration will be described in D6.2 along the functional blocks defined in D3.2 for uAUT and uDPM. Also, based on the development of uAUT and uDPM described in this deliverable, Task 3.2 will describe, analyse and compare alternative network scenarios for functional convergence, based on the NG-POP concept, which was introduced in D3.1. The NG-POP is a location in the network, where the operator could implement multiple functions, including the IP edge for all network types. T3.2 will show how the technical solutions for horizontal targets could be organized and implemented in actual 5G networks considering two alternative scenarios.

One alternative is a centralised solution, with a small number of NG-POP locations, typically at the sites of core Central Offices (COs), which are the edges of the current fixed aggregation network. The other alternative relies on a larger number of NG-POP locations, located in the current main COs; this location corresponds to an extension of the IP backbone towards the access network. In both cases, the advantages brought by the SDN and NFV concepts shall be adopted and assessed. Specifically, this development of two alternative network scenarios for functional convergence will be the subject of deliverable D3.5.

List of authors

Full Name – E-mail	Company – Country Code
Lander Alonso – lander.alonso@fon.com	FON – ES
Dirk Breuer – D.Breuer@telekom.de	DTAG – DE
Selami Çiftçi – celami.ciftci@argela.com.tr	ARGELA
Tibor Cinkler – cinkler@tmit.bme.hu	BME – HU
Souheir Eido – souheir.eido@telecom-bretagne.eu	IMT-TB – FR
Onur Eker – Onur.Eker@argela.com.tr	ARGELA
Annie Gravey – annie.gravey@telecom-bretagne.eu (editor)	IMT-TB – FR
Stéphane Gosselin – stephane.gosselin@orange.com	Orange – FR
Stefan Höst – stefan.host@eit.lth.se	ULUND – SW
Younes Khadraoui – younes.khadraoui@telecom-bretagne.eu	IMT-TB – FR
Xavier Lagrange – xavier.lagrange@telecom-bretagne.eu	IMT-TB – FR
Zhe Li – zhe.li@jcp-connect.com	JCP – FR
Tahar Mamouni – tahar.mamouni@orange.com	Orange – FR
Thomas Monath – Thomas.Monath@telekom.de	DTAG – DE
Peter Olaszi – polaszi@aitia.ai	AITIA
Serban Purge – serban.purge@orange.com	Orange – FR
Jose Torrijos Gijón – jgijon@tid.es	TID – ES

List of reviewers

Full Name – E-mail	Company – Country Code
Achille Pattavina	PoliMi – IT
Ricardo Martinez	CTTC – ES
Serban PURGE	Orange – FR

Approval

Approval	Full Name – E-mail	Company – Country Code	Date
Task Leader	Annie Gravey – annie.gravey@telecom-bretagne.eu	IMT-TB – FR	30/04/2015
WP Leader	Dirk Breuer D.Breuer@telekom.de	DT – DE	30/04/2015
Technical Leader	Stéphane Gosselin stephane.gosselin@orange.com	Orange – FR	30/04/2015
Project Coordinator	Jean-Charles Point - pointjc@jcp-connect.com	JCP – FR	30/04/2015
Other (PMC, SC, etc)			

Document History

Edition	Date	Modifications / Comments	Author
v0.00	18/01/2015	First version	Annie Gravey
V0.01	05/02/2015	Input to section 2.4	Thomas Monath
V0.02	06/02/2015	Inputs to section 2	Stéphane Gosselin
V0.03	06/02/2015	Inputs to 2.2.3, 2.2.4, 4.2.2, 4.3	Zhe Li
V0.04	09/02/2015	Inputs to 4.1, 4.1	Onur Eker
V0.05	11/02/2015	Inputs to 3.4	Jose Torrijos Gijón
V0.10	12/02/2015	Version to be discussed during telco 12/05/2015	Annie Gravey
V0.11	12/02/2015	Version after telco, including input by Orange	Tahar Mamouni and Annie Gravey
V0.12	13/02/2015	Input to 2.2.1	Souheir Eido
V0.13	15/02/2015	Input to 3.2 and 3.3	Lander Alonso
V0.14	18/02/2015	Input to 2.2.2	Stefan Host and Xavier Lagrange

V0.15	18/02/2015	Input to 3.4.3	Onur Eker
V0.20	19/02/2015	Version to be discussed during telco 19/02/2015	Annie Gravey
V0.21	20/02/2015	Creation of new section 5, new section 2.2.5	Annie Gravey
V0.22	21/02/2015	New text for 5.1, initial text for 5.3	Jose Torrijos Gijón
V0.30	03/03/2015	New text for 2.4, new text for 2.3, initial text for 4.2.1, new text for 5.2 and 5.5 from initial 4.3 (removed)	Stefan Host, Xavier Lagrange, Younes Khadraoui, Tahar Mamouni, Peter Olaszi, Lander Alonso, Tibor Cinkler, Zhe Li
V0.31	04/03/2015	Modified text for 5.5; suppression of 5.4	Jose Torrijos Gijón
V0.32	04/03/2015	Review of section 2.2	Achille Pattavina
V0.33	04/03/2015	Text for 4.2.2.2 and removal of 4.2.2.3	Zhe Li
V0.34	04/03/2015	Revision of section 2.1	Stephane Gosselin
V0.35	05/03/2015	New text for 2.4	Stefan Host, Xavier Lagrange, Younes Khadraoui,
V0.40	05/03/2015	Text for 4.2.3	Stefan Host, Xavier Lagrange, Younes Khadraoui, Thomas Monath
V0.41	11/03/2015	Text for 4.1 and 4.2	Annie Gravey
V0.42	11/03/2015	Revision of 5.1 and 5.3	Xavier Lagrange
V0.43	11/03/2015	Revision of 5.4	Zhe Li and Jose Torrijos Gijon
V0.44	11/03/2015	Revision of 4.2.1	Lander Alonso and Tibor Cinkler
V0.45	11/03/2015	Revision of 4.2.2	Zhe Li
V0.50	12/03/2015	New version to be discussed during Telco	Annie Gravey
V0.51	12/03/2015	Revision of 4.2.3	Stefan Host and Annie Gravey

V0.52	12/03/2015	Inclusion of decisions taken during telco	Annie Gravey
V0.53	16/03/2015	New text for 3.1	Peter Olaszi
V0.54	16/03/2015	New text for 2.3	Xavier Lagrange, Younes Khadraoui,
V0.55	16/03/2015	New text for 2.2.5	Tahar Mamouni
V0.56	16/03/2015	New text for 3.2	Lander Alonso
V0.60	16/03/2015	To be discussed during telco	Annie Gravey
V0.61	18/03/2015	New text for 4.1 and 4.2	Annie Gravey, Stefan Host
V0.62	19/03/2015	New fig 4	Souheir Eido
V0.63	21/03/2015	Intermediate conclusions of chapter 2	Annie Gravey
V0.70	24/03/2015	Intermediate conclusions of chapters 3 and 4	Annie Gravey
V0.71	24/03/2015	Modifications to section 2.2	Zhe Li, Lander Alonso
V0.72	26/03/2015	Modifications to sections 2 and 3	Tahar Mamouni
V0.73	26/03/2015	Editorial work on Tahar's proposal	Annie Gravey
V0.80	30/03/2015	Clean version to discuss before review	Annie Gravey
V0.81	30/03/2015	Revision of material regarding HT1 Revision of 4.2.1 Revision of 5.2 Revision of Section 2	Tahar Mamouni, Lander Alonzo, Peter Olaszi, Annie Gravey, Zhe Li, Jose Torrijos Gijón
V0.82	02/04/2015	Revision of Section 3. Clean version	Tahar Mamouni, Peter Olaszi, Lander Alonso
V0.83	04/04/2013	Revision of Section 2	Xavier Lagrange
V0.84	08/04/2015	Revision of Section 4 Revision of Sections 3.3.2, 3.4 Revision of Figure 10	Xavier Lagrange Tahar Mamouni Younes Khadraoui
V0.85	08/04/2015	Clean Version without revision marks	Annie Gravey

V0.86	09/04/2015	Proposal for 4.2	Tibor Cinkler
V0.87	12/04/2015	Review of Sections 1, 2, 3	Achille Pattavina Xavier Lagrange
V0.88	13/04/2015	Review of Sections 2 and 3 Initial text for exec summary and conclusion	Jose Torrijos Gijon Stephane Gosselin
V0.89	13/04/2015	Clean version without revision marks. Delivered for internal reviewing.	Annie Gravey
V0.90	14/04/2015	Partial review of sections 1 and 2	Peter Olaszi
V0.91	19/04/2015	First external review	Lucian Suciu, Stephane Gosselin, Annie Gravey
V0.92	20/04/2014	Review of document	Stefan Host, Selami Siftci, Peter Olaszi, Ricardo Martinez
V0.93	22/04/2015	Clean version to discuss	Annie Gravey
V0.94	28/04/2015	Editorial modifications Review of section 5 Section 4.2.2 Section 4.3 and Appendix 2 Section 4.4 Appendix 1 Section 4.5 Section 4.2.1 Section 4.2.3	Annie Gravey Jose Torrijos Gijon Lander Alonso, Annie Gravey Xavier Lagrange, Younes Khadraoui Souheir Eido Yue Li Zhe Li Lander Alonso, Tibor Cinkler Stefan Host
V0.95	28/04/2015	Version without revision marks for reviewing	Annie Gravey
V1.0	30/04/2015	Version 1.0 for release	Annie Gravey

Distribution List

Full Name or Group	Company	Date
PMC	Public deliverable (will be made available through COMBO website)	
SC		
Other		

Table of Content

List of Tables

List of Figures

Glossary

Acronym / Abbreviations	Brief description
AAA	Authentication, Authorisation, and Accounting
AP	(Wi-Fi) Access Point
ARPU	Average Revenue Per User
BBU	BaseBand Unit
BNG	Broadband Network Gateway
CapEX	CAPital EXpenditures
CC	Cache Controller
CCN	Content Centric Networking
CDN	Content Distribution Network
CDS	Content Distribution System
CeAP	Cellular offload Access Point
CN	Core Network
CO	Central Office
CPRI	Common Public Radio Interface
DWDM	Dense Wavelength Division Multiplexing
EPC	Evolved Packet Core
eNB	E-UTRAN Node B, Evolved Node B
ePDG	evolved Packet Data Gateway
FDE	Forwarding Decision Entity

FE	Front End
FMC	Fixed-Mobile Convergence
FTTH	Fiber To The Home
GTP-U	GPRS Tunnelling Protocol (User Data Tunnelling)
HPLMN	Home Public Land Mobile Network
HSS	Home Subscriber Server
HT	Horizontal Target
IoT	Internet of Things
LGW	Local GateWay
LIPA	Local IP Access
LMP	Local Management Primitive
LTE	Long Term Evolution
M2M	Machine to Machine
MCDM	Multi Criteria Decision Making
MN	Metro Network
NFV	Network Function Virtualisation
NG-POP	Next Generation Point of Presence
NS	Network Scenario
OpEX	OPerational EXpenditures
OTT	Over The Top
PDCP	Packet Data Convergence Protocol
PDN	Packet Data Network
PGW	Packet data network GateWay
PON	Passive Optical Network

PPP	Point to Point Protocol
QoS	Quality of Service
RGW	Residential Gateway
SDN	Software Defined Network
SGW	Serving GateWay
SIPTO	Selected IP Traffic Offload
SSID	Service Set ID
TWDM-PON	Time and Wavelength Division Multiplexed PON
UAG	Universal Access Gateway
uAUT	universal subscriber and user AUThentication
UC	Use Case
UDC	User Data Convergence
uDPM	Universal Data Path Management
UDR	User Data Repository
UE	User Equipment
ULF	User Location Function
WDM	Wavelength Division Multiplexing

1 Introduction

The fifth generation of mobile technology is positioned to address the demands and business contexts of 2020 and beyond. In this regard, 5G networks will operate in a highly heterogeneous environment characterised by the existence of multiple types of access technologies, multi-layer networks, multiple types of devices, multiple types of user interactions, multiple forwarding modes, etc. Fixed Mobile Convergence (FMC) is one of the enablers that shall allow dealing with that scenario [45].

Deliverable D3.1 [1] described current architectures for fixed, Wi-Fi and mobile networks together with the roles of key equipment. Key network functions were classified in eleven functional groups, capturing the main functional areas, which are relevant for future converged networks. Based on a comparative analysis of fixed, Wi-Fi and mobile networks, deliverable D3.1 pointed out that some of the functional groups need strong efforts to reach convergence: Forwarding, Automatic Configuration and Management, Policy & Charging, Subscriber Data and Session Management, Mobility.

This functional convergence will be fostered by new architectural enablers and trends such as Network Function Virtualization and Software Defined Networking, but also by technological triggers related to advanced mobility, offloading and even connection control functions.

The use cases defined by WP2 [2] were also considered in [1] which also identified what is missing in the current state of the art to realise them. This “gap analysis” was used to derive key FMC architectural targets, as combinations of both architectural and technological concepts, aiming at providing technical solutions to these use cases. The complete methodology and derivation process that were applied are illustrated in Figure 1.

The derived FMC architectural targets identified by the “gap analysis” carried out during the first year of COMBO are of two different types:

- Horizontal Targets (HT), defined as consistent sets of FMC generic functions allowing functional convergence and solving some key “horizontal” end-user related tasks in an FMC context such as universal authentication and interface selection control. The HTs thus focus on advanced functional features of future converged networks. These functional features can be implemented differently depending on the supporting network scenario.
- Network Scenarios (NS), defined as consistent combinations of architectural and technological concepts providing technical solutions to FMC use cases and targeting structural and/or functional convergence. Network Scenarios specifically focus on the organisation and overall architecture of the FMC network.

Horizontal targets encompass functional developments, which are required whatever the future organisation of the FMC network is. This is why they have to be described and developed in details before the analysis of network scenarios is elaborated. The present D3.2 deliverable aims precisely at describing and

developing the proposals made by COMBO to solve these FMC Horizontal Targets.

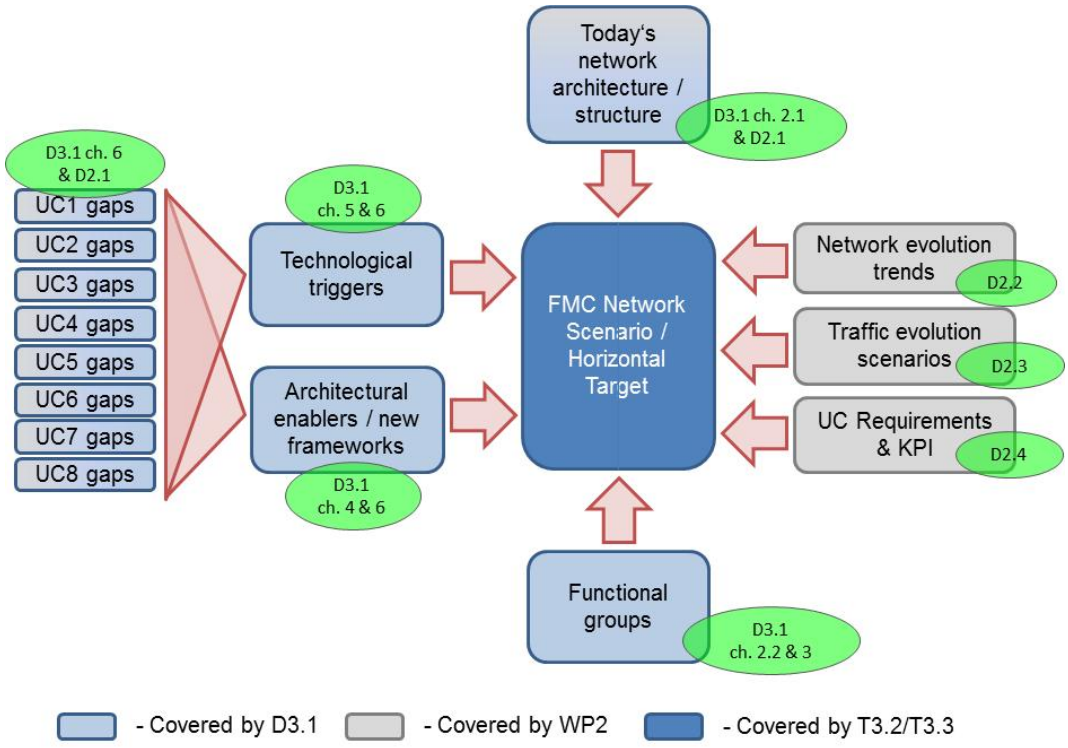


Figure 1: Methodology for FMC architecture development

So as to justify Horizontal Targets as intermediate goals for FMC, Section 2 of this document explains the motivations and expected benefits from FMC and gives a description of the overall FMC target along with the main characteristics of HTs proposed by WP3 for functional convergence. It finally describes current developments in the scope of these HTs.

Section 3 describes and develops technical solutions and implementation options proposed by COMBO for the first Horizontal Target (HT1) entitled "Converged Subscriber and Session Management". HT1 is first described in terms of problems to be solved and requirement for the solution. Then, a description of the target solution is made, which we call universal subscriber and user AUThentication (uAUT). Lastly, a progressive migration path for realizing uAUT in a realistic manner is presented.

Section 4 focuses on the second Horizontal Target (HT2) entitled "Advanced Interface Selection and Route Control". HT2 is not independent from HT1, as the network operators only provide network connections to the UEs of authenticated subscribers. HT2 is also described in terms of problems to be solved and requirement for the solution. So as to fulfil HT2, we then propose a functional block that realises a "Universal Data Path Management" (uDPM). The main functional sub-blocks of uDPM are described and analysed, including interactions between the different sub-blocks. Three specific COMBO proposals are then detailed. Two of them focuse on off-loading the LTE network, the last one focuses on the interaction between FMC network control and content distribution.

Section 5 then applies HT1 and HT2 solutions to specific FMC use cases, which have been defined in D2.1 [2]. As the two horizontal targets do not have the same impact on all use cases, Section 5 only focuses on use cases 1, 2, 4 and 8 [2], which are the most relevant from the HT1 and HT2 points of view. For each of them, it is analysed why HT1 and HT2 are needed and how uAUT and uDPM can be applied.

Section 6 concludes the document and summarizes key outcomes and achievements. It also highlights that the addressed novel technical solutions for HT1 and HT2 could be implemented and rolled out differently in FMC networks, through distributed and centralized flavours of the NG-POP concept. It thus paves the way to architectural blueprints of FMC networks leveraging on HT1 and HT2 proposed solutions. These architectural blueprints will be developed in Task 3.2 “Convergence of fixed / mobile network functions” and will be the subject of COMBO deliverable D3.5.

2 Justification of HTs as intermediate goals for FMC

This section gives a description of the overall FMC target along with the main characteristics of HTs proposed by WP3. It explains the motivations for FMC, the current developments and the expected benefits from the proposed Horizontal Targets.

2.1 Motivation of FMC

Internet traffic keeps growing rapidly as a consequence of the steadily increasing number of users and the adoption of new bandwidth-intensive services (such as video services) by these users. According to recent studies [12], global IP traffic has increased more than fivefold in the past 5 years, and will increase threefold over the next 5 years. Moreover, due to the increasing popularity of smart phones and emerging mobile applications, mobile Internet is dramatically expanding. It is predicted that Internet traffic from wireless and mobile devices will exceed traffic from wired devices by 2016 and that nearly half of Internet traffic will originate from non-PC devices by then [12]. Global mobile traffic will increase by nearly 11-folds between 2013 and 2018.

Fixed network resources do not always efficiently carry mobile traffic, as fixed and mobile network architectures are typically separated at both structural and functional levels. Mobile traffic is typically tunnelled through the fixed network, as it has to access IP networks (including the Internet) through a PGW. As mobile operators typically deploy a small number of PGWs, the path between UE and the Internet is usually much longer for a mobile subscriber than for a fixed subscriber.

The present section illustrates through a few examples the benefits brought by an FMC network architecture. It is shown that substantial bandwidth gains can be obtained by facilitating mobile data traffic offloading (Section 2.1.1), that optimised data path control can help offloading traffic from congested areas and facilitating load balancing (Section 2.1.2), that a holistic network control provides new business opportunities thanks to network sharing (Section 2.1.3), that delivered QoS can be improved by intelligent content distribution techniques (Section 2.1.4), and that FMC provides new opportunities for partnerships between telecom operators and OTT service providers (Section 2.1.5).

2.1.1 Bandwidth gain in the core and metro network

In the last few years, fixed network operators have deployed caches and data centres close to the end-users in order to efficiently serve traffic demands by limiting the bandwidth requirements on the backbone and metro networks. According to Bell Labs' paper [29], the use of distributed caches within the Metro/Aggregation and Access network segments resulted in offloading into the metro caching up to 57% of the total fixed traffic in 2012. This amount could grow to 75% by 2017.

The use of caches in the mobile network is not as simple as in the fixed network due to the tunnelling between the UE and the PGW in both directions. This tunnelling implies that, even if a video server is geographically close to a UE, the downloaded video stream has to go through the EPC in order to enter the tunnel available between the UE and the PGW, which may be far from the UE (there are currently a

few PGWs per mobile network). However, mobile data offloading approaches such as LIPA and SIPTO [40] have allowed the UEs to access the external IP network using distributed Local Gateways (LGWs) closer to the UE without traversing the EPC network as shown in Figure 2.

In order to assess the positive impact of these architectures on bandwidth demands, [30] assumes that content popularity is identical whether the subscriber uses a fixed or a mobile network, that optimal traffic control relying on SIPTO is implemented, and that session continuity is preserved (which may not be the case with the current standards as pointed out in Section 2.4). Under these assumptions, the following results are obtained:

- In 2012, the volume of traffic supported by the Core Network (CN) could have been reduced by less than 3%. This gain is limited compared to the investment (in terms of LGW and distributed PGWs deployment) that needs to be done in order to achieve it. Therefore, offloading the mobile data traffic is not essential at the present time.
- In 2017, more than 30% of CN bandwidth as well as 15% of the metro network bandwidth can be offloaded if a distributed mobile LTE architecture is considered. This is a significant gain that can justify the CapEX necessary for distributing the LTE architecture.

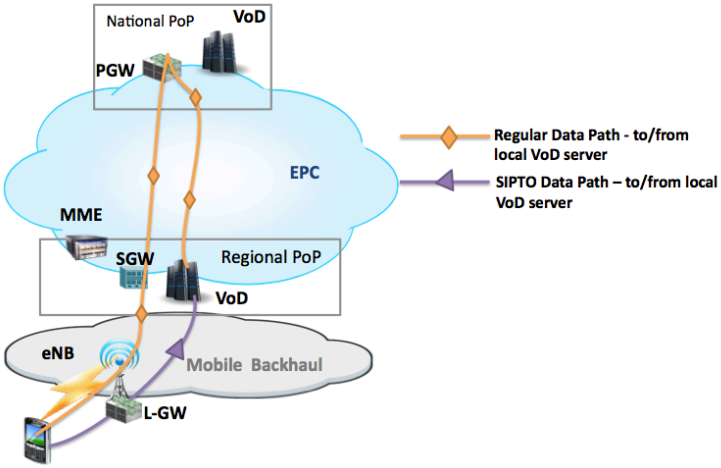


Figure 2 : Accessing a local cache with SIPTO architecture

2.1.2 Optimised traffic control

In order to face the unprecedented increase of mobile data traffic with a limited spectrum, it is proposed to increase the number of base stations and more precisely to deploy a large number of small base stations (i.e. micro, pico, femto). In addition, as mentioned by [31], both Wi-Fi and cellular technologies can be used/combined to cope with the traffic increase. In this regard, the number of Wi-Fi access points has been dramatically increased since 2002 as illustrated in Figure 3, which represents the number of Wi-Fi networks versus time since 2002.

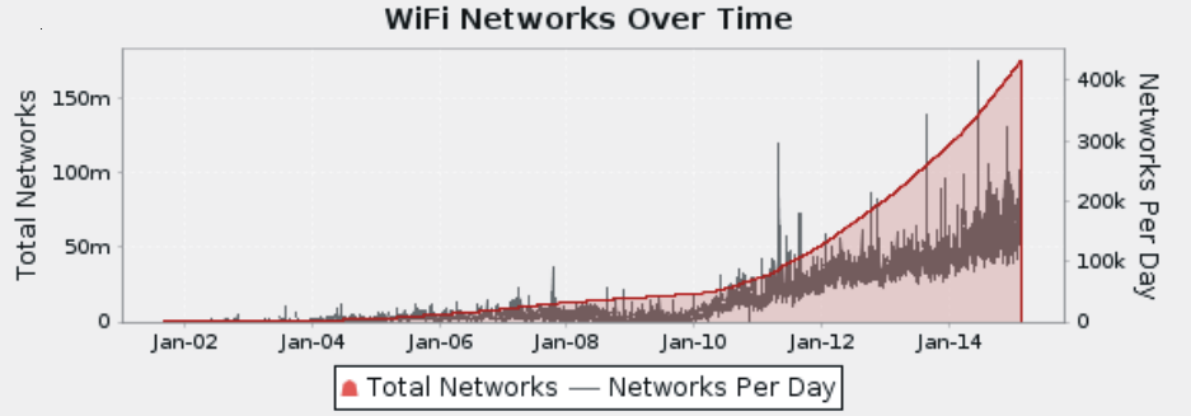


Figure 3: Evolution of the number of Wi-Fi access points (source [32])

According to [32], there are 177 million Wi-Fi networks registered by the site, whilst 4.3 million cell towers have been deployed.

Nowadays, most terminals have several interfaces (mobile and Wi-Fi). In order to benefit from the capacity offered by that huge number of access points, traffic control should be optimised aiming at achieving the following pair of functions:

- Advanced interface selection, or attachment functions (e.g. session mapping) which automatically connect the UE to physically available networks e.g. performing LTE Wi-Fi handover in both directions;
- Data path control functions, which provide the FMC network operators with optimal traffic control functionality. This function will dynamically control data paths, which may carry sub-flows, in case of universal access bonding (mobile and fixed) when available.

Control is currently performed by the operating system of the UE. This may prevent an optimal use of the available access technologies as well as preclude important features such as session continuity in case of handover between LTE and Wi-Fi. However, both aspects are essential parts and reasons for migrating towards a FMC network. In such a network, it will also be possible to handle connection control from a network unit, and thereby enable control over the usage of the terminal equipment. The traffic flow between a UE and a content server (CS) could also be split and adapted to attained path performance according to QoS requirements. It will also be possible to increase the mobility and assure session continuity throughout the network in a more natural way by an improved routing control in the network.

Advanced interface selection and route control functions should rely on criteria such as available data paths, channel properties (i.e. delay, jitter, etc.), load conditions and possible other policies (as described in Section 4). These functions have to be implemented on both sides, i.e. at both the UE and the traffic control location.

2.1.3 Network resource sharing

The increase in the number of OTT players as well as the expansion of the variety of the services offered by them push network operators to look for new strategies, creating more cost-effective networks, to reduce their costs and thus to increase their profits.

A first strategy aims at reducing CapEX and OpEX, thanks in particular to the virtualization techniques: network virtualization (NFV) and cloud based mechanisms. They provide means to increase the utilisation of network resources by applying partitioning and slicing of physical nodes and links; they also enable replacing costly dedicated hardware (e.g., router) by generic and programmable high power processing units.

A second and complementary strategy builds upon new value-added services and pricing schemes. In particular, quality-based differential pricing might be seen as promising, although it might require reshaping the whole business logic of network operators, and may be considered as going against Net Neutrality.

Another approach is based on network operators offering resource-based “bundled services” to other network operators and service providers. This could be considered as “Resources as a Service”, or “Network as a service” and can be achieved by sharing infrastructure resources in a controlled manner.

With their structural and functional convergent solutions, FMC networks present distinguished features that can be exploited in various advanced services. While the structural convergence is seen as a way of maximising the utilisation of network elements and resources (i.e., transmission, switching, forwarding, etc.), functional convergence solutions provide unified mechanisms/operations to seamlessly serve fixed and mobile users while also providing new resource-based bundled services to other network operators and content providers.

With careful resource management strategies, an FMC access network provider or an infrastructure provider could thus globally maximise the utilisation of the network resources it owns. As a result, new business models targeting new revenue channels and lower CapEX/OpEX may yield higher network average revenue per user (ARPU).

2.1.4 Content distribution

In order to reduce the amount of traffic in the network backbone, and to improve QoS for end users, caching schemes have been proposed in Content Delivery Networks (CDNs). To meet the growing content related requirements of mobile users, various technologies such as caching in the air [13] and traffic offloading [14] have emerged as potential solutions. For Content Distribution Services (CDS), a major benefit of FMC will be to facilitate the sharing of caching/storage facilities between network types.

When FMC is not implemented, the network distance between a subscriber and its requested content is always larger, as e.g. the LTE subscriber cannot take advantage of a shorter Wi-Fi network distance due to the tunnelling of data between the UE and the PGW.

A preliminary evaluation shows that FMC networks can significantly improve QoS for content distribution. The scenario simulated includes 100 mobile users who are consuming short videos coded at low bit rate (200 kb/s to 400 kb/s). UE is either connected to Wi-Fi access point (AP) or to eNB. When the UE moves, it will switch between LTE and Wi-Fi network according to the coverage of Wi-Fi APs. 8 APs and 2 eNBs are deployed over a geographical space around 2 km². The downlink bandwidth of these APs is limited to 10 Mbit/s, and the same for the mobile backhaul.

Figure 4 reports our simulation results, detailed in [52] during which the proportion of discontinued sessions is measured. The simulation shows that an FMC network without caching can already improve the percentage of smoothly streamed sessions to more than 55%, which means 55% of the videos accessed by the users in the network can be played back without any interruption. Conversely in the current network the percentage of continuity is only around 30%.

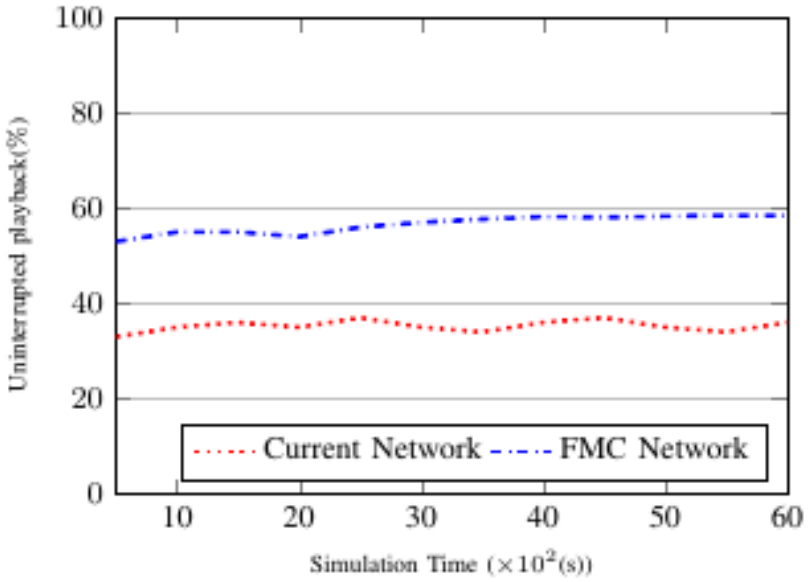


Figure 4: Video playback continuity in heterogeneous mobile network

2.1.5 FMC and OTT business opportunities

In the last decade, OTT players have fought with telecom operators about supporting the cost of increased resources requested to support OTT services.

Yet, OTTs do not compete for telecom operators service revenues; actually, they only wish to control key links in the digital value chain, with business models that span consumer electronics, online advertising, software licensing, e-commerce and more [43]. Therefore, unlike telecom operators, OTTs do not bear the burden of providing access to Internet service. According to VisionMobile [43]: “Connectivity may be as important to their business model as gas to a car; yet, it’s the telcos which supply it, not the OTTs themselves.” The asymmetry illustrated in Figure 5 makes it difficult for telecom operators to protect the profitability of some legacy business models.

On the other hand, telecom operators present assets that OTT do not have: they have an access to the subscribers with a direct billing facility and they have access to their private data, including those about their network usage. With respect to private data protection rules, the operator can help in providing a better service experience, thanks to its knowledge of the network type used, its quality at a given time, location etc. This is clearly seen when comparing the QoS of Triple Play services (i.e. television and telephony over IP) with the QoS of similar OTT services (i.e. video or telephony over the Internet). It is indeed difficult today for OTT providers to adapt distributed content to real-time variations of network conditions (mobile cell load, available bandwidth etc.).

Therefore, an FMC operator implementing a convergent subscriber data system could offer a seamless authentication service to OTT providers. As long as the user has subscribed to the OTT service, the FMC operator could make this service available on any network with any device. FMC thus offers opportunities for telecom operators to monetize the mechanisms implemented for providing the convergence of subscriber data.

This could be beneficial from an OTT service provider point of view, as the QoS delivered to its users would be improved, while a seamless access to the OTT services over all types of access networks would also be facilitated. These opportunities are also beneficial for the FMC operators since they provide their customers with seamless access to OTT services while differentiating their offers from those offered by their competitors.

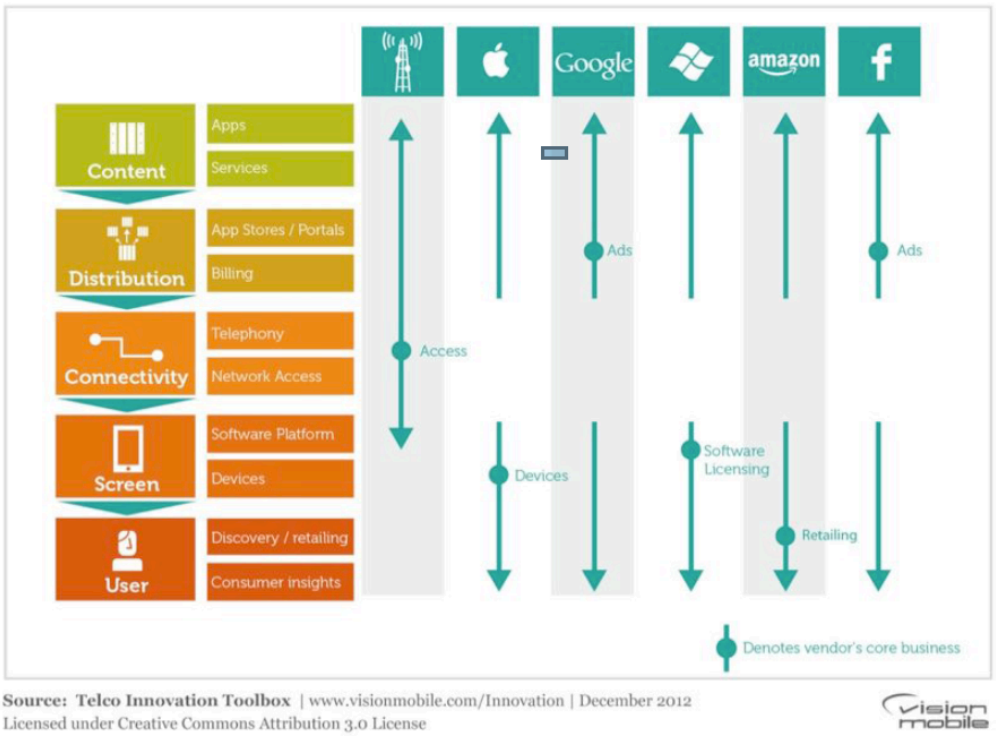


Figure 5 OTT compete with telecom operators for control, not profits

2.2 A high-level approach for reaching FMC

2.2.1 Overall goals for FMC in the communication ecosystem

In the communication ecosystem, many types of actors collaborate, interwork and compete to deliver services to subscribers:

- Infrastructure operators, which manage network and/or storage resources, or both;
- Fixed, mobile and hotspot (Wi-Fi) access network operators, which, when integrated (i.e. operate more than one type of network), can be FMC telecom operators;
- Over The Top (OTT) service providers.

Although a service control layer (such as IMS) can conceal heterogeneous network architectures and infrastructures from the users, it does not help in optimising the usage of the deployed infrastructures; an inefficient use of resources leads to highly-priced services with an unsatisfying QoS. One of the main goals of FMC is to cope with the always-increasing capacity demands in a unified way on the various access network technologies (i.e. mobile, fixed, Wi-Fi). Another and also important goal is to help the telecom operators to efficiently deal with the stagnating revenues and high cost pressure for the infrastructure as well as reducing the operational costs required for maintaining multiple/independent (technological) access networks. FMC thus potentially impacts on the users and on all the types of operators and service providers identified above.

2.2.2 Functional versus structural convergence

As explained in D3.1 [1], two types of convergence have to be addressed to reach a real integration of fixed and mobile networks:

- The convergence of fixed and mobile network functions, called **functional convergence**, is defined as the implementation of generic functions to realise similar goals in different network types (fixed, Wi-Fi, mobile). This includes replacing functions designed for a specific network type by generic functions that support all network types;
- **Structural convergence** is defined as pooling or sharing of network and infrastructure resources (cable plants, cabinets, buildings, sites, equipment, links, technologies) among several network types (fixed, mobile, Wi-Fi). It aims at defining joint fixed/mobile equipment and infrastructures for access and aggregation networks, thus allowing streamlining of broadband network infrastructures.

This document addresses the functional convergence more specifically. It aims at defining and developing consistent sets of FMC generic functions attaining functional convergence and solving some key “horizontal” end-user related tasks in an FMC context. These consistent sets of FMC generic functions are the so-called Horizontal Targets (HT).

2.2.3 Identification of Horizontal Targets

D3.1 in its Section 3 has analysed the main differences between the three network types (i.e. fixed, mobile, Wi-Fi) according to the networking functional groups. It has identified that five functional groups need strong efforts to reach the targeted convergence: Forwarding, Automatic Configuration and Management, Policy & Charging, Subscriber Data and Session Management, Mobility.

Commonalities and differences, for each functional group, between the functions operating in the various network types are identified in the second column of Table 1.

Subsequently, a “gap analysis” intending to identify what features are missing in the existing functional groups to allow FMC has been carried out. The results from the gap analysis are reported in the last column of Table 1.

Several missing features in many of the five key functional groups are similar, or at least quite related. This suggested to tentatively group them into a smaller set of two functional

blocks, called “Horizontal Targets”, as reaching these targets would represent a major step toward a true FMC. The mapping of the missing features from Table 1 on the two HTs is illustrated in Figure 6.

Table 1: Extension of Table 13 from D3.1 to identify missing features in functional groups

	Commonalities and differences	**What is missing to support FMC?**
Forwarding	Basically the same mechanisms; Mobile traffic transport via tunnelling protocols	Mechanisms sharing the interface selection control between user and network Global control of all available data paths that can be used by the sub-flow(s) of a session, whether tunnels are used, or not. Such a control would allow e.g. load balancing between data paths from different network types.
Automatic Configuration Management	Different, incompatible mechanisms and standards	A mechanism to bind a single user accessing one network through a given UE to multiple subscribers identities
Policy & Charging	Different mechanisms and standards, Activities between BBF and 3GPP to harmonize approaches	A global framework that takes into account policy rules relative to all network types, depending on the multiple subscribers identities of a single user
Subscriber Data and Session Management	Different, incompatible mechanisms and standards	A global authentication mechanism, enabling the access to multiple subscribers profiles A holistic management of a session over multiple network types
Mobility	Only implemented in mobile networks	A global support of vertical handover between multiple network types A support of content distribution over fixed and mobile networks

Five key functional groups for FMC

Two Horizontal Targets	Forwarding	Automatic Configuration Management	Policy & Charging	Subscriber Data and Session Management	Mobility
Converged subscriber and session management		Used to bind a single identity (the user) accessing one network to multiple subscribers identities (all network types)	Accesses policies specific to each user type, binds them to a single user	Global authentication, enabling access to multiple profiles Holistic management (high level) of session over multiple networks	Identifies a user as unique over several networks: - makes « vertical handover » between LTE and Wi-Fi possible - Facilitates load balancing between servers distributing content
Advanced interface selection and route control	Advanced interface selection Global control of available data paths Load balancing		Takes account of policies specific to each network type (user specific and network specific)	Applies high level session management to existing data paths	Handover extended to all components of a converged network Allows optimising server for content distribution by activating new data paths when necessary

Figure 6: Mapping of FMC functional groups to two horizontal targets

The next subsections fully describe the two HTs:

- HT1: Converged subscriber and session management;
- HT2: Advanced interface selection and route control.

2.2.3.1 HT1: Converged subscriber and session management

A network user can have multiple identities related to its subscription(s) depending on the access type, the service and the device used. Moreover, almost each user or subscriber identity is associated with a subscriber profile that is stored in a network database. These profiles host security credentials, authorisation data related to services and features that the user is allowed/denied to use and also miscellaneous data related to its current location, the device being used, etc. Regarding the authentication for the access to the network or to the service, various mechanisms are in place depending on the network type or service used, ranging from fixed lines' basic login/password to mobile networks' strong mutual authentication mechanisms (authenticating both the user and the network) with ciphering and integrity checks.

The drawback of this proliferation of identities, subscriber profiles and authentication mechanisms is that a subscriber using a given access network, different from the one corresponding to his/her subscription, is in general not seamlessly recognized in the visited network. So the user has to choose the network or the service manually and should authenticate again, with the visited network's mechanism. Moreover, adapting the service to the access network or to the device being used is thus difficult to manage or quite impossible.

HT1 addresses converged subscriber data and session management in such a way that a convergent operator or an operator working with various partners (other

network operators or OTT players) can manage all its/their subscribers in a unified way. At the network level, a novel **universal subscriber and user AUThentication** (uAUT) mechanism will provide a common subscriber authentication platform regardless of the access network. This grants the end-user with a seamless authentication for the access to any network type. Moreover such a platform will also be suitable for cooperation with OTT players for seamlessly authenticating the user when he/she is accessing the OTT service.

HT1 also addresses the unification of subscriber data management in such a way that all subscriber and user profiles are either physically hosted in a logically unique subscriber database or present a single interface to the convergent operator. Thus, the uAUT server would be the entry point to the set of subscriber profiles to be used by the FMC operator. The operator will then be able to provide service level adaptation information based on access network type and capabilities.

uAUT is part of the control plane, and relies on data from the management plane.

2.2.3.2 HT2: Advanced interface selection and route control

LTE routes data traffic by default on the following three data paths:

- the first one between the User Equipment (UE) and the eNB, carried over the Packet Data Convergence Protocol (PDCP),
- the second one, a tunnel between the eNB and a Serving Gateway (SGW),
- the third one, another tunnel between the SGW and a Packet Data Network Gateway (PGW).

The default data path has been designed in order to facilitate AAA in the mobile network and to ensure session continuity in case of UE mobility.

The currently deployed LTE networks typically centralise the SGW and PGW functions into a small number of locations, which may lead to inefficient routing, by “tromboning” mobile traffic through the PGW. Distributing the SGW and PGW functions over a larger number of locations could limit this drawback.

Moreover, a mobile UE currently presents several interfaces (LTE and Wi-Fi) and can potentially take advantage of alternate connections provided by the fixed network to access Internet e.g. through a Residential Gateway (RGW) or a LGW. However, when a session is moved from one interface to another, session continuity may not be maintained as each interface is usually reached through a specific IP address. The decision to use an interface or another is mostly left to the UE. This precludes the network to take part in the decision, which would help in some cases, e.g. implementing load balancing strategies or avoiding congested data paths.

The motivation for HT2 is to provide the FMC network operator with new means to participate, together with the UE, in the dynamic control of the data paths used by mobile traffic, when several such paths are available, while maintaining session continuity whenever necessary.

Such data path control would blend network-driven “traffic offloads” (from the mobile network to the fixed or Wi-Fi network) to user-initiated offload (when a user selects its Wi-Fi interface instead of selecting its mobile interface, without the network operator’s

intervention). A novel **universal Data Path Management** (uDPM) will allow the FMC operator to take advantage of path diversity in order to implement traffic and QoS control policies and perform appealing traffic engineering tasks such as load balancing, congestion avoidance, energy consumption limitation, etc.

uDPM presents functions from both the control and the data plane, and relies on data from the management plane.

2.2.3.3 Approaches for fulfilling HT1 and HT2

Implementing uAUT and uDPM implies finding alternatives for the current implementations of the five functional groups as identified in Figure 6.

Sections 3 and 4 will describe and develop the high level proposals made by COMBO to achieve both HT1 and HT2 through the design of uAUT and uDPM functional blocks.

Note that there are possibly multiple methods for implementing both HT1 and HT2 technical solutions, which could be devised and adopted so as to deploy the targeted functional FMC solution. In particular, how much the functions should be distributed between the UE and the network elements needs to be addressed in a later deliverable D3.6.

We have previously shown that functional network convergence takes benefit from the Next Generation Point of Presence (NG-POP) concept [1]. The NG-POP is a location in the network, where the operator implements multiple functions, including the IP edge for all network types. Task 3.2 will show in D3.5 how the technical solutions for horizontal targets could be organized and implemented in actual 5G networks considering two alternative scenarios.

One alternative is a centralised solution, with a small number of NG-POP locations, typically at the sites of core Central Offices (COs), which are the edges of the current fixed aggregation network. The other alternative relies on a larger number of NG-POP locations, located in the current main COs; this location corresponds to an extension of the IP backbone towards the access network. In both cases, the advantages brought by the SDN/NFV shall be assessed. Specifically, this development of two alternative network scenarios for functional convergence will be the subject of deliverable D3.5.

2.3 State Of the Art for HT1

The present Section describes the main points in the state of the art of subscriber data convergence between fixed broadband, mobile (3GPP) and Wi-Fi hotspot networks, in order to clarify which innovations are brought by the proposal made by COMBO in Section 3.

The convergence or unification of subscriber data and session management is not a new topic and some partial solutions, listed below, are already available, either commercially or in standards.

2.3.1 Subscriber data convergence for mobile traffic offload in Wi-Fi networks

In order to offload mobile networks, operators have deployed Wi-Fi networks. An interaction between network elements dealing with subscriber data in both networks is thus required. Indeed, the visited Wi-Fi network should check the mobile identity and apply the right rules according to the offloading type. We can identify two main types:

- Mobile radio and core network offloading: after a strong SIM-based authentication, all user traffic is offloaded through the Wi-Fi network without crossing the mobile core network. This mechanism is also known as "*local WLAN break-out*".
- Mobile radio only offloading: the traffic uses the Wi-Fi radio network but crosses the mobile core network. The mobile operator then keeps control of all user traffic so that functions like deep packet inspection or lawful interception are possible. This is defined by the 3GPP as non-3GPP access architectures. The entry point of the mobile core network is then either directly the Packet Data Network (PDN) Gateway in a trusted mode, or the evolved Packet Data Gateway (ePDG) in an untrusted mode, with which a secured tunnel is in place.

In both cases, the Home Subscriber Server (HSS) provides derived keys to an Authorisation, Authentication and Accounting (AAA) server, which interacts with the AAA client of the Wi-Fi operator. No request is sent to the HSS for security reasons. Figure 7 shows the network elements involved in non-3GPP accesses to a Home Public Land Mobile Network (HPLMN).

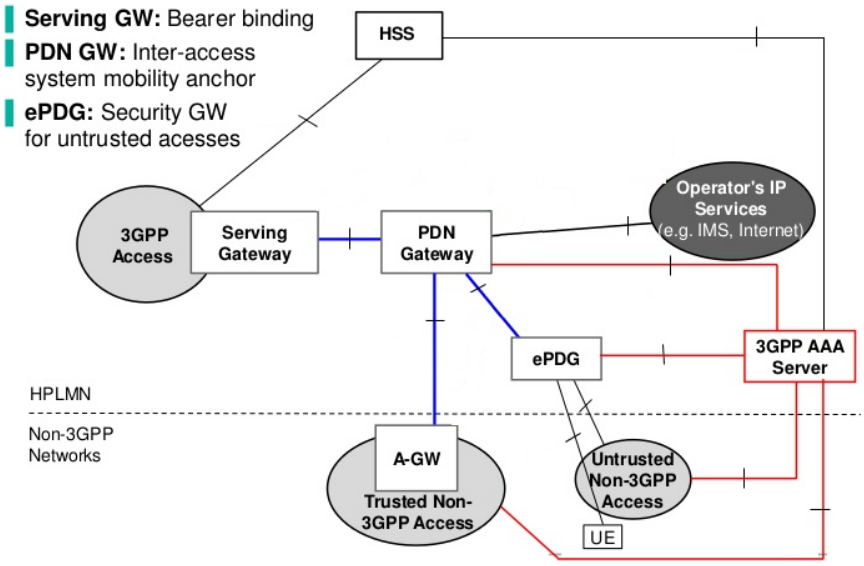


Figure 7: LTE architecture for non-3GPP access

On the Wi-Fi industry side, the Wireless Broadband Alliance (WBA), in 2010, started to define a set of Wi-Fi Alliance (WFA) standards called Hotspot 2.0, which attempt to bring 3G-like end-user experience to Wi-Fi authentication and roaming.

For SIM-based authentication methods two main mechanisms are available:

- EAP-SIM (IETF RFC 4186) for 2G SIM cards.

- EAP-AKA (IETF RFC 4187) for 3G USIM cards. There is also a variant of EAP-AKA for LTE.

These standard initiatives lead to subscriber convergence in the sense that there are interactions between mobile and Wi-Fi subscriber databases, but there is no global view of a given user's identities for the benefit of a single operator that would be managing both types of networks.

Some commercial mobile data offload over Wi-Fi Access Networks are already available. For example, Aptilo Mobile Data Offloading in trusted and untrusted 3GPP Wi-Fi access [50]. Another example is provided by Orange Romania, which provides seamless public Wi-Fi with Cisco Hotspot 2.0. The service enables Orange Romania users to move between cellular and Wi-Fi networks with seamless authentication, combining a cellular-like roaming experience with the capacity of Wi-Fi, typically in public buildings, venues and within enterprises [51].

2.3.2 Fixed networks and community Wi-Fi

In many countries, several "*community Wi-Fi*" services are available. For a fixed subscriber, such a service consists in opening the access of its own Internet connection to nearby visitors. Technically, the access point broadcasts an additional Service Set ID (SSID) that is specific to the community network. The benefit is that this subscriber is also able to access to the Internet through any other Wi-Fi access point that is a part of the same community Wi-Fi service.

While small communities of users first initiated this kind of service, different operators adopted the concept over time. In particular, Wi-Fi only operators (such as Fon) and many fixed broadband operators propose this type of service to their customers. This is not a true convergence between fixed and Wi-Fi networks as Wi-Fi connections here are all related to a fixed line. However, some operators (such as Orange in France) have extended the community Wi-Fi to public hotspots managed by the operator. Fixed subscribers can thus access such hotspots with their residential credentials (login/password). In this case the authentication mechanism relies on a captive web portal accessible from an open SSID (without protection) Wi-Fi access point.

2.3.3 User Data Convergence

The most developed initiative on subscriber data unification is probably "User Data Convergence" (UDC) specified by the 3GPP [22]. UDC is the logical representation of a layered architecture that separates user data from application logic in such a way that user data is stored in a logically unique data repository, the User Data Repository (UDR). This database is generally distributed in different locations and duplicated by means of active replications for redundancy reasons. Thus, in case of outage of an instance of the data repository, all user data will be preserved.

Dedicated entities handling application logic, named "application front-ends" (FE) provide the links between the user database and the network elements that need to access user data including:

- the legacy 2G/3G Home Location Register;

- the 4G Home Subscriber Server;
- the IMS Home Subscriber server: e.g. for managing the Voice over LTE service;
- Mobile Number Portability: for keeping the same public number (a.k.a. MSISDN) when the user changes the operator;
- Equipment Identity Register: for restricting access to black-listed devices (stolen devices or non-compliant devices);
- Policy Control and Relay function.

Figure 8 illustrates the UDC concept with the replicated UDR database and the FEs.

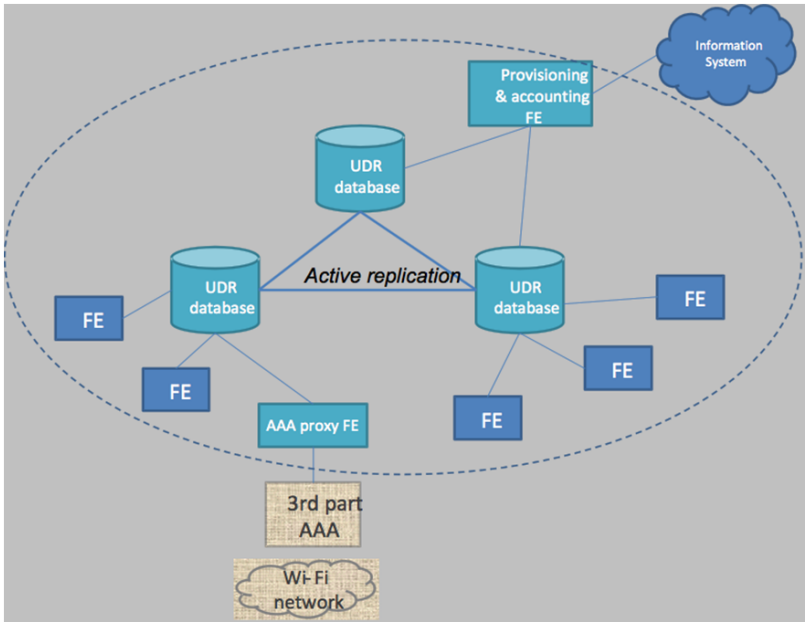


Figure 8: Schematic representation of the UDC concept

Several FE instances can be deployed for the same network function (e.g. the Home Location Register – HLR), while the HLR is seen by the other network elements as a unique logic network element. The scalability of network signalling is ensured by adding additional FEs or by providing the FEs with more processing capabilities.

The provisioning of users in the database can be done from a dedicated FE. Moreover, for addressing users who connect through non-3GPP accesses, a specific FE can act as a 3GPP AAA server and communicates with a AAA client that is in charge of authenticating users behind a third party (e.g. Wi-Fi network).

UDC implementations are already available in commercial products: NT HLR from Nokia Networks [56], ZXUN USPP from ZTE [57] or SDM from Alcatel-Lucent [58]. All implementations claim to be compliant with the 3GPP UDC standard.

Currently, the standard only specifies the UDC the separation between the backend (the UDR subscriber database) and the Front Ends (FEs) (the applications communicating with the network elements). Neither the data model nor the management of heterogeneous subscriber profiles are yet specified.

Moreover, UDC as defined by the 3GPP focuses on mobile subscribers only. In particular, 3GPP does not distinguish a “user” from a “subscriber” whereas in fixed networks for example, the users behind a Home Gateway are not necessarily identical to the subscriber of the line (the entity who or which pays for it). In this framework, different users related to the same subscription may have different rights according to their own user profile.

UDC is an interesting approach for unifying subscriber data and session management in the sense that it separates a logically unique subscriber/user database from independent application FEs.

In Section 3, we extend the UDC concept to provide solutions for HT1.

2.4 State Of the Art for HT2

The present Section describes the main points in the state of the art related to interface selection and route control, in order to clarify which innovations are brought by COMBO in Section 4.

Today most terminal devices have multiple interfaces (i.e. Wi-Fi, cellular and sometimes Ethernet). As we assume that all applications are IP-based, they can transparently use any interface. Moreover, there are potentially multiple paths available to transport the information between a terminal and servers in the network, depending on the network type. Furthermore, for content distribution services, a content requested by the user may be stored in multiple locations, which implies that many alternative paths can be used to deliver the requested content.

Using several paths, either sequentially or simultaneously, to satisfy a user’s communication request involves several functional blocks, which are detailed in the following subsections. The first one corresponds to the decision of actually selecting a given data path (“decision process” described in Section 2.4.1); a second one corresponds to the activation or de-activation of the data path(s) (“creation/destruction” of data paths addressed in Section 2.4.2), a third one to the coordination of data carried over several data paths (“data path coordination” explained in Section 2.4.3), and the last one to the actual sending of the packets on the selected data paths (“session mapping execution”, Section 2.4.4).

Identifying separated functional blocks within HT2 allows classifying various aspects of the State of the Art. It also helps in specifying the chained functions used to realize a given traffic offload strategy. As an illustration, Figure 9 represents the call flow corresponding to the activation of SIPTO [40] for a specific session. In this particular example, there is no need for the block “data path coordination”.

Although the full process is not always decomposed as above (and as in Section 4), it is possible to classify the state of the art regarding HT2 along these blocks.

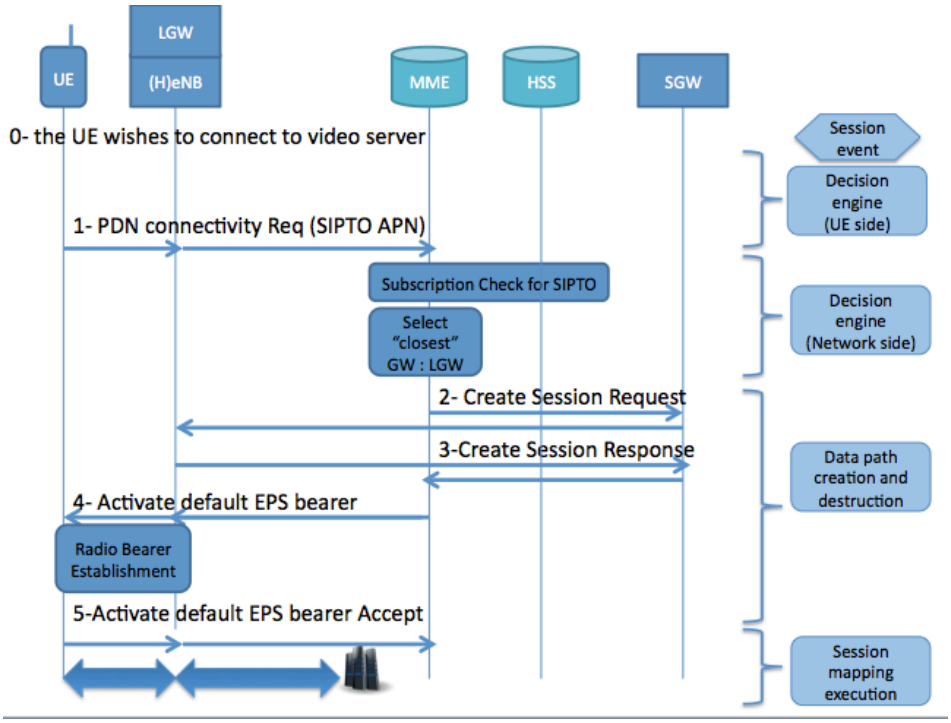


Figure 9: Activation of a SIPTO connection

2.4.1 Decision process

The decision process is relative to several types of decisions:

1. access selection, whenever there are several available paths between a UE and a PDN;
2. mobility management, which controls how a UE moves from one eNB (or access point) to another;
3. lastly, when the service requested by the user requires some type of content distribution (e.g. VoD), the decision process includes the selection of the content repository.

Concerning the first decision type, access selection, all current solutions regarding access choice are host-controlled. This means that the terminal autonomously decides to move from one access to another, or to set up additional connections in case of multipath.

However, it is possible to use mechanisms such as IEEE.11v [53], Access Network Discovery and Selection Function (ANDSF) [28] or HotSpot 2.0 [38] in order to help the terminal to discover new networks or to provide it with a set of policy rules to apply for selecting the right access. Policy rules could depend on the user profile, which can be accessed as part of authentication process. They can also take into account the state of the network (e.g. load of the network), which should thus be broadcasted to the terminal. This process can potentially generate a lot of signalling messages and thus result in an extra latency, which limits the possibilities for optimisation. In particular, ANDSF is not suitable for frequent/dynamic updates of policy and cannot be used to steer access selection in real time. Furthermore, the

user equipment would still be responsible for applying the policy and for deciding which interface to use. At the moment, both implementations of ANDSF and HotSpot 2.0 in UEs are such that they do not allow the operators to fully control the UE behaviour. In Section 4, we assume that this is not the case, and that the network operator can, in some cases, fully control the UE.

Regarding the second decision type, mobility management, the equipment that triggers it depends both on the type of systems and the activity. In LTE, a terminal with no active transmission autonomously chooses the eNB to be connected to. When a transmission is active (or was active typically during the previous ten seconds), the terminal is under the control of the network but assists it by providing quality measurements. The handover is thus mobile-assisted and network-triggered. In Wi-Fi, it is always the terminal that chooses the access point it is connected to.

Concerning the last decision type, selection of the appropriate repository, content distribution is usually controlled above the IP layer. For example, in a traditional CDN centric scenario, there may be several servers, from which the content can be obtained and the selection of the appropriate server is part of the CDN's logic.

2.4.2 Data path creation/destruction

When there is no mobility, data paths are implicitly activated through IP address allocation by the network operator. As the IP address is allocated to the interface device between the subscriber and the network, a given user can access its services through different IP addresses if he/she subscribes to different networks.

For example, a user of the mobile network can potentially receive traffic from a given service over three different paths:

- the default mobile path going through eNB, SGW and PGW;
- a path using a Wi-Fi access;
- a path authorized by the mobile architecture and taking advantage of LIPA and/or SIPTO [40] to avoid the default path going through the SGW and the PGW.

All three types of data paths are activated through standard procedures applied once an IP address is allocated to the user.

This is also the case when mobility mandates a relocation of the UE, the activation/deactivation process of data paths is fully standardised [22]: activation of a tunnel for temporarily transferring data to the UE, activation of a new data path from the target eNB and the (potentially new) SGW, and, when the UE is fully attached to the target eNB, deactivation of the tunnel and of the old data path

The policies governing data path creation and destruction therefore depend on standard-based procedures although the network is responsible for selecting the triggers.

In the case of Wi-Fi access networks, the UE's connection manager controls data path creation. That is the case for both SIM-enabled and nonSIM-enabled devices. The interface can be switched on and off completely manually and users are used to performing this regularly when they need to.

When the Wi-Fi interface is enabled, the device periodically scans for available networks and connects to those that have been previously been registered. This procedure is the only one that creates a new data path. The user performs data path destruction by any of the following three actions:

- moves away from the coverage area (the terminal detects that the path is no more present and declares it as lost);
- manually connects to another network in the area;
- switches the interface off.

The local mobility between APs of the same network is managed in the standard and it is also part of the connection manager.

ANDSF and Hotspot 2.0 partly automate the network discovery part and the data path creation but they do not influence the data path destruction. However the network does not have a complete control of the process since there are several evaluations performed by the device prior to the data path creation.

In case of a content service, the creation of the path leg to the appropriate server in case of a content service relies on CDS specific procedures that point the user towards the appropriate server using either an HTTP redirection-based or a DNS-based process.

2.4.3 Coordination of Data Paths

A typical characteristic of an FMC network is that multiple paths can be used, depending on the chosen access network, between a user and a site on the Internet. Content distribution services also natively involve multiple paths to deliver a given content to a user; this is due to the caching policies and content replication policies managed by the CDS operators.

Whenever it is necessary to change the path during a session (due e.g. to mobility), or to simultaneously use several paths during a session, some procedures are requested to coordinate the emission/reception processes on these multiple paths. The methods available for this coordination are grouped within the “Coordination of Data Paths” functional block, which is part of the control plane.

An extensive state of the art on mobility and multi-homing was published in [39]. We use here a similar approach to classify the different architectures and protocols and extend the analysis to both fixed and mobile networks. Figure 10 lists the different solutions proposed for mobility and multi-homing support and classifies them according to the features they provide and their location in the OSI protocol stack. Solutions that have not received much interest during the last five years are listed in small characters. We indicate the most popular solutions in bold characters and underline the solutions that are already standardised by either the IETF or the 3GPP.

We distinguish between site multi-homing and host multi-homing. In site multi-homing, the different network portions are called sites (e.g. the edge network and the core network) and the solutions allow a user to be reachable through different sites (e.g. different core networks). In host multi-homing, it is the user terminal that can

simultaneously use several interfaces or have several addresses without impact on upper layers, i.e. the application still sees only one connection.

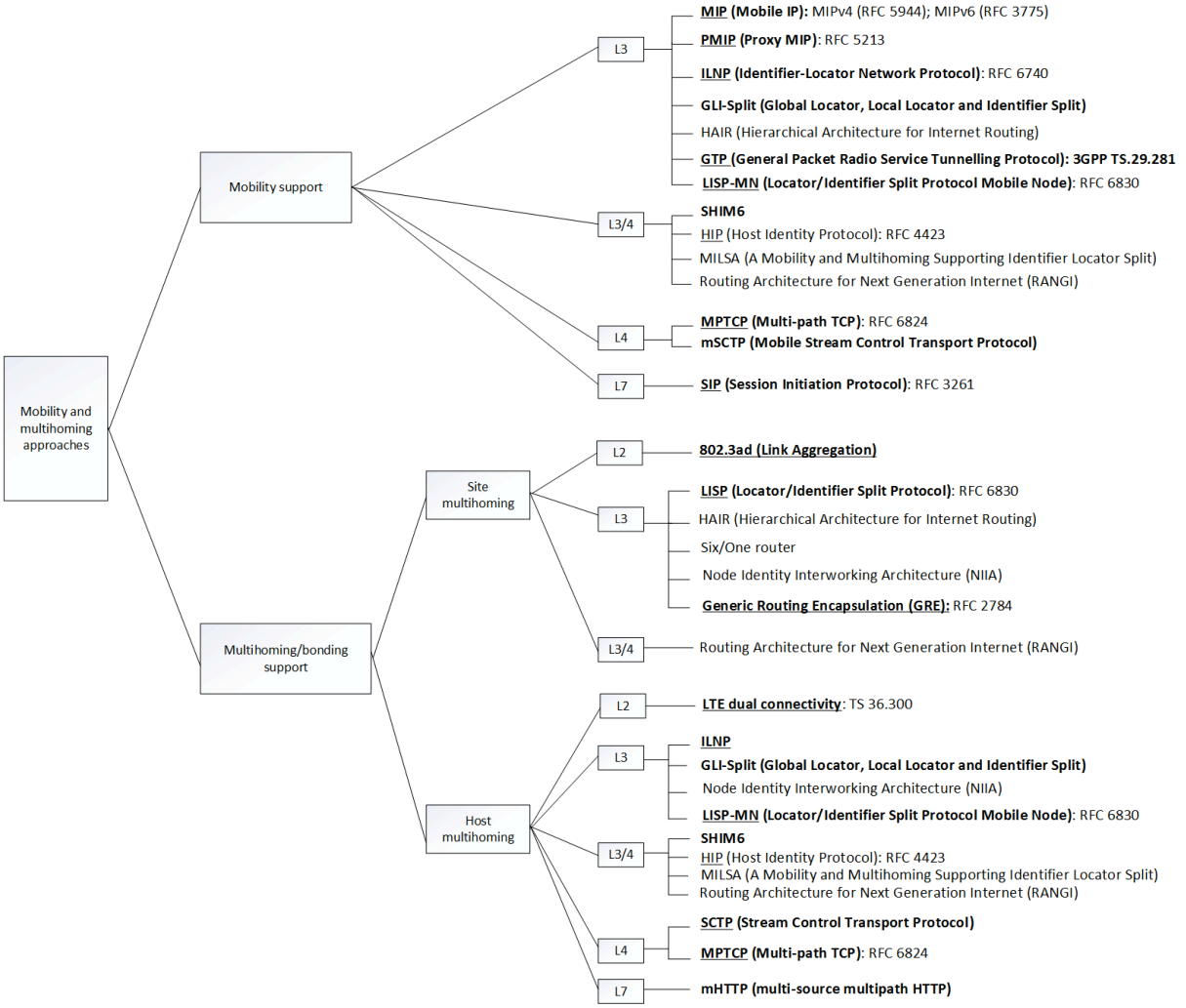


Figure 10: Mobility and multi-homing/bonding approaches

The possibility to use several interfaces at the same time is referred in the following as “bonding”[1].

At layer 2, only bonding is provided. Solutions are also called trunking or bundling. They are generally restricted to site multi-homing because current terminals are generally not designed to allow layer 2 bonding. One of the solutions was standardised by the 3GPP and is currently used for LTE dual connectivity. It allows a terminal to simultaneously use the radio resources provided by two eNBs. It is used for dual connectivity between standard eNBs as well as between a macro and a small eNB. A dual connectivity between LTE and Wi-Fi is also currently under study.

At layer 3, different solutions were proposed to provide either mobility or multi-homing and even both features for some of them. Some solutions are based on the

[1] Bonding at layer 3 is also known as multi-homing

identifier/location split paradigm. In current architectures, an IP address is used to identify a session and to locate a user in the network at the same time. An identifier/location split-solution usually proposes to have an identifier used by the transport layer to identify the session and a locator used for location. The identifier remains unchanged until the end of the session while the locator can change if the user moves to another network. Some of layer 3 solutions are implemented as sub-layers while others propose to change the network architecture. On the other hand, layer 4 solutions are more end-to-end and are independent from the network architecture. Some of them propose new protocols while others propose only sub-layers. Some application-level solutions provide mobility support and attempt to support multi-homing. In Table 2, we analyse the approaches with different criteria [4][5][33][34][35][36][37].

Table 2: Classification of mobility and multi-homing/bonding solutions

OSI layer	Criteria / solution	Modification of the host protocol stack	Modification of the network architecture	Transparency to current applications	Transparency to network elements	Control entity (host or network)	Mobility area
Layer 2	802.3ad	Yes	No	Yes	Yes	Host	/
Layer 3	MIP	Yes	Yes	Yes	No	Host	Full
	PMIP	No	Yes	Yes	No	Host	Local
	ILNP	Yes	No	No	No	Host	/
	GLI-Split	Yes	Yes	Yes	No	Host	Full (using MIP)
	NIIA	Yes	Yes	No	No	Host	Full
	LISP	No	Yes	Yes	No	Host	/
	HAIR	Yes	Yes	Yes	No	Host	Full
	Six/One router	No	Yes	Yes	No	Host	/
Layer 3/4	SHIM6	Yes	No	Yes	No	Host	Full
	HIP	Yes	Yes	No	No	Host	Full
	MILSA	Yes	Yes	No	No	Host	Full
Layer 4	MPTCP	Yes	No	Yes	No	Host	Full
	SCTP	Yes	No	No	No	Host	/
	mSCTP	Yes	No	No	No	Host	Full
Layer 7	SIP	No	No	/	Yes	Host	Full
	mHTTP	No	No	/	Yes	Host	/

There are thus multiple solutions, however, if one looks for solutions that do not necessitate to modify the network architecture and that are transparent to current applications, the number of available solutions is drastically reduced:

- 802.3ad (link aggregation)
- SHIM6
- MPTCP

All the other solutions require a “clean slate approach”. We have chosen in COMBO to (at least initially) focus on “evolutionary” scenarios, which restricts us to the above

three solutions. Moreover, SHIM6 operates only in an IPv6 network, whereas most networks still operated in IPv4. This restricts the number of solutions to two. In the innovative approaches proposed by COMBO in Section 4 to address HT2, we mostly focus on these two solutions (see Sections 4.2.2.2, 4.3 and 4.4), although a “clean slate” approach is envisaged in Section 4.5.

The objective of most of the solutions reported in Table 2 is to define protocols; the proposed solutions do not take into consideration the architecture of the access/aggregation network. More efficient solutions could possibly be proposed if both the architecture and the protocol point of view are taken into account. This is the objective of COMBO.

2.4.4 Session mapping execution

In a host-based context, the sockets opened for the application mandate data forwarding rules within the IP layer. Layer 2 connection then fully determines the data path(s) on which a given session is to be mapped. Each network type relies on some specific layer 2 protocols (e.g. PPP for fixed networks, PDCP for the RAN, GTP-U for LTE traffic). User plane forwarding is specified for each access network, identifying which equipment is responsible for encapsulating and forwarding user data traffic.

When a UE is connected to both 3GPP and non-3GPP accesses at the same time, the ANDSF can provide rules where one type of traffic is routed on one interface and another type on another interface. This is achieved through the provisioning of Inter-System Routing Policies (ISRPs) to the device. These policies are configurable and overall they indicate the following:

- under which conditions do the policies apply;
- the conditions that the traffic needs to meet (known as filters);
- the data path that the traffic needs to follow when the above conditions are satisfied.

Session mapping execution is a functional block of the data plane. It applies the decisions taken by the data path creation and destruction block, which is part of the control plane. Therefore, session mapping execution is the direct application of static rules that are not changed in real time.

2.5 Conclusion of Section 2

A few applications of FMC have been briefly outlined, ranging from an optimal usage of radio frequencies to potential network sharing scenario.

Section 2 has then identified two sets of functional blocks, which are named “Horizontal Targets” in this deliverable, as necessary targets to reach a true FMC network:

- HT1: Converged subscriber and session management;
- HT2: Advanced interface selection and route control.

HT1 and HT2 have first been shown to solve most of the gaps existing between the current non-converged situation and a situation where an FMC operator can take

advantage of a global control of its resources and a global management of its subscribers. Lastly, Section 2 presents a (non-exhaustive) state of the art for both HT1 and HT2.

Regarding HT1, two main approaches have been identified within 3GPP:

- The UDC concept specified by 3GPP [22], which separates user data from application logic. This is achieved in such a way that user data is stored in a logically unique repository allowing access by multiple applications.
- Secured Wi-Fi offload for mobile communications, either pure “WLAN break-out” in which user data traffic does not cross the EPC, or “mobile radio offloading” which is allowed by 3GPP in the framework of “non-3GPPP access architecture”.

In both cases, the HLR/HSS provides derived keys to AAA process, which interacts with the RADIUS server of the Wi-Fi operator.

Another approach to be considered is the WBA developed Hotspot 2.0, which is intended to bring mobile 3G-like end-user experience to Wi-Fi authentication and roaming.

All these initiatives lead to subscriber convergence in the sense that there are interactions between different subscriber databases, but there is no global view of a given user’s identities for the benefit of a single FMC operator. Additionally, a convergence between community Wi-Fi and public Wi-Fi hotspots has been noticed, with captive portal authentication forms enabling fixed subscribers to use such Wi-Fi hotspots. However, this type of solutions may not meet the security requirements expected from the mobile network.

Regarding HT2, related frameworks are multiple, as they build on existing mobility management (whether IP-based or LTE controlled), and/or on multi-homing management. In both cases, an important feature is session continuity as all mechanisms potentially involve either interruptions (when moving data from one data path to another), or de-synchronization (when using multiple addresses).

For both HT1 and HT2, existing frameworks (UDC [22], ANDSF [28], HotSpot 2.0 [38], LIPA and SIPTO [40], MPTCP [54]) are re-used as much as possible , in the following Sections 3 and 4 to build complete and realistic solutions to these horizontal targets.

3 Description and Analysis of HT1

The convergence of subscriber and session management (HT1) is a major objective of FMC. An FMC operator, having a unified management of subscriber and user profile can potentially provide a seamless customer experience for the use of various network types and services. This will be done notably through the unification of authentication across heterogeneous network types.

The present section first describes HT1 in terms of the problems to be solved and identifies the requirements for the proposed solution. Then, a description of the target technical solutions proposed by COMBO is provided. Lastly, a progressive migration path for realising these solutions in a realistic manner is presented.

3.1 Description of HT1

HT1 develops solutions for two levels of convergence:

- the convergence of subscriber data across heterogeneous networks;
- the convergence of authentication management between these networks.

These convergence levels are related to fixed, mobile and Wi-Fi network operators, but also to OTT players, in order to address business opportunities as highlighted in Section 2.1.5.

The developed solutions will support “hybrid access” facilities, in which a user authenticates on a given access and is then transparently allowed to use access networks and/or services associated to other network types or service providers. This shall be made possible by linking all subscriber data related to a given user; access to this global set of subscriber data could be limited by explicit agreements between network operators and service providers.

3.1.1 Subscriber versus user

A network user can have different “identities” related to its (multiple) subscription(s), depending on the access type, the service and the device used. These identities are classified into: mobile identities (i.e. public phone number, private and temporary identities), fixed identities (i.e. fixed line number, line ID), other identities related to ISP’s services (such as email account, TV account, customer portal etc.) and OTT logins (at least one for each service provider). This presents several drawbacks, as pointed out in Section 2.2.3 in the definition of HT1.

Let us first clarify the difference between a “subscriber” and a “user”, as it impacts on how to deal with identities and subscriptions in HT1.

- A user is a physical person or a machine using the access/service at a given time.
- *“A Subscriber is an entity (associated with one or more users) that has subscribed a service with a provider. The subscriber is allowed to subscribe and unsubscribe services, to register a user or a list of users authorised to enjoy these services, and also to set the limits relative to the use that associated users make of these services”.* [55]

The user may differ from the subscriber; for example a mobile user owning the subscription is both the user and the subscriber of the mobile service. But in the case of a fixed broadband access in a house, the subscriber for the fixed services can be one of the parents of the family, whereas the children are users of the services.

In the scope of HT1, a subscriber is a person who owns the mobile, fixed, Wi-Fi or OTT service subscription and who pays for the line or the account.

3.1.2 User data consolidation

Users' rights may differ depending on what the subscriber allows them to do (e.g. parental control). This leads to complex subscriber and user profiles to be managed by a service provider.

As pointed out previously, true convergence requires a global view of a given user's identities. It may be neither realistic (e.g. for security reasons) nor desirable (e.g. when providing a partial access to user's credentials to an OTT provider) to host all user data in a single user-centric database.

The issue is thus to design an entity that can access all user databases to satisfy the requirements imposed by a converged network operator. These requirements include both the consultation and the modification of user data, which may have to be modified.

3.1.3 Authentication functions

Each network type uses specific mechanisms to authenticate users and to secure their communications and services. This has been detailed in [1]. Indeed, whereas it is difficult to intercept wireline data traffic in copper or in fibre, the air is considered un-trusted and mobile operators need to deploy mechanisms to explicitly secure mobile data traffic. Furthermore, as all networks will not evolve at the same speed, legacy authentication mechanisms should still be supported even in 5G networks, which may have to interwork with legacy network technologies.

Therefore, in order to solve HT1, it is not proposed to define a unique authentication method, but rather to propose a solution for managing various authentication mechanisms in a unified way, in each case with the required security level.

The technical solution for HT1 in FMC networks is developed in Section 3.2 as a "universal subscriber and user authentication" (uAUT). It relies on a common layer 2 mechanism for carrying authentication messages, on top of which various security mechanisms are available. The solution also allows the use of authentication mechanisms based on Web technologies in order to easily support collaboration with both OTT players and with Wi-Fi access points that only use Web portals.

3.2 Global HT1 target

The present Section describes the target technical solution proposed by COMBO, with its expected functionalities, to be achieved in HT1, as the ultimate solution for HT1 in FMC networks. A user and credential structure is first introduced. Then, a high level overview of the technical solution is given, providing another layer of detail, by presenting the different modules and protocols involved.

3.2.1 Subscriber, user and credential scheme

The following user and credential structure is proposed. This structure is the one that will be present in the single database entity that will be used for authentication purposes for all types of access networks.

In the proposed structure, the base node is the subscriber, who holds the legal relationship with the operator through a service contract. Within each subscriber account it will be possible to hold many user accounts or nodes, each of them independently identifying a unique user for the network. Each user account may present several credentials available for authentication, each of these credentials being related to a given network access type.

Separate user accounts will facilitate accounting for billing purposes. On the other hand, a similar QoS-oriented profiling system will be available for controlling the QoS of users along time. Certain services or access accounts that are not linked to a mobile user directly and are shared among the users of the same subscriber will be covered as an independent user account that depends directly on the subscriber node. This allows applying independent billing and QoS profiles.

In relation to operator services, the authorisation for each user is performed in a per user basis. This allows the FMC operator to grant access to services independently for each user in order e.g. to enable individual billing.

The uAUT in itself will not be in charge of all authorisation processes in the network, but will hold the relation of the services that each user is authorised for as permanent storage. This is illustrated in Figure 11 on an example of a family with a mother (ALICE) and a child (BOB). ALICE holds the contract with the FMC operator and is charged for the services consumed by the whole family. So ALICE is registered as the subscriber and she is also one of the users of this subscription. In general, her user profiles (in green in the figure) will be less limited than other users, as she pays the bill (highlighted by the green dotted line in the figure).

ALICE and BOB are both users of the services subscribed by ALICE. Each of them has a user account with dedicated credentials for his/her mobile and Wi-Fi hotspot usage. Technically, the ALICE's user profile is linked to other sub-profiles related to each network type and service: a Wi-Fi hotspot profile, mobile profiles (HLR/HSS) and IMS profiles for particular mobile services such as Rich Communication Suite (an enriched SMS/MMS service) or Voice over LTE. BOB (in red in the figure) has similar profiles and sub-profiles but with rights that are more limited than ALICE's. Regarding the fixed broadband line of the home, each of the users can access the Internet and can make VoIP calls but with different rights according to his/her profiles. For example ALICE can make any voice call whereas BOB can make calls to fixed numbers only. Likewise, ALICE has an unlimited Internet access, whereas BOB's profile imposes a parental protection with web sites to be blocked. When BOB connects to the Internet through a community Wi-Fi with his login/password, he still has Internet access with the same parental control.

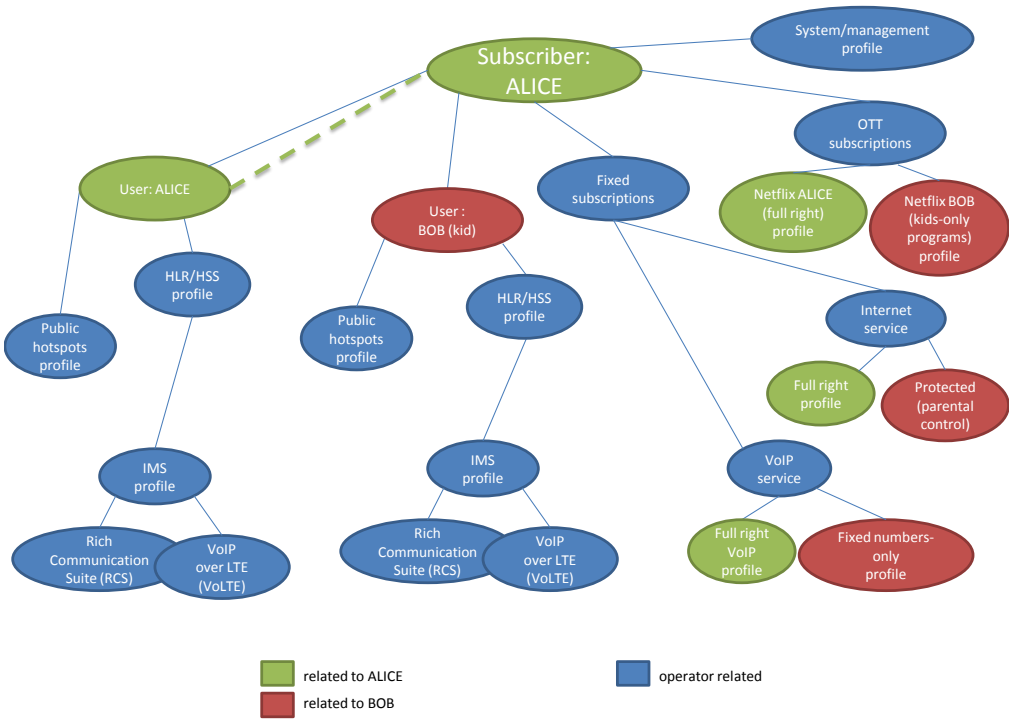


Figure 11: Organization of user data profiles for a FMC subscriber

The same applies for an OTT service that is included in the subscription with the FMC operator: a Netflix VoD service can be provided to each user with different access right to contents. That is ALICE has a full right profile that grants her to rent any video, whereas BOB can watch kids-only programs and cannot rent additional videos.

As an implementation detail, the so called "system/management profile" can be considered for facilitating e.g. software updates or remote management. This profile would be hidden from the subscriber and is only available to the operator.

3.2.2 Unification of subscriber data

We build on the example presented in the previous section to now present one of the main advances proposed in the present deliverable to solve HT1.

Our proposed solution for unifying subscriber data is built upon the 3GPP's UDC architecture, with the following enhancements:

- consideration of data model (3.2.2.1);
- new Front End (FE) applications (3.2.2.2);
- database access optimisation (3.2.2.3);

FE scalability and introduction of user location function (3.2.2.4).The 3GPP UDC architecture considers the split between network applications (FE) and databases (UDR back ends) and specifies the interfaces between them with the *Ud* reference point. We kept them unchanged:

- data access messages on the *Ud* interface will make use of Lightweight Directory Access Protocol (LDAP) (IETF RFC 4510)
- subscription messages and notification messages on the *Ud* interface shall make use of SOAP/XML.

Another progress beyond the state of the art, described in Section 3.2.3.3, is the extension of the UDC concept to integrate OTT service providers. This extension allows in particular new business opportunities for FMC operators as pointed out in Section 2.1.5.

Figure 12 shows how each network type has a different FE available and how each of them communicates with the database. The new FE needed for supporting FMC and OTT needs are highlighted in Figure 12. Some legacy protocols should still be supported temporarily for specific FEs:

- the MAP protocol for Home Location Registers used in 2G/3G [46];
- the RADIUS protocol for managing AAA in legacy fixed or Wi-Fi networks [47].

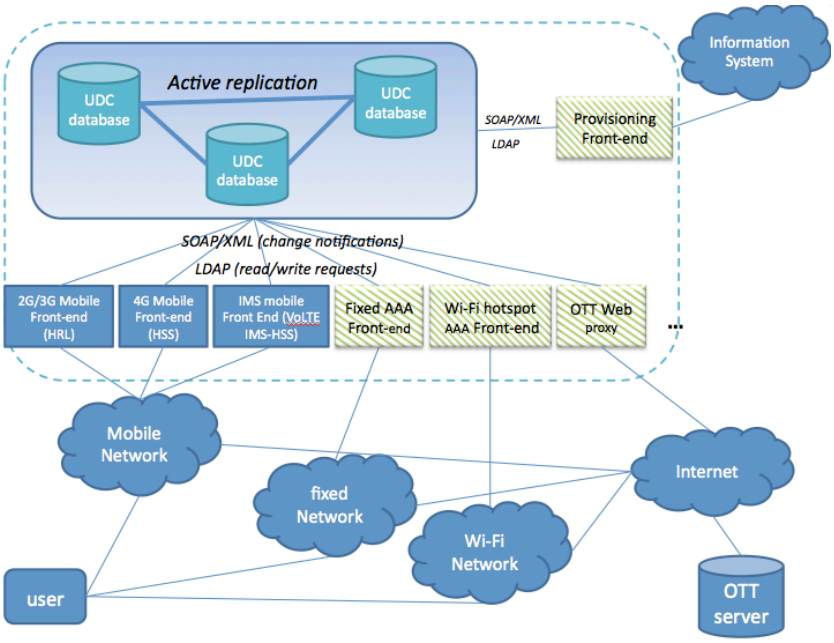


Figure 12: FMC subscriber database with dedicated Front Ends

DIAMETER [48] is recommended as the main protocol to be used in the interfaces between FEs and future 5G network elements. Although RADIUS is still massively used today, DIAMETER was designed by IETF to enhance Radius and to replace it.

Representing a logically single UDR does not mean that all subscriber data are centralised; however this entity will have a view on all user databases in place in the network. For example, the FE that is involved in the communication with the 3rd party operator or service provider (OTT) should synchronize subscriber data with this partner and should subscribe to a notification mechanism for any change of user data, in order to have an up-to-date consolidated user data.

3.2.2.1 Data model consideration

The internal structure of UDR is out of the 3GPP standardization scope. However, in the scope of HT1, it is worth considering a relational database approach in order to address the complexity of user and subscriber profiles. Such an implementation will help in linking the various user profiles with the main subscriber profile.

3.2.2.2 New Front Ends applications

As UDC is mobile-oriented, it does not consider FMC application. Some additional entities for the following network functions are necessary:

- RADIUS AAA supporting the RADIUS protocol between the FE and the other network elements (BNG, Wi-Fi controller)
- a specific FE for managing OTT authentication and subscriber/user data exchange (user profiles, accounting data etc.)

The multiple FE approach will allow adding additional FEs in the future without impacting operational solutions. This also enables making each of the FEs evolve in terms of features independently without affecting the rest of the system

3.2.2.3 The optimisation of database access

The UDR database may either be distributed over different locations or centralized. For ensuring high availability, it is worth deploying it with geographical redundancy and active real-time replication mechanisms. Moreover, given the criticality of such database for the availability of the FMC network, some back-up functions to secure the storage of data are necessary.

The FE are not aware of this internal structure, but to limit the access to the data repositories, some mechanisms can be implemented inside the FE as shown below.

3.2.2.4 Front End scalability

In order to ensure the scalability of application FEs, a load balancer inside the FE entity distributes the signalling across multiple instances of the application to be served.

Regarding the access to user data profiles in the database, to avoid numerous useless accesses to the UDR database, a User Location Function (ULF), (inspired from the IMS's SLF function) and a user data cache are introduced. The purpose of the ULF function is first to indicate whether the requested user data is available in the local cache of the FE entity, and second to indicate in which data repository of the UDR the requested user profile can be found.

A notification mechanism with the UDR database should also be in place in order to notify FE caches of any change of user data. Figure 13 details the structure of a mobile FE inside a UDC architecture for FMC needs.

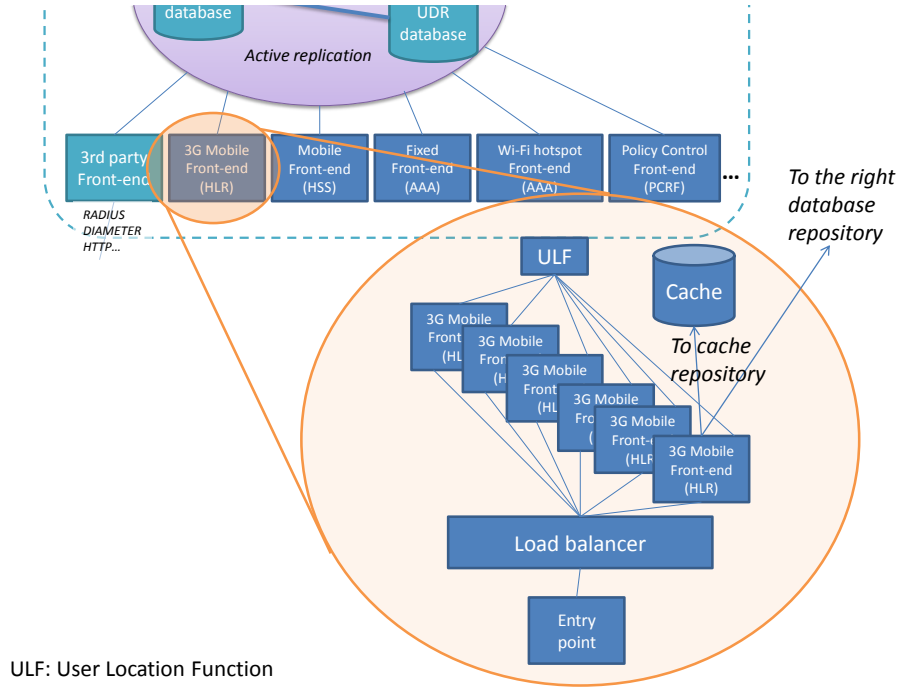


Figure 13: "Zoom" on a specific Front End to illustrate Front End Scalability

3.2.3 Authentication convergence

On authentication convergence, the main idea is that a subscriber using any authorised network access type (e.g. fixed, Wi-Fi, LTE) should be able to authenticate seamlessly. Moreover, the authentication made initially for accessing the network should also be usable for accessing added value services, provided either by the operator itself, or by an OTT service provider through a partnership.

When looking at the numerous authentication methods that are available today or will be available in future 5G networks, it is worth to have a common mechanism for the negotiation of the authentication method to be used.

A similar type of requirement is fulfilled by EAP [49] in IEEE 802.1X. EAP defines the encapsulation of the Extensible Authentication Protocol (EAP) at layer 2, whatever the transporting protocol (PPP, Wi-Fi or Ethernet). EAP typically runs directly over the link layer without requiring IP and therefore includes its own support for in-order delivery and retransmission. EAP is an authentication protocol that supports multiple authentication mechanisms, including 3GPP mobile authentication mechanisms that involve SIM cards. These 3GPP mechanisms supported in IEEE 802.1X are called EAP-SIM for 2G, EAP-AKA for 3G and there is also a variant of AKA for 4G (LTE). Anyway these mechanisms based on EAP are only used today in Wi-Fi authentications.

3.2.3.1 Application to mobile authentication (LTE and beyond)

For the support of traditional UE (typically those carried by human users), a 5G network should support EAP for carrying authentication messages between the MME and the UE for the mobile access authentication. The current LTE security

mechanisms can be used in future networks as well or even be enhanced, but it should always be carried over EAP.

This could be different for new types of UEs, e.g. those participating in M2M/IoT scenarios.

3.2.3.2 Application to fixed authentication (xDSL, FTTH)

In order to also support fixed networks, EAP should be extended with an additional authentication mechanism that should support validating physical lines and RGW identifiers.

AAA servers, e.g. FreeRADIUS[2] can already check MAC (Media Access Control) addresses against a database and can thus be extended to support line ID and RGW ID. For this last one, as a RGW is a network device, the MAC address can be this identifier.

Moreover, if the FMC operator prefers a strong authentication for the fixed access, a SIM card can be embedded in the RGW and an EAP-SIM-like mechanism can easily be put in place. This solution is interesting in particular for FTTH deployments with GPON where there are multiple subscribers connected to the same GPON tree. The EAP authenticator should then be hosted in the Broadband Network Gateway (BNG).

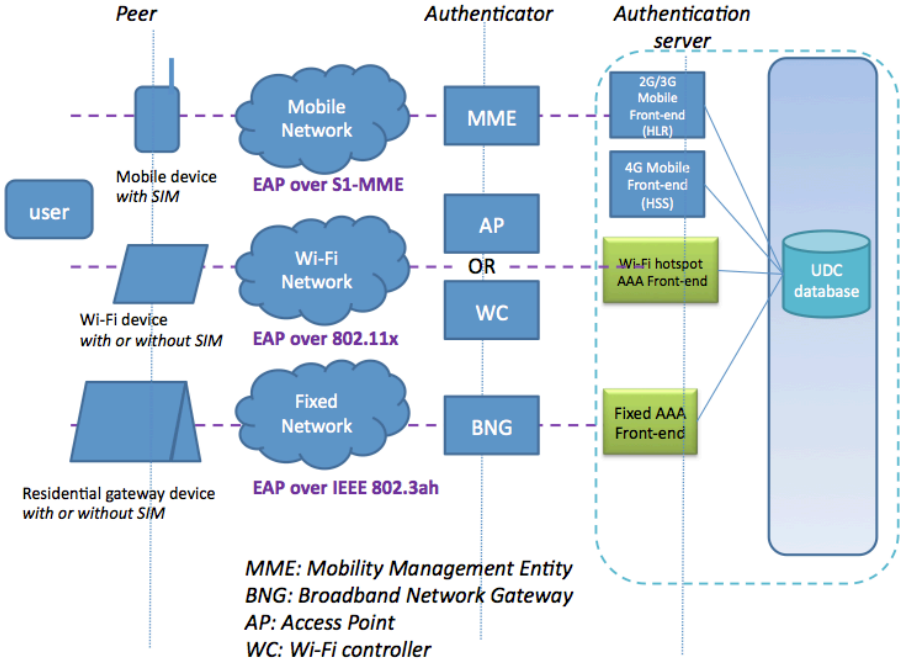


Figure 14: uAUT transport over EAP protocols

Whatever the access technology, it is only necessary that the device is able to setup an EAP session with the access point; then, the desired authentication mechanism will run on top of the EAP session. Obviously, the choice of the authentication method should always be under the control of the FMC operator. The authentication

[2] http://freeradius.org/

server will either be the uAUT server directly or get its credential data from the uAUT server. The authenticator will be different depending on the network type. This is illustrated in Figure 14, where each method supporting EAP is shown, depending on access type.

3.2.3.3 Application to OTT service providers

An FMC operator has a special relationship with its subscribers thanks to several subscription features: flat rate subscriptions, automatic billing mechanism, network usage and location data etc. This enables the FMC operator to provide interesting features to OTT service providers, such as:

- A seamless authentication of the mobile subscriber using an OTT service included to his/her network subscription.
- A strong authentication at network access level, by keeping confidential all the private mobile subscription data that should not be disclosed to the OTT partner. It will thus respect the private data protection rules.
- An accounting made by the OTT partner for the usage of the service and the corresponding billing made by the mobile operator
- A preferential accounting for OTT related traffic in the monthly data volume limit.

For the service access, the proposal is to use a standard like SAML (Security Assertion Markup Language). SAML is an XML-based, open-standard data format for exchanging authentication and authorisation data between parties, in particular, between an identity provider and a service provider. It uses web browser single sign-on (SSO). SSO solutions are widely used by OTT service providers, generally using cookies. This is illustrated by Figure 15, which represents the call flow of the setup of a user session for accessing an OTT service through a FMC network.

The successive steps of the procedure shown in Figure 15 are the following:

1. After a network access authentication, the user requests an access to the OTT service.
2. A Web proxy located at the egress of the FMC core network intercepts the HTTP request. The proxy requests the user data token for this service from the uAUT.
3. The uAUT provides the token, extracted from the right user profile in the corresponding subscriber tree; the token provides the following information:
    a. the user identity, which is linked on operator side to a subscriber and to a subscription to the OTT service.
    b. the user profile: what the user is allowed to do according to the subscription and to the user rights (i.e. the user “BOB” linked to the subscription of his mother “ALICE”).
    c. the IP address of the current session
    d. a certificate for the current session

e. miscellaneous other data: device capabilities, user location, load of the network etc.

4. Then the Web proxy informs the OTT service provider that a user wants to access to the service. A SAML message carries the token which contains all data needed to authenticate and authorise the user according to his/her rights.
5. The access to the service is then granted to the user for this session.

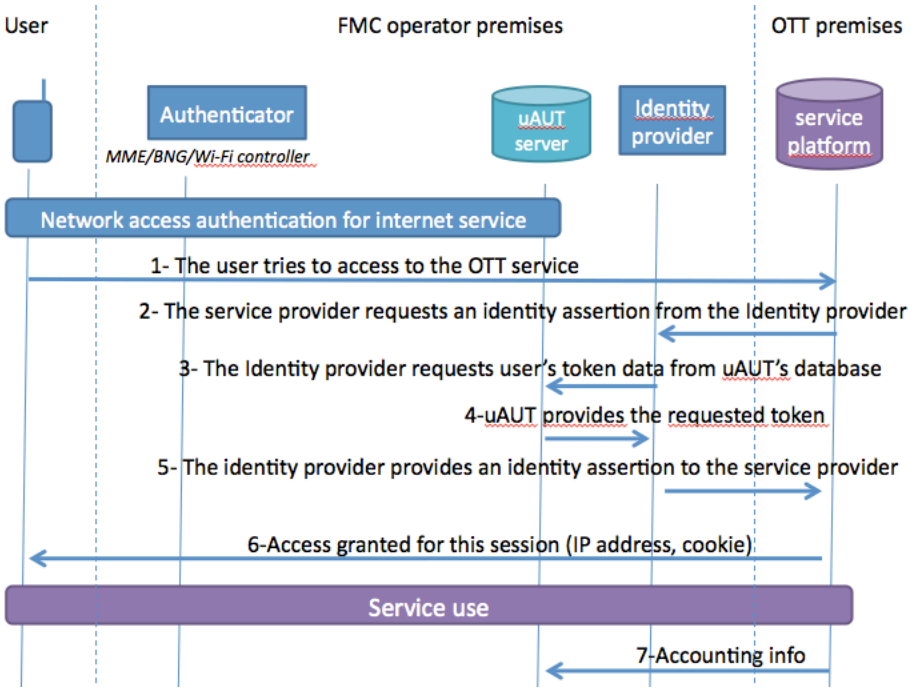


Figure 15: Authentication of a mobile user to an OTT service

3.3 Migration paths to HT1 target

The target technical solution for HT1, as described in the previous section, tries to fulfil the needs of an FMC network in 2020 and the requirements that 5G mobile networks might have. There is a significant gap between this ultimate solution and the currently deployed architectures. Therefore, it is useful to develop "migration paths" approaches towards that target to allow a step-by-step evolution.

In the current Section, a "short-term" view is first identified, which would include the first changes necessary to support some of the converged use cases with a low impact on existing equipment and deployments. Then, a progressive method to reach a fully converged uAUT solution is identified, which would bring the first converged user database and multiple FEs whilst not providing all functionalities required by the ultimate solution.

3.3.1 Short-term view

To avoid any major change in legacy subscriber data network elements in terms of functions and protocols, these network elements are kept as they currently are and a new network element called uAUT is introduced for all new functionalities.

A way to quickly implement a uAUT to support the so-called “hybrid access” is now provided. Firstly, the uAUT has a UDC-based architecture, with one or several FE which support the protocols needed for communicating with other network elements: the mobile HSS (DIAMETER), the OTT proxy (with single sign-on standards like SAML or OpenID), the Wi-Fi hotspot AAA (RADIUS) or the broadband access AAA (generally RADIUS as well). In particular, it is necessary to go through an intermediate uAUT network element before accessing data from the origin subscriber due to security reasons.

The OTT case is particular in the sense that the uAUT mechanism does not authenticate the access to the network but the access to the service. However, the uAUT uses the authentication previously done for the network in order to provide a seamless access to the OTT service and to obtain, from the FMC network, accounting data relative to the usage of the OTT service.

To perform correctly the above procedures, there are some important requirements for the uAUT:

- There should be some specific data obtained from the subscriber profile stored in the origin subscriber database to ensure that the user is allowed to use the network or the service.
- The uAUT should implement notification subscription mechanism to get notification of updates on subscribers’ profiles from other databases.

Figure 16 shows how a uAUT can be deployed in short-term by an operator. A “Hybrid” access user is a user who tries to connect to a network that is not linked to its base subscription, e.g. a mobile user with SIM card who use its mobile credentials to connect to a Wi-Fi hotspot. Note that it is not mandatory for the operator to manage all network types. It is possible to provide such service with third party operators with which there is a partnership (e.g. a Wi-Fi operator). The uAUT then behaves as a mediator between them for all subscriber management tasks: seamless authentication and potentially, profile management (authorisation) and accounting.

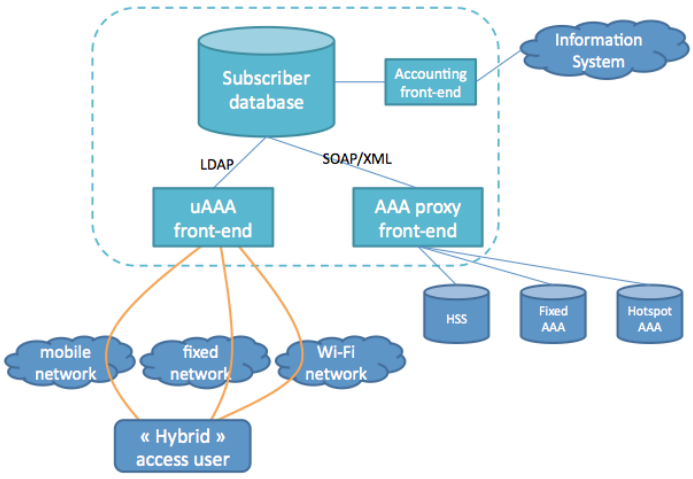


Figure 16: Short-term approach for uAUT

3.3.2 Steps towards the ultimate target

The short-term view is a way to provide seamless authentication to users immediately and possibly authorisation and accounting for various access networks and services. However there is no pooling in operator's infrastructure as there are still several subscriber databases, all linked to the uAUT server.

- For an operator managing different network types, a unified subscriber management system can be built progressively by merging the different databases into a single uAUT. This is illustrated in

Table 3, which represents one possible migration path from the current situation to a situation where a uAUT solution is fully deployed.

- The uAUT server is first deployed as a new standalone network element linked to legacy subscriber databases (HLR, HSS, AAA). It has one dedicated FE behaving as the *"Identity Provider"* for OTT collaboration and one FE for *Hybrid Accesses*, e.g. when a user connects to a network different from his base subscription.
- The network functions related to subscriber management will then progressively be ported in the uAUT server, by merging subscriber data within the UDR database and by deploying new FE for each network element to be replaced. The legacy subscriber databases will be then dismantled.

Table 3: Possible steps towards HT1 target

Steps to the target solution	Hybrid access	OTT Identity provider	Mobile HSS/HLR	Mobile IMS-HSS: for VoLTE, RSC	AAA for Wi-Fi hotspots	Fixed IMS-HSS for VoIP	AAA for fixed access
Current situation	Partial support	Dedicated AAA	Dedicated HLR/HSS	Dedicaded HSS	Dedicated AAA	Dedicated IMS-HSS	Dedicated AAA
Short term HT1 solution	uAUT Front End	uAUT Front End	Dedicated HLR/HSS	Dedicaded HSS	Dedicated AAA	Dedicated IMS-HSS	Dedicated AAA
Mid term 1	uAUT Front End	uAUT Front End	uAUT Front End	uAUT Front End	Dedicated AAA	Dedicated IMS-HSS	Dedicated AAA
Mid term 2	uAUT Front End	uAUT Front End	uAUT Front End	uAUT Front End	uAUT Front End	Dedicated IMS-HSS	Dedicated AAA
HT1 Target	uAUT Front End	uAUT Front End	uAUT Front End	uAUT Front End	uAUT Front End	uAUT Front End	uAUT Front End

Thus, the uAUT server will provide progressively a unified view for subscriber management to the information system (customer care, billing, usage statistics, etc.). In particular, all subscribers' profiles related to a given user (i.e. mobile and fixed broadband subscriber) will be either merged in a unique one or explicitly linked together.

3.4 Conclusion of Section 3

Section 3 has addressed HT1 “Converged Subscriber and Session Management” to alleviate the current situation where users hold different subscriptions, identities and authentication methods to be allowed to use fixed, mobile and Wi-Fi telecommunication services.

We first highlighted the need for user data consolidation. This leads to the requirement that true convergence requires a global view of a given user’s identities, even if all user data do not have to be hosted in a single user-centric database. Regarding the convergence of authentication, a common security architecture is necessary but it has been shown that all networks do not have the same requirements in terms of security level, even in future convergent 5G networks. Lastly, the solution should also allow the collaboration with OTT players and with Wi-Fi access points that rely on Web portals.

The proposed technical solution for HT1 relies on improvements to the 3GPP’s User Data Convergence [44] concept, namely splitting subscribers’ data repository from the application logic specific to each access type. A single functional block, the “universal subscriber and user authentication” (uAUT) server links several application logics, called “Front Ends” (FE) in the UDC framework, with a single global UDR. The UDR hosts and organizes user and subscriber data in a convergent manner so that a unified view of users and subscriptions can be provided to the operator. This architecture is completed with proposals regarding the optimisation of database access and FE scalability.

The target solution for authentication convergence is based on the idea that a subscriber using any authorised network access type (i.e. Wi-Fi, LTE, FTTH) should be able to seamlessly authenticate, using a common transport layer (EAP) for authentication mechanisms. On top of this layer, the choice of the security mechanism would be negotiated between the authenticator and the supplicant according to the requirements of the network and the capabilities of the device. Moreover, the authentication made initially for accessing the network should also be usable for accessing added value services, provided either by the operator itself, or by an OTT through a partnership.

Finally, a migration path for realizing the target solution in a realistic manner is presented. It starts with a short-term view in which an operator can provide seamless authentication and a first level of unification of user data by introducing a new network element for hybrid accesses, with little changes in the existing networks. Then the operator could smoothly deploy the targeted universal subscriber and user authentication solution.

The solution proposed for universal subscriber and user Authentication in this deliverable is currently being investigated in WP6 for testing purposes. Within WP6 the authentication of a device with the same credential set is being covered. In this way, a single mobile device is expected to authenticate through different access technologies of the FMC network, LTE and Wi-Fi, with a single identity.

4 Description and Analysis of HT2

The present Section focuses on the second HT that has been identified as key to providing FMC to network operators operating fixed, mobile and Wi-Fi networks. HT2 is intended to provide solutions for “Advanced Interface Selection and Route Control”. HT2 is not independent of HT1, as the network operators only provide network connections to the UEs of authenticated subscribers.

However, although it depends on HT1, many capabilities, beyond those fulfilling HT1, are requested to fulfil HT2, as shown in the following.

We propose to deal with HT2 thanks to a set of functional blocks that realise a “Universal Data Path management” (uDPM). Instances of these blocks could be used to help reducing FMC operators’ OpEx thanks to cross-layer optimisation techniques and to improve delivered QoS (e.g. reduce latency and limit the impact of congestion) by providing mechanisms for seamless service delivery.

4.1 Description of HT2

Mobile traffic offload is a major objective for FMC. As shown in Section 2.1, it can help in avoiding radio resource congestion and can potentially bring bandwidth gain in the metro and core network. This is especially valid for some services such as video distribution or cloud storage, in which the requested content can be found in several locations on the network.

The objective of HT2 “Advanced interface selection and route control” goes beyond mobile traffic offload on Wi-Fi. Indeed, multiple data paths could potentially be used to serve a UE:

- In case of mobility, the UE can change its attachment (moving from one eNB to another), and the UE can transiently use two data paths during its re-attachment procedure;
- A specific content requested by an UE can potentially be served by several servers; in some cases, it would be convenient to change the server within a session (e.g. for load balancing purposes, or because the UE has moved and is closer to another server)

The generic issue to be solved is to allow UE requests to be served on several data paths while ensuring session continuity, even in case of mobility. In some mobility cases, session continuity is not guaranteed with the current 3GPP procedures as it generates an interruption delay longer than can be supported by the application. QoS requirements regarding different classes of traffic are specified in [44], where in particular the delay between the UE and the PDN is specified. These characteristics are reported in Table 6.1.7 of [44] which is replicated in Table 4.

3GPP also mandates that the interruption delay be limited to 300ms for real time services and 500ms for non-real time services.

Although new services are envisaged for 5G [18], no new QoS requirements are yet specified for these services.

The problems to address in order to solve HT2 are the following:

- A data path at the UE side ends at an IP address that is allocated either by the mobile or by the Wi-Fi network; these addresses may differ, even if the FMC operator has bound both addresses to the same subscriber. A session is usually associated to a single IP address. Some applications present time constraints (such as video streaming) that preclude changing the IP address within the session.
- Multiple servers, with different IP addresses, should be able to collaborate to distribute content to a single UE, whether the UE is fixed or mobile;
- The data paths controlled by the mobile operator should comply with the constraints set by 3GPP, in terms of tunnelling the traffic towards a SGW, and then a PGW. Moreover, as SGWs are mobility anchor points, a UE is not allowed to connect simultaneously to multiple SGWs.

Table 4: Standardised QoS Class Identifier (QCI) characteristics

QCI	Resource Type	Priority	Packet Delay Budget	Packet Error Loss Rate	Example Services
1	Guaranteed bit rate (GBR)	2	100 ms	10^{-2}	Conversational Voice
2		4	150 ms	10^{-3}	Conversational Video (Live Streaming)
3		3	50 ms	10^{-3}	Real Time Gaming
4		5	300 ms	10^{-6}	Non-conversational Video (Buffered Streaming)
5	Non-GBR	1	100 ms	10^{-6}	IMS Signalling
6		6	300 ms	10^{-5}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		7	100 ms	10^{-3}	Voice, Video (Live Streaming) Interactive Gaming
8		8	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		9			

The network operator should be able to either partially or fully control the data path taken by the data to and from a UE, including over the first leg of the end-to-end data path (mobile versus Wi-Fi interface), although the UE's OS currently performs the control.

COMBO proposes a functional block, the universal Data Path Management (uDPM), for solving HT2. uDPM is described in the next Section.

4.2 Universal Data Path Management

In the current networks, each operator handles a single data path for each user at any given time. However, in the case of an FMC network, multiple access technologies are available to UEs, and different data paths might be available at a given time for a user thorough the network. Under these circumstances, the issue to be solved consists in providing the tools to map a given session on one (or several) data path(s), while ensuring session continuity. The generic solution proposed by COMBO to solve this issue is illustrated in Figure 17.

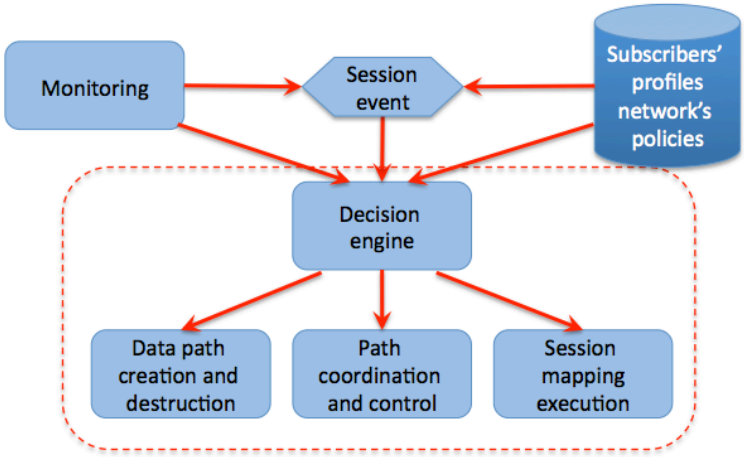


Figure 17: Universal Data Path Management (uDPM) as chained functional blocks

The uDPM functional block is triggered by a "session event", which represents any singular event relative to the activity of a particular UE; such an event can be the request to launch a new application, as well as the need to change the interface used by the UE, or the data forwarding process, due e.g. to an eNB detecting that handover is necessary. A "session event" can be generated by the "monitoring" functional block; for example:

- either the network or the UE measures a degradation of the received signal strength;
- the UE detects a new Wi-Fi access point.

A session event can also be triggered by the application of a subscriber's profile or network's rule (e.g. offload residential LTE traffic on Wi-Fi during business hours). The set of rules to take into account are stored in a repository space named "Subscribers' profiles and network's policies" on top of Figure 17. As illustrated by the above examples, a session event can be generated either by the UE or by the network.

The decision to map a given session on a particular (set of) data path(s) should be controlled, at least partially, by the network operator. Let us call "Decision Engine" the set of functions used to reach a session mapping decision. The decision engine is typically fed by (proprietary) policy rules that the network operator chooses to

enforce. These rules may be related to global network management issues such as load balancing (between network types, or between mobile cells, or between LGWs…). Some rules also may be related to enforcing service features specified in user profiles, and accessed by the FMC operator thanks to the uAUT. The decision engine thus relies on both the monitoring processes that shall help in identifying specific events (overload or congestion over network areas or data paths, UE mobility, etc.) and the set of rules stored in the “Subscribers’ profiles and network’s policies”. The processing of multiple rules and translation of these rules in terms of decision(s) can typically be modelled as a multi-criteria decision-making (MCDM) problem.

In some cases (e.g. UE mobility), the decision can be to create new data paths, and/or to destroy others; in Figure 17, these functions are grouped into the “data path creation and destruction” block. The mapping processes may differ in upstream and downstream directions. In the upstream, it is likely that the UE shall have at least partial control of path creation/destruction using its various interfaces, although the network operator, by remotely activating interfaces, and allocating IP addresses, shall also participate in this control. A fuller control of the data paths in the downstream direction by the network operator is to be expected (e.g. selection of a given server, or of a given data path between the server and the UE).

As the session may rely on several data paths, either concurrently or successively, some mechanisms may be necessary to ensure that concurrent data paths smoothly deliver the packets corresponding to the session, and that session continuity is guaranteed, even in case of UE mobility. These functions are gathered within a “Path coordination and control” functional block.

Both the “data path creation and destruction” and the “Path coordination and control” functional blocks are part of the control plane of the FMC network.

The last set of functions to consider is the “Session mapping execution”. As part of the transfer plane, it enforces the session mapping decision taken by the decision engine, and relies on the control performed by both “paths creation and destruction” and “path coordination and control” blocks. Packets corresponding to the session are filtered and forwarded on the data paths selected by the decision engine; sub-flows may be merged when multiple paths are concurrently used

The various functional blocks of the uDPM are shown in Figure 17, and further described below.

While uAUT can be considered as a unique solution for HT1, uDPM has to deal with a multiplicity of specific problems. This section proposes a set of solutions to specific problems; three such solutions are currently identified: very tight coupling between Wi-Fi and LTE access, which allows a user to seamlessly move from LTE to Wi-Fi (Section 4.3), smooth SIPTO-based mobile access (Section 4.4), which allows a user to seamlessly access the IP backbone and the Internet thanks to LGWs, and reactive content placement (Section 4.5), which is based on the Content Distribution Service (CDS) management to react to user location in order to improve content placement. In each case, the high-level implementation principles of the particular uDPM solution are given. Functions forming the uDPM are not required to be placed into a single and centralized uDPM entity. Indeed, in some cases, functions are implemented in

the network while others are implemented in the UE or are distributed between both network and UE.

4.2.1 Decision Engine

The decision engine module is a part of the uDPM system, and it is in charge of making decisions based on different sources of information in order to modify the behaviour of the different network elements related to HT2. With these decisions it manages the data paths and the interface selection process by taking into account all FMC network management related aspects.

In this Section, a general overview of this module is first made, where the internal structure of a typical decision engine is presented. Afterwards, the network management possibilities obtained through the module are shown. Additionally, the input information that is required and the actions that can be taken as output are also covered. Then, the decision algorithm is described and finally application examples of the uDPM illustrate some possible workflows and show potential benefits in the network.

4.2.1.1 General overview

The decision engine module takes care of evaluating the network and its status in order to make changes that improve, or even optimise, the network's usage in terms of various criteria such as resource usage, delivered QoS, OpEx, etc. The engine is composed of different elements that take this kind of decisions and propagate them in order to obtain the expected results.

The two main entities involved in the decision engine are the decision logic within the UE and the decision algorithm that is located within the network. The former is linked to the Wi-Fi connection manager of the device, which is capable of employing policies in its decision process. This is already the case for the ANDSF and HS2.0 frameworks. The latter is more complex since it should be capable of modifying the policies that are sent to, or retrieved by, UEs. Additionally, the decision algorithm takes into account not only the nearby access conditions but also the state of the network, or of at least an area of the network.

The decision algorithm can typically be modelled as a MCDM problem. Multiple methods have been proposed to handle such problems, and each network operator may implement its preferred method. Appendix 1 explains, as an example method, how control theory can be used to maximize a utility function characterizing the quality of a decision. Some problems may be very complex, due to the number of rules or the number of parameters to consider. It may be necessary to either use heuristics or to separate the problem into simpler instances. In this last case, this may correspond to distribute the decision process between several functional entities, each taking its decision independently of the others as it is explained in Section 4.2.1.4.

The following is the list of the elements that are part of the decision engine at the network's side:

- **Optimisation target:** The operator can set an optimisation target to be considered as a network management strategy. Different target options that

are relevant to operators have been considered in the present deliverable are listed below in the next Section. This target would be provided to the decision algorithm as a "*configuration*".

- **Decision algorithm:**
  - It is the logic that, by analysing the data of the network topology and the current network status, performs actions in order to optimise the network's usage. With these actions the algorithm seeks to achieve the target that has been set as the network management strategy. The actions, which are listed in a following Section, allow modifying the behaviour of UEs and other network elements. For this purpose the Decision Engine interacts with the Data Path Creation/Destruction and the Session Mapping Execution blocks as shown in Figure 17.
  - The decision algorithm can run both periodically or asynchronously. In some cases, both types of execution may happen in parallel to analyse different parts. The former allows periodic reviewing of the network status to check if any change is necessary, while the latter is available to be triggered by a network event that requires specific attention.
  - One of the key points of the Decision Engine is that it is capable of responding to several situations in real time. Existing policy based management systems, are not well suited for real time management and thanks to this extension, the FMC operator obtains new management capabilities.

4.2.1.2 Example of optimisation targets

There are different network management objectives that can be used to configure the decision engine. For example, it could be possible to define the decision engine to optimise the network for the following aspects:

- Optimise the network operation cost (OpEx);
- Optimise the use of network resources in order to improve energy efficiency;
- Optimise the efficiency of cache usage and improve the QoS of content delivery;
- Optimise for providing the best possible QoS to all users or to a user class;
- Optimise for providing enhanced availability to all users or to a class of users;
- Optimise other indicators (e.g. performance indicators, such as the signalling load);
- Any combination of the previous ones.

With the optimisations listed above as targets to achieve, the decision algorithm tries to achieve them by e.g. performing changes in the behaviour of the UEs, locate content to distribute in different cache nodes, and activate/deactivate the different network features.

4.2.1.3 Inputs and Outputs of the Decision Engine

Input Information

The decision engine employs different sources of information to feed the hosted decision algorithm. On the one side, the algorithm may receive events from the network that need to be attended. On the other, regular polling over certain (monitored) performance indicators can be requested in order to assess the network's status. Several aspects can be considered:

- Network related information:
  - The location of each access node (e.g. Wi-Fi access points, mobile base stations).
  - The measured traffic load of each of the stations/access points.
  - The energy consumption of each network element involved in the actions that the decision algorithm can consider.
- Device related policies, such as the ones related to ANDSF and HS2.0, which are in force at the given time of execution for mobile devices and other network elements. Different policies may be in force in a per user class or per user level in some cases.
- Content related information, e.g. identification of the content currently stored in the caching system and the content that is currently being accessed by users.
- Subscriber related information, as the network has access to the subscribers' profiles to take them into account in the evaluation process. For example, the QoS or service class that a group of users belongs to can be taken into account when moving their traffic from an access technology to another. Based on this information the Decision Engine might decide not to perform the handover, or to pre-fetch contents to maintain the QoS level.

Output actions

Given the optimisation target the Decision Engine has received, the decision algorithm can perform different actions in order to achieve it. The following is the list of actions that can be performed and that constitute the output of a decision cycle:

- Remotely activate or deactivate interfaces or send additional commands (this is fully described in Section 4.2.2). For example, it is possible to disconnect a user from Wi-Fi due to network congestion in order to guarantee video call quality. The execution of this type of commands is not currently possible.
- Activate and deactivate dual attachment. It is possible to increase the QoS as well as the availability provided to a user by enabling dual attachment. By doing so, it is possible to obtain a higher throughput or lower service interruptions. For such a purpose, the UE will only connect to Wi-Fi networks that allow this feature. When deactivated the device may connect to any Wi-Fi network approved by the operator even though it does not provide any mobility features.
- Modify user / device policies.

- For example, it is possible to modify policies belonging to ANDSF and HS2.0 so that a group of users offload onto Wi-Fi while others do not. This provides a better QoS to the top user class while reducing the cost (combined target).
- Content can also be pre-fetched according to the UE interface and network topology changes so as to improve the QoS of content delivery service.
- A combination of the examples yields the following: when the decision engine decides to offload a group of users onto Wi-Fi, the content currently required by these users can be pre-fetched on the cache nodes close to the Wi-Fi APs in such a way that content can quickly be retrieved right after the handover process.

4.2.1.4 Decision Algorithm description

The decision algorithm, as stated before, is a case of MCDM problem. The evaluation criteria are the targets that have been provided to the algorithm as configuration. Additionally, the data to evaluate is the one obtained from the monitoring module or the generated session events.

Upon the occurrence of each event, the algorithm will evaluate whether it requires any output in order to obtain the goal set. If it is the case, it can make use of the output actions that are available.

An example workflow of the decision algorithm behaviour is illustrated in Figure 18.

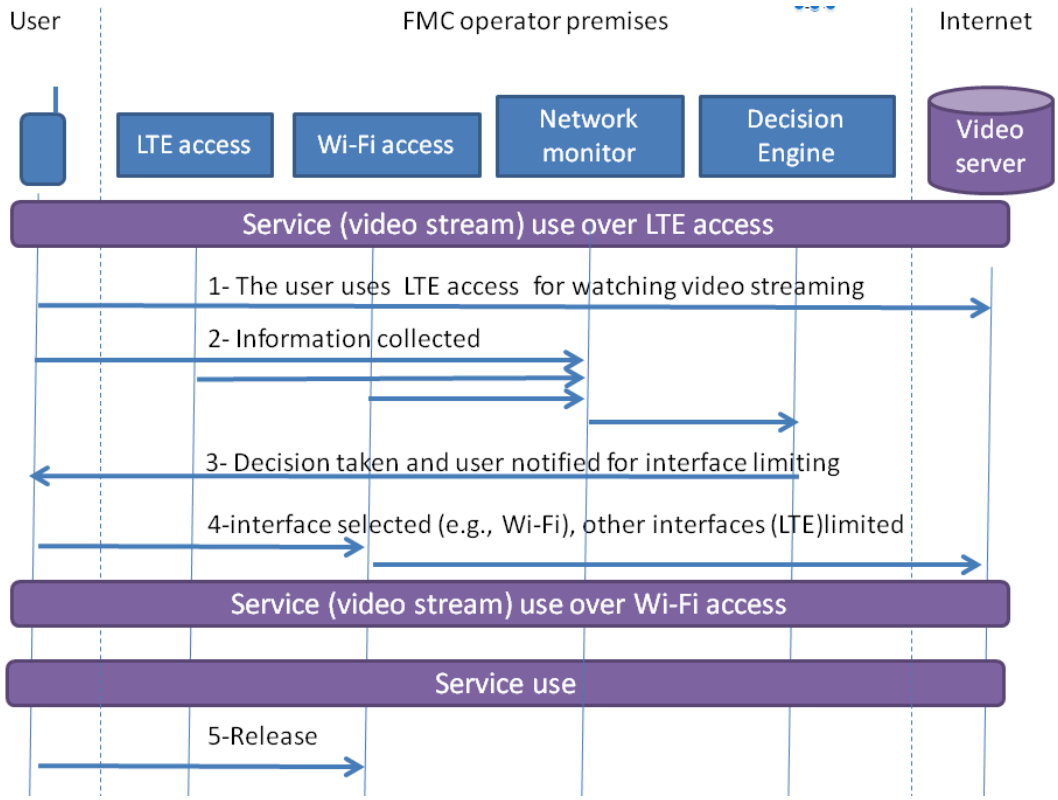


Figure 18: An example workflow of how the decision engine operates

In Figure 18, the Decision Engine makes use of the capabilities of the FMC network to divert traffic from the LTE access to the Wi-Fi access and back. In this case:

- First, the user connects to LTE and starts consuming a video streaming service over this access.
- Then, the Decision Engine retrieves the current status of the network and evaluates it. The Decision Engine notices that the cost of the UE's traffic can be reduced since there is non-congested Wi-Fi access available.
- The Decision Engine notifies the UE to connect through Wi-Fi to decrease the network operation cost.
- Finally, the UE continues to consume the video service over Wi-Fi.

In this example, the Decision Engine has optimised UE behaviour taking just cost into account.

As another example, the following scenario can further illustrate the role of the decision engine.

- Several users of the FMC operator arrive to an area where there is both LTE and Wi-Fi connectivity. There is a low traffic being generated at that moment, the Decision Engine decides to use the Wi-Fi access for all users since this lowers the operational cost for the operator.
- However, as time advances, the number of devices and the traffic in the area increases, partially saturating the Wi-Fi access. Because of this, some users start experiencing a decrease/degradation in their connection throughput.
- In such a case, the Decision Engine decides to keep all newcomers in LTE and evaluates the conditions in the Wi-Fi access in order to move some of the users back to LTE. For that purpose, it forces their disconnection from the Wi-Fi access. In this case the Decision Engine takes into account the QoS level that each user profile accessed by the uAUT system in order to guarantee them while performing the handover process.
- When the traffic in the area decreases again, the Decision Engine then moves all possible traffic to Wi-Fi again since quality levels are acceptable for all users.

In this second example, the FMC operator optimises the use of its resources to minimise the OpEx while providing a good quality in its service to all the users. For that purpose, a combined target has been is used.

Decision engine impacts in content delivery

There are different cases where the content delivery system can interact with the Decision Engine.

The first case is to help the optimal selection of UE interface for traffic offloading. Since the monitoring module keeps the mapping of all the content available in the caching system and its location, it can provide the information to the Decision Engine. When a user is asking for a piece of content that is cached in a cache node embedded in the Wi-Fi network via mobile network, according to the content location

information provided by the monitoring module, the Decision Engine can modify the UE policy so that the user will retrieve the content from the local cache via Wi-Fi network and the traffic in the mobile network is offloaded.

On the other hand, the decision made by the Decision Engine can also impact the content allocation. For example, when an offloading decision is made, before sending the directive to the UE, the Decision Engine will first inform the content delivery system so that it is aware of the upcoming handover process and which access point the UE will connect to. Then, the content delivery systems let the cache node that is close to the access point pre-fetching the content that will be requested by the user. In this case, when the user switches to the new access point, the content is already available in the local cache and can be accessed immediately.

In these cases, the Decision Engine has the optimisation of the content delivery system marked as one of the targets in its configuration.

4.2.2 Data Path Creation and Destruction

In this section, the extensions to how data paths are created and destroyed in an FMC network are described in comparison to existing networks. The section covers two different parts of the data path, the RAN segment and the above the RAN segment, where appropriate proposals are described for each case.

4.2.2.1 RAN based processes

Existing technologies (such as ANDSF and Hotspot 2.0), provide the tools for operators to better manage the Data Path Creation in the RAN segment from a network perspective.

While the aforementioned technologies mainly facilitate the data path creation process for the case of Wi-Fi access, there is no feature or command available for disconnecting a UE from a specific Wi-Fi network or remotely switching the Wi-Fi interface off. The disconnection from a given network is a feature that is currently relied on the UE.

This kind of behaviour does not allow the Decision Engine to fully control the traffic steering process when taking into account the paths to be used in the RAN segment. For this purpose, a new set of commands is proposed, which allow managing the Data Path Destruction from a network perspective.

The proposed approach is to extend the UE's connection manager system (considering that ANDSF and HS2.0 extensions are already included) so that it can receive a number of new commands. These commands will be real-time ones and they are covered in the following list:

- Switch the Wi-Fi interface ON.
- Disconnect the UE from the current Wi-Fi connection.
- Switch the Wi-Fi interface OFF.

It is necessary that the FMC network controls the status of the Wi-Fi interface of each device. That way, it can switch interfaces on when the network knows that offloading is possible and disconnecting the user and switching the interface off when the Wi-Fi

connection is no longer needed or it is not usable. Therefore, the UE needs to inform the network of its status:

- When the interface is ON.
- When it is OFF.
- When it is connected to a network (operator managed or non managed).
- When it is not connected.

For these commands to be consistent with the existing use of the Wi-Fi access and the device behaviour, the operator will not disconnect the UE from a Wi-Fi network that the user has connected to. That is, the operator should not interfere with connections to non operator-managed Wi-Fi accesses. This has been considered so, since the user may have manually connected to a network that provides a specific service and disconnecting the UE from the network would mean disrupting the use of that service.

In order to introduce this behaviour, an extension of the S14 interface, which belongs to ANDSF, is proposed. The interface is extended with the capability of sending messages from the UE to the network in relation to the status changes, and to receive commands from the network, which are executed by the UE. This information exchange does not require requesting a new Management Object (MO) which is quite big in size, allowing these operations to be bandwidth efficient. Basically, in the exchange, a message will be sent through the connection, and its reception will be acknowledged.

If all devices increase the management traffic they generate in the S14 interface, a relevant overhead in the network would be created. This situation is not scalable without affecting the network and it might not be beneficial in some cases. In order to avoid this drawback, the following possibilities are included within the proposal:

- Each device can be remotely set to switch this feature ON or OFF through the S14 interface. This allows the network to use the current policies and procedures as such, when it does not make sense to assume the extra traffic exchange.
- Each device can report to a different ANDSF entity depending on the location it is in. The information for reaching the correct entity would be provided by the network through S14. This allows distributing the load on different entities that are not necessarily centralised.
- The ANDSF Management Object, which includes all the ANDSF policies, is extended so that the device can turn this feature on based on the time of day or location area, the same way it is done with other policies. This information would be included in a fourth type of information, named Remote Assistance Level Policy. This separation, would allow its future extension.

In summary, with the inclusion of the newly described feature, the FMC operator obtains further traffic steering capabilities. For example, this is the case of:

- Disconnecting some of UEs that are in a congested hotspot (and effectively moving them to LTE), thus providing a better QoS to all UEs.

- Activating Wi-Fi for offloading when the UE has the interface turned off and the user is not aware of the benefit.
- Taking into account the battery consumption of UEs and the consumption of each access networks to make energy efficiency based decisions.

4.2.2.2 Above the RAN processes

Two cases have to be considered. The first one corresponds to mobile traffic using other paths than the default one: using either a HeNB (thus part of a LGW), or exiting the default path at the eNB. Both cases correspond to LIPA and/or SIPTO procedures [40]. 3GPP has fully specified the data path creation/destruction processes for LIPA and SIPTO, but some cases may lead to session discontinuity in case of mobility as each LGW allocates an IP address to the appropriate UE interface. When a UE leaves the area covered by a LGW and requests attachment to another LGW, the initial address is lost. The solution proposed to maintain continuity is fully described in Section 4.4.

The second case consists in the path creation and destruction related to CDS. The creation of the path leg to the appropriate server in case of a content service relies on CDS specific procedures that currently point the user towards the appropriate server using either an HTTP redirection-based or a DNS-based process. When the decision engine can interact closely with the CDS manager, the selection of the appropriate server can be facilitated. Solutions specific to this particular issue are reported in Section 4.5.

In both cases, either standard procedures (protocols running between functional entities) or SDN based procedures (where forwarding tables and rules are computed centrally, and then distributed e.g. by OpenFlow to data plane equipment.

4.2.3 Path Coordination and Control

When the UE is connected to the network using several data paths, these paths will merge at some point called the Multi-Path Entity (MPE). In Figure 19 a general view of the data paths is shown. In this figure the UE denotes the user side termination point for the connections, e.g. a handheld device like a phone or a tablet, a computer or the RGW in the home network.

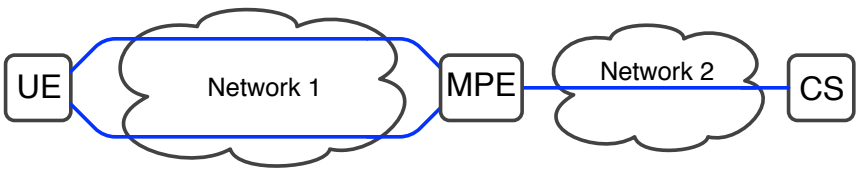


Figure 19: a generic view of dual interfaces in UE and data path coordination

In order to keep the content server (CS), unchanged, there should only be one address for the UE, meaning its point of presence will be the MPE. The MPE is the merge point for the dual paths, where there is a common protocol for the data paths. Thus, the dual connections to the UE can be used for traffic loading by data path control functionality. In this way the server can be completely unaware of the existing data paths between the MPE and the UE, which also gives a natural way to achieve

session continuity. While the data paths between the UE and the MPE can vary due to user mobility, the connection between the MPE and the CS is fixed. Note that an alternative to the MPE is for the CS to maintain a set of IP addresses for a single user (as e.g. in MPTCP used in Section 4.4).

In a typical FMC network the data paths between the UE and the MPE can go via a Wi-Fi access point or via the eNB. Without too much alteration of the UE hardware, it can connect to these two network architectures simultaneously. If the MPE can set rules for the outgoing traffic over the data paths, the network can take full control of the decision to map the session on the existing data paths.

Currently, the data paths are controlled by the end points. However, in an FMC scenario, it is useful that the network can coordinate the multiple data paths and session mapping decisions. This makes developing more efficient algorithms for off-loading or traffic management possible.

The position and implementation of the MPE can vary between network architectures and operators. It can be located close to the user, in the eNB, in the NG-PoP, or some other reference point defined in the network by the operator or service provider. It can also be located in the Content Server, and thus controlled by the OTT operator by e.g. MPTCP. All solutions have in common that, in order to control the traffic, all data paths have to use a common protocol. For example, if the eNB is used as MPE, the coordination is performed over a layer 2 protocol used by the mobile traffic, and if the MPE is within the NG-PoP, all data paths are over IP.

4.2.3.1 Coordination in eNB

Very tight coupling between LTE (Long Term Evolution) and Wi-Fi is a new concept developed by COMBO and presented in [14]. It is detailed in Section 4.3.

4.2.3.2 Coordination in the NG-PoP

If the coordination is performed in the NG-PoP the content server must see one IP address from the IP core. One suitable location for this can be typically the BNG at the IP edge. IP can be the common protocol for the two data paths in this case.

TCP or UDP can be considered as alternatives to IP as the common protocol for multiple paths. In this context the MPE can serve as a Multi Path Proxy [54]. The UE connects here via multiple paths, e.g. over Wi-Fi and mobile, and the traffic is forwarded to the server.

4.2.3.3 Discussion

The above solutions can all serve their purpose in different scenarios. For a mobile operator serving also a Wi-Fi network, it is natural to deploy a solution with the MPE implemented in the eNB to increase the infrastructure utilisation; this solution if considered in Section 4.3. In the case of a mobile and fixed access operator, or cooperation between two operators, potential locations to concentrate resources in an FMC access/aggregation network are the RGW (if it is under the control of the network operator) and the IP edge (which is always controlled by the operator). For an operator delivering a service, e.g. IPTV, as well as for an OTT service provider,

native MPTCP support or the use of a multi-path proxy can be suitable; servers supporting MPTCP are considered in Section 4.4.

4.2.4 Session Mapping Execution

Whereas the three previous blocks required methods to go beyond the state of the art, this is not the case for the Session Mapping Execution functional blocks.

Although the methods for creating data paths, and coordinate data transmission over these data paths, may be modified by the procedures designed for uDPM, the forwarding processes are not modified. This is because no new forwarding paradigm is actually considered for LTE (except possible for the RAN part, which is not addressed within COMBO).

4.3 Very tight coupling between LTE and Wi-Fi

Very tight coupling between LTE and Wi-Fi is a new concept developed by COMBO and presented in [14]. The idea is to apply LTE security procedures provided by the PDCP for Wi-Fi transmissions. To do so, Wi-Fi APs are connected to eNBs. The integration between Wi-Fi and LTE is done below the IP layer. Thus, the UE uses a single IP address for both Wi-Fi and LTE interfaces. Layer-2 bonding, seamless handover between the two accesses and session continuity are thus provided. Since PDCP security procedures are used for Wi-Fi, there is no need for Wi-Fi security mechanisms and the attachment of a UE to an AP can be very fast. Hence, it is possible to use Wi-Fi even if an AP covers the terminal for only few seconds. User's mobility is handled by regular LTE procedures.

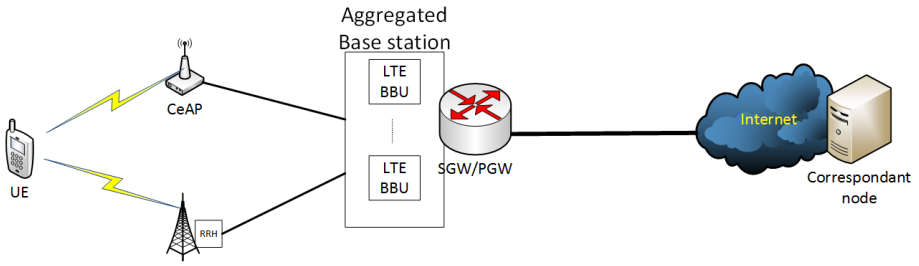
Very tight coupling can be used to offload the LTE network (switching traffic from the LTE path to a Wi-Fi path) or to increase the user transmission bit rate (if both LTE and Wi-Fi paths are used).

4.3.1 Topological aspects of very tight coupling

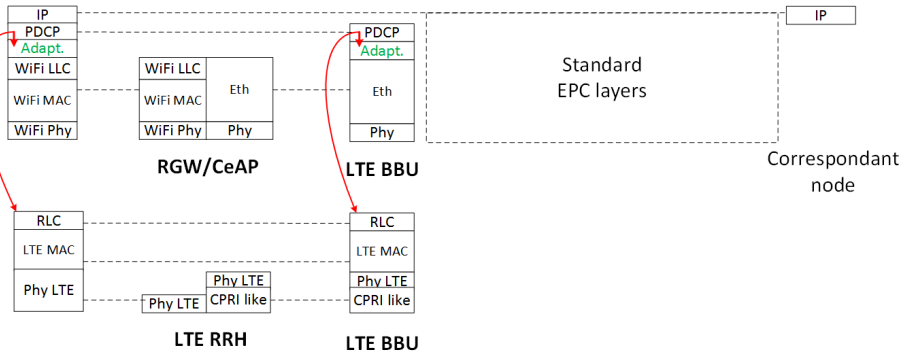
Figure 20 shows the architecture of a network implementing very tight coupling and a typical protocol stack for each possible path (i.e. Wi-Fi and LTE). The EPC is kept unchanged. A new functional entity called Cellular offload Access Point (CeAP) is added. A CeAP can be a Wi-Fi access point deployed and managed by the operator or a residential gateway with some additional functions: the CeAP broadcasts a SSID specified by the operator and is connected to one or several eNBs. The CeAP is a layer-2 entity and its functions are limited to only forward layer-2 frames between the UE and the eNB. From a security point of view, no Wi-Fi authentication mechanism is implemented in the CeAP since PDCP is already used as a security layer. This means that no specific security procedure (involving e.g. a pre-shared key) is needed for the UE to attach to the CeAP.

The functions of the uDPM are mostly implemented in the eNB. The decision to activate and to de-activate the Wi-Fi interface is taken by the eNB. The sender (e.g. the UE on the uplink, the eNB on the downlink) is responsible for Session mapping Execution, but the network is always in charge of the policy decision. The eNB implements the data path coordination function, as convergence point between Wi-Fi

and LTE. Since most uDPM functions are implemented by the eNB, it is called an "Aggregated Base Station".



a) Architecture for very tight coupling



b) Protocol stacks for very tight coupling

Figure 20: Main principles of very tight coupling

An aggregated base station is typically a cabinet grouping several LTE Baseband units (BBU hostel) as shown in Figure 20-a. We do not show details about the EPC since no modification affects it. The SGW and PGW can be separated but for the sake of simplicity, we show them as co-located. We assume that an FMC network is already deployed and that fixed and cellular networks share the same layer 2 access or/and aggregation network. In this case, RGWs are connected to the aggregated base station through a layer 2 access network. It can be an Ethernet link as shown in Figure 20-b or any type of transport network.

4.3.2 Offload initialisation and release

Figure 21 shows the different steps to initialise the offload process to a Wi-Fi CeAP. Based on periodical measurements and location information sent by the UE and on the CeAP information from the CeAP database, the decision engine is able to identify the CeAPs that are relevant to the UE. The data path creation and destruction module is then responsible for sending the different information (SSID, BSSID, frequency,…) to the UE. With this information, the UE tries to attach to the CeAP even if it does not detect it yet. If the attachment is successful, the UE sends a test message to the eNB through the CeAP (over Wi-Fi). Upon reception of the test message, the decision engine computes the new policies, sends them to the UE. Wi-Fi and LTE paths are then bonded. The UE directly applies the new policies and can start using the CeAP. Note that Wi-Fi is only used for the data plane. All control-plane

messages are systematically transmitted over LTE. Note also that ANDSF can be used to define and send the policy to the UE.

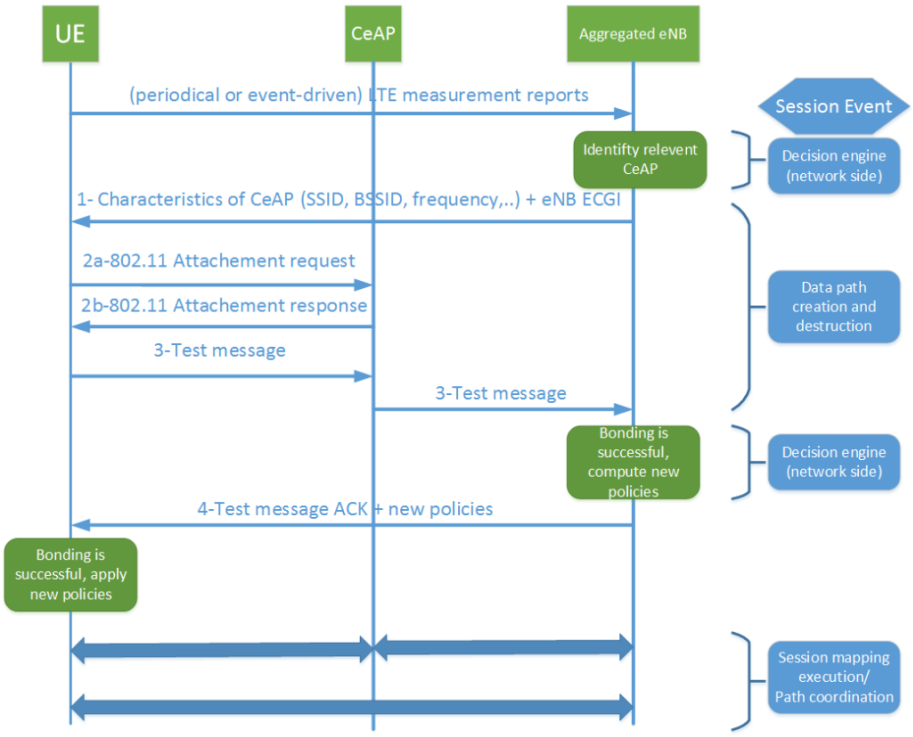


Figure 21: Message sequence chart of offload activation with very tight coupling

The frame error rate on Wi-Fi is monitored. When it exceeds a given threshold, a session event is triggered: the decision engine computes a policy that is adapted to the new situation and transmits it on LTE; the Wi-Fi leg is released by the data path creation-and-destruction entity.

4.3.3 Migration path

Very tight coupling can be deployed with legacy eNB (in which BBU and RRH are co-located). A transport service should thus be available between the CeAP and the eNB site. This can be considered at the firs step of the deployment process. In that case, the route for Wi-Fi can be non-optimal because the access and aggregation network can be crossed twice by the data path: a first path is defined between the CeAP and the aggregated eNB while a second path is defined between the eNB and the SGW (see Figure 22).

When the very tight coupling architecture is fully deployed, a BBU hotel located at the main or the core CO is the point of convergence between all access technologies (see Figure 20-a). In that case, all RGWs are easily connected to aggregated base station. Better performance is expected compared to step 0 as the decision policy and all related functions (data path creation and destruction, path coordination) are centralised in the aggregated base station.

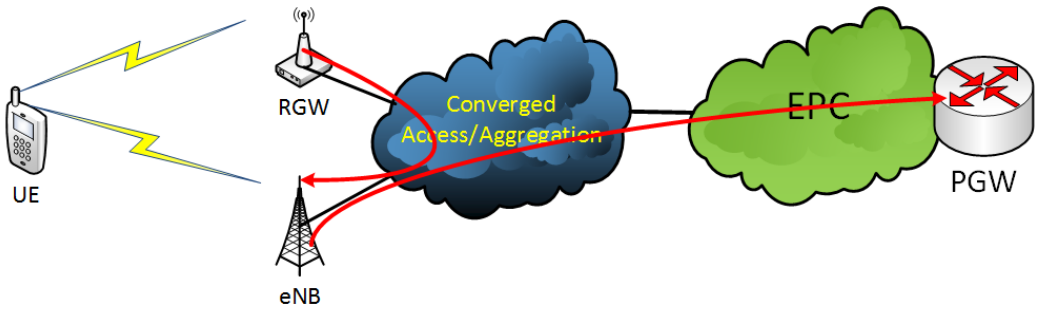


Figure 22: Deployment of very tight coupling: step 0

4.4 Smooth SIPTO-based mobile access

Smooth SIPTO-based mobile access is a new set of procedures that has been developed by COMBO in order to maintain session continuity in case of user mobility between different LGWs. It focuses on offloading LTE traffic over the fixed network by relying on a HeNB co-located with RGW.

In this scenario, a LGW is part of the RGW and LIPA or SIPTO can be used to route part of the mobile traffic on the fixed network through the LGW. In a SIPTO framework, the UE routes regular LTE traffic through the HeNB towards EPC while offloading part of its IP traffic through the co-located LGW. This is in particular appropriate for downloading high bitrate flows (e.g. streamed video) as the available bandwidth on fixed access is usually larger and more stable than the one on the mobile access. However, if the user moves from one source LGW to a target LGW, a gateway relocation procedure is provided by the MME, which results in losing the IP address allocated for the UE by the source LGW. The MME disconnects the impacted SIPTO PDN connection with re-connection clause required during which the UE gets a new IP address from the target LGW [23]. Since the VoD server is unaware that the source IP address and target IP address correspond to the same user, the download is interrupted between the deactivation and reactivation procedures, and the application may have to resynchronize its streaming mechanism.

4.4.1 Implementing SIPTO-based mobile access

The first enabler for a Smooth SIPTO-based IP access is relying on MPTCP to connect the UE and the server in the IP backbone. The MPTCP connection will enable the server to synchronize the user's traffic, when the user relies on multiple IP addresses distributed on MPTCP sub-flows.

The second enabler for a Smooth SIPTO-based IP access is modifying the role of the LGW. SIPTO considers a LGW as a PGW with some additional functions, and not as a co-located SGW/PGW. This is because a UE can be connected to only one SGW at a time [23]. In order to ensure session continuity in case of mobility, we have to circumvent this restriction. We thus introduce a new functional entity, the "Proxy-SGW" (see Figure 23). A Proxy-SGW is a 3GPP gateway that is seen as:

- a SGW by both the HeNB and the LGW
- a HeNB by the SGW identified for the UE by the MME.

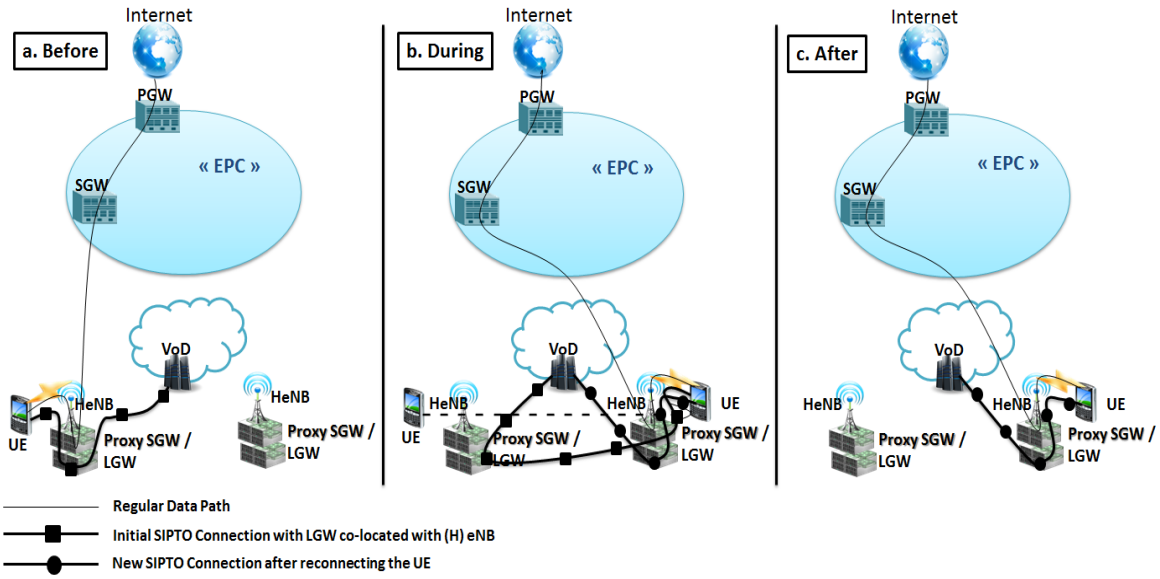


Figure 23: Providing session continuity in a SIPTO-based mobile access

4.4.2 Setting up an MPTCP connection between UE and server

We now explain how the MPTCP connection is setup between the UE and a server, and how the various available data paths are used. This is illustrated in Figure 24.

1. The UE first receives an IP address by the default LTE PGW; let us call this address “default IP address”;
2. Using the default IP address, the UE connects to the VoD service and is pointed to an appropriate VoD server. We assume here that this server is MPTCP-capable. This data path, build between the UE and the server, shall be used for all signaling messages related to the MPTCP connection; it is called “default data path”; however, it is not used to download the stream from the server.
3. Over the default data path, the UE establishes an MPTCP connection.
4. The UE requests the establishment of a SIPTO data path to the server, and thus obtains another IP address, called “local IP address”.
5. The UE sends its local IP address to the server that updates the list of addresses it uses to communicate with the UE.
6. Using the MP-Join option of MPTCP, the UE requests the creation of an MPTCP sub-flow between the server and this local IP address.
7. The UE then declares, with the MP-PRIO option of MPTCP, the sub-flow over the default data path as “backup path” and the sub-flow over the SIPTO data path as “regular path”.
8. Downstream traffic from the server arrives to the UE through the regular MPTCP path.

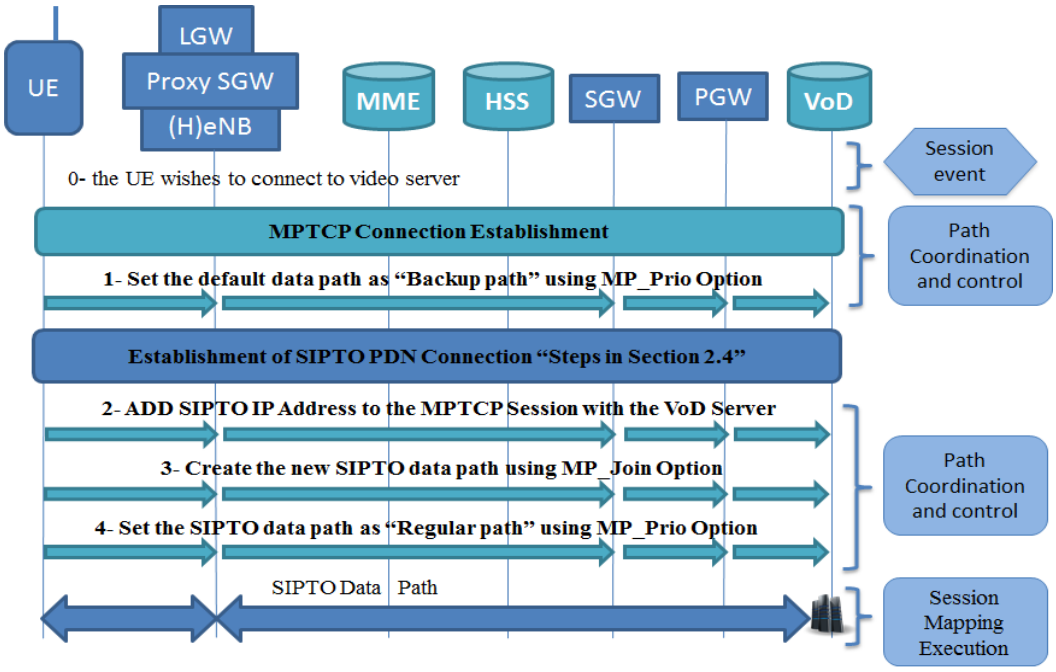


Figure 24: Establishing an MPTCP connection in the SIPTO-based mobile access scenario

4.4.3 Ensuring session continuity for SIPTO-based mobile access

Assume that the UE has to move from a source HeNB to a target HeNB as illustrated in Figure 23; this implies that both proxy-SGW and LGW have to be relocated (from "source" to "target" locations), and that the MPTCP regular path has to be modified. The successive steps are illustrated in Figure 25.

1. When the UE starts its mobility procedure, an S1-based HO procedure [23] is initiated to forward the traffic received from the server by the source Proxy-SGW (S-Proxy-SGW) to the target Proxy-SGW (T-Proxy-SGW). Packets received by the source HeNB are transiently stored within the HeNB and then forwarded to the Target HeNB when the tunnel is established.
2. When the UE is attached to the target HeNB, it exchanges traffic with the server thanks to this tunnel.
3. The UE now initiates a new SIPTO connection to the server using it new local IP address, which is allocated by the target LGW.
4. Over the default data path, the UE sends its new local IP address to the server that updates the list of addresses it uses to communicate with the UE. Using the MP-join option, it requests the creation of a new "regular path" between the server and this new local IP address.
5. The server now sends its traffic over the two regular paths.
6. At the completion of the HO procedure, the tunnel is disconnected, and the UE can ask the server to remove its initial local IP address from the list of available addresses.
7. The data traffic between server and UE now use the new SIPTO connection.

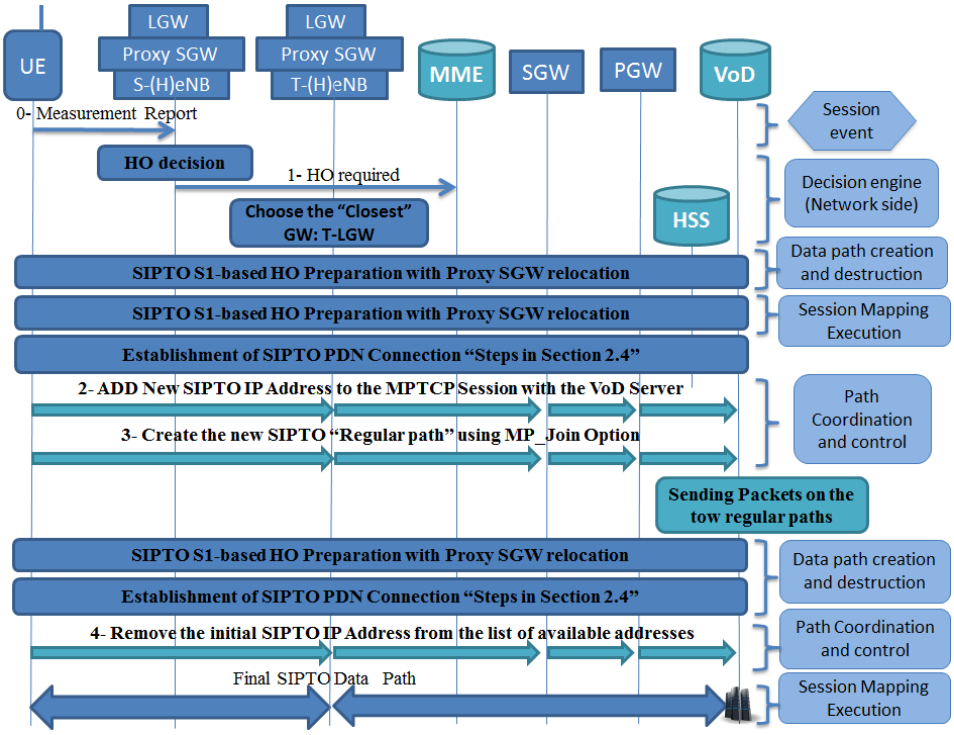


Figure 25: Ensuring session continuity in SIPTO-based mobile access

4.5 Reactive content placement

The objective of the Content Delivery System (CDS) is to take advantage of the FMC network to optimize the QoS delivered to end-users.

4.5.1 High level description or reactive content placement

An entity called Cache Controller (CC) is used to enable the interaction between the Decision Engine and the CDS. Specifically, it knows the current content location, and it is able to control the behaviour of cache nodes in the network. There are different cases where the CDS can interact with the Decision Engine to improve the QoS.

The first case is to help the optimal selection of UE interface for traffic offloading. Since the CC keeps the mapping of all the content available in the caching system and its location, it can provide the information to the Decision Engine. When a user is asking for a piece of content that is cached in a cache node embedded in the Wi-Fi network via mobile network, according to the content location information provided by the CC, the Decision Engine can modify the UE policy so that the user will retrieve the content from the local cache via Wi-Fi network and the traffic in the mobile network is offloaded.

The decision made by the Decision Engine can also impact the content storage policy. For example, when an offloading decision is taken, before sending the directive to UE, the Decision Engine will first inform the CDS so that it is aware of the upcoming handover process and which access point the UE will connect to. Then, the CDS can make the cache node that is close to the access point pre-fetch the content that will be requested by the user. In this case, when the user switches to the

new access point, the content is already available in the local cache and can be accessed immediately.

In these cases, the Decision Engine has the optimisation of the CDS marked as one of the targets in its configuration.

Two technical approaches can be envisaged to realise the CDS, which yield slightly different system architectures, and have pros and cons. The first one is a SDN based solution while the second one is Content Centric Networking (CCN) based.

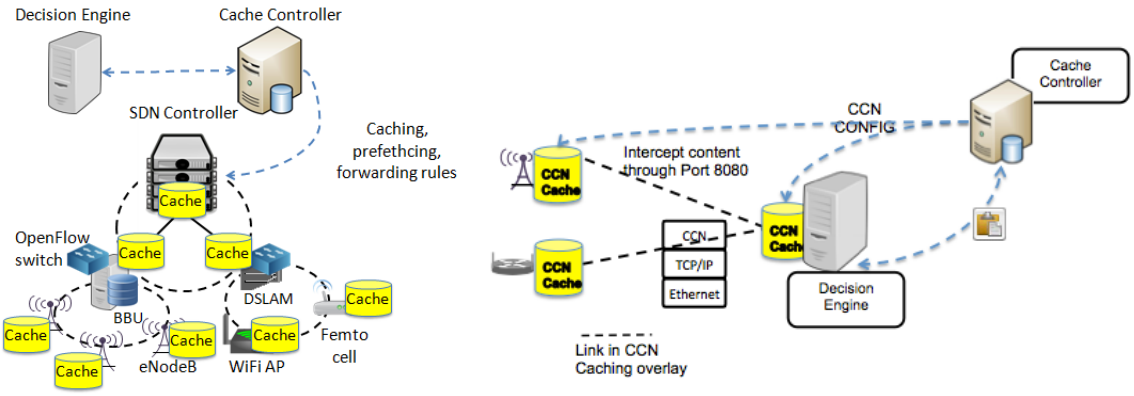


Figure 26: Content Distribution Service architectures (a) SDN based (b) CCN based

Figure 26(a) represents the SDN based solution. Cache nodes are deployed on various network equipment to form a hierarchical caching system. On each cache node, there is an entity called Local Management Primitive (LMP) that manages the local storage, listens to the CC and reports to the CC the content location update. The CC maintains the mapping of content and its location (i.e., in which cache node the content is stored). The Decision Engine listens to the monitoring module and session events, takes decisions, and sends the decisions to the CC. Upon receiving the information, the CC updates the caches according to the decisions. Finally, the caching plan is transformed into forwarding rules and caching (pre-fetching) directives and sent to the SDN controller to deploy the new cache settings. The SDN controller is responsible for setting the appropriate forwarding rules in the OpenFlow-enabled switches so that the user requests will be forwarded to the right cache nodes. It should also be able to configure caching policies applied by cache nodes via Netconf protocol.
Figure 26 (b) depicts the CCN based solution. A CCN cache overlay is build on top of the access network that intercepts the user requests. The cache overlay transforms the user's request to a CCN Interest and processes it using CCN protocol [60]. If no required content is found in the cache overlay, the request will be sent to Internet. Again the CC interacts with the Decision Engine and LMP to deploy content, and modify the caching policies applied on each node. New message types that do not exist in the original CCN design are introduced to realize the cache control, particularly, the *Config/Control* message for caching policy control and the *Config/Prefetch* message for pre-fetching control. Moreover, in CCN network there is no need to have SDN controller that configures the forwarding of the switching element to route the users' requests since it is tackled automatically by CCN protocol.

In the SDN based approach, the cache nodes offer standard Netconf-based control interface so that the CDS can be easily orchestrated by any SDN controller that

implements Netconf client. However, the SDN controller has to configure the switching equipment to route user request, which is an extra-overhead. On the other side, in the CCN based solution, the CC does not take care of routing. But specific software has to be installed in the CC to realize the control over the cache nodes. Another possible approach is to combine these two solutions, i.e. to control the CCN cache nodes via SDN controller.

4.5.2 Flowchart for reactive content placement

In order to better illustrate the proposed solutions, we specifically focus on the CCN based solution. A flowchart regarding content distribution is illustrated in Figure 27.

When a cache node is initiated, the LMP is registered to the CCN daemon (a CCN software implementation). Thus, the CCN daemon knows the face toward the LMP. At a certain moment, the Decision Engine detects some session event and makes an interface switching decision for an end user. This decision, including where the end user will connect to, and which content is currently required by the user, are sent to the CC. Then, the pre-fetching process is triggered by sending the Config/Prefetch message to the cache node close to the access point to which the user will be connected, and the message is forwarded to the LMP. Once the pre-fetching directive is received by the LMP it sends a CCN Interest to retrieve the content. Thereafter, the content is sent back to the LMP and cached in both of the CCN daemon and LMP. This is to make sure that even if the content cached in CCN daemon is evicted according to certain replacement policy, the content can always be retrieved from LMP. When the content stored by the LMP, an acknowledgement is sent back to the CC, so that the CC can register the {content:location} mapping. The CC will inform the Decision Engine about the mapping. This way, the Decision Engine can take into account the content location when it makes decision next time.

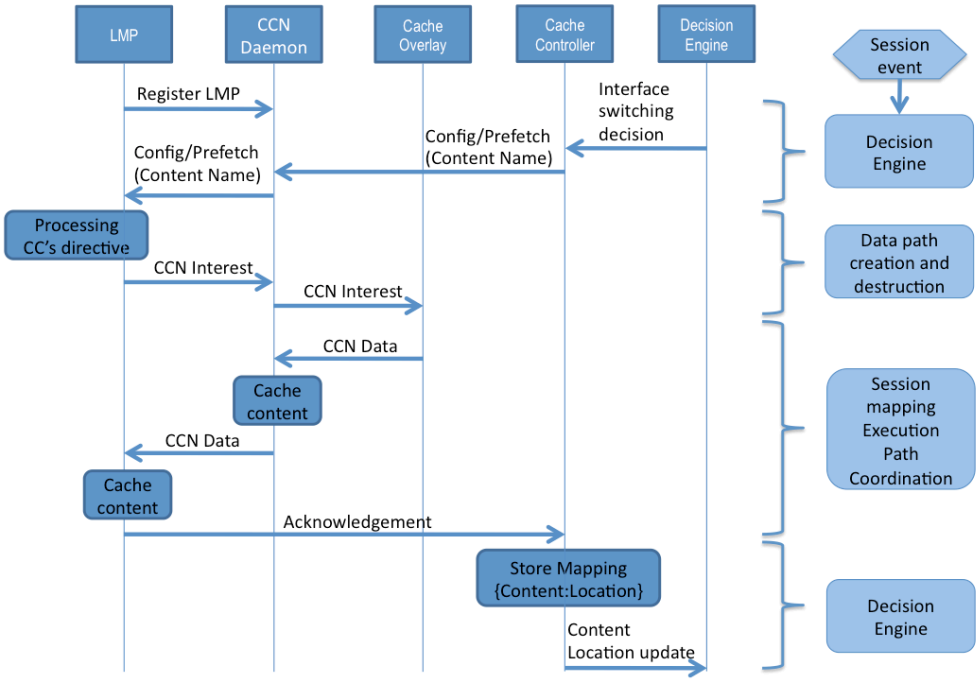


Figure 27: Flowchart for CDS, CCN based approach

4.5.3 Migration path

The innovation in the proposed CDS is that the system offers full control of the behaviour of cache nodes. To the best of our knowledge, current study about SDN based in-network caching system [59] can provide only limited control over cache nodes (e.g. no control on the replacement policy applied by cache nodes). For CCN, only transparent caching is available in the current implementation [60]. Following is the migration path for the development of CDS.

- Realise a full control over both the SDN based and CCN based CDS.
- Combine the two approaches. That is to use SDN to control CCN cache node.
- Finally, evaluate the performance of different approaches and select the most profitable one.

4.6 Conclusion of Section 4

Section 4 has addressed the second HT that has been identified as key to providing FMC capabilities to network operators operating fixed, mobile and Wi-Fi networks. HT2 is intended to provide solutions for “Advanced Interface Selection and Route Control”.

We proposed to fulfil HT2 thanks to a set of functional blocks that realise a “Universal Data Path management” (uDPM). The objective is to allow a FMC network operator to take advantage of all access resources to better serve the users. As the data paths operated by fixed and Wi-Fi access networks are static, the issue is to better accommodate mobile data traffic, by redirecting (part) of this traffic over the fixed/Wi-Fi data paths from the default LTE path. A major issue is to maintain session continuity, in particular during mobility. This implies finding alternatives for the current implementations of the following functions: AAA (specifically addressed in HT1), handover and mobility support, forwarding (and especially tunnelling), charging and route control.

A generic specification of the uDPM has been proposed, and described. The uDPM is triggered by any “session event” (application launching, mobility event, policy rule activation, etc.). The uDPM is composed of four functional blocks. Its intelligence is located in the “Decision Engine”, which relies on monitoring information, on user or subscriber’s profile and on network policies to select how the session is to be mapped on data paths. The problem of identifying the appropriate decision to take upon a session event in a given context can be modelled as an MCDM, and each network operator is likely to select its own resolution method.

Once the decision has been taken, its enforcement in the network depends on three slave functional blocks:

- Data path creation and destruction, as it may be necessary in some cases to create new data paths, and to destroy others. In the upstream direction, it is likely that the UE shall have partial control of path creation/destruction using its various interfaces, although the network operator by remotely activating interfaces, and allocating IP addresses, shall also participate in this control. A fuller control of the data paths in the downstream direction by the network

operator is to be expected (e.g. selection of a given server, or of a given data path between the server and the UE).

- Path coordination and control, which ensures that concurrent data paths smoothly deliver the packets corresponding to the session, and that session continuity is guaranteed, even in case of UE mobility. An important element, the MPE has been identified, which is in charge of synchronizing session data carried over multiple paths.
- Session Mapping Execution, which applies the session mapping decision taken by the Decision Engine. Packets corresponding to the session are filtered and forwarded on the data paths selected by the decision engine.

Lastly, the high level description of uDPM solutions to three specific sub-problems is given: very tight coupling between Wi-Fi and LTE access, which allows a user to seamlessly move from LTE to Wi-Fi, smooth SIPTO-based content distribution, which allows a user to seamlessly stream video traffic thanks to LGWs, and reactive content placement, which is based on the Content Distribution Service (CDS) management to react to user location in order to improve content placement. In each case, the high-level implementation principles of the particular uDPM solution are given. In some cases, functions are implemented in the network while others are implemented in the UE or are distributed between both network and UE.

In order to demonstrate some of the concepts described in Section 4, different activities are being carried out within WP6. For example, the Decision Engine concept is to be demonstrated by showing the relationship with the rest of the uDPM modules and by realising the required output actions necessary to achieve the optimisation goal sought. Additionally, a dual attachment procedure is expected to be available so that the UE can take advantage of having a multiple interface attachment to the network. Finally, the reactive content distribution will be in place so that content is available from the CDS (or cache node) when the UE requires it.

5 Application of HT1 and HT2 to WP2 use case instances

This section analyses the impact of HT1 and HT2 on the WP2 use cases defined in D2.1. As these horizontal targets do not have the same impact on all use cases, this section only focuses on the use cases 1, 2, 4 and 8, which are the most relevant from the HT1 and HT2 points of view. For these use cases, it is analysed how uAUT and uDPM approaches can be used and the benefits they can bring in the scope of HT1 and HT2 are identified.

5.1 UC1 – Unified FMC access for mobile devices

UC1 allows mobile devices to use Wi-Fi access in combination with mobile access in an FMC network with advanced cooperation. Three main aspects are considered in this cooperation between Wi-Fi and mobile networks: simultaneous attachment (including transmission on several networks at the same time), seamless handover, and smart network assistance for the selection and utilisation.

Figure 28 shows how UC1 could take benefit of the uAUT and uDPM approaches such as described in Sections 3 and 4. In that Figure the coordination point in the network side deals with the user data traffic coming from all access networks and relies on uDPM. uAUT has two main types of interfaces: the interface to the provisioning system and the interfaces towards the Wi-Fi and mobile network elements interfacing AAA (e.g. the AP controller in case of Wi-Fi or the MME in case of LTE).

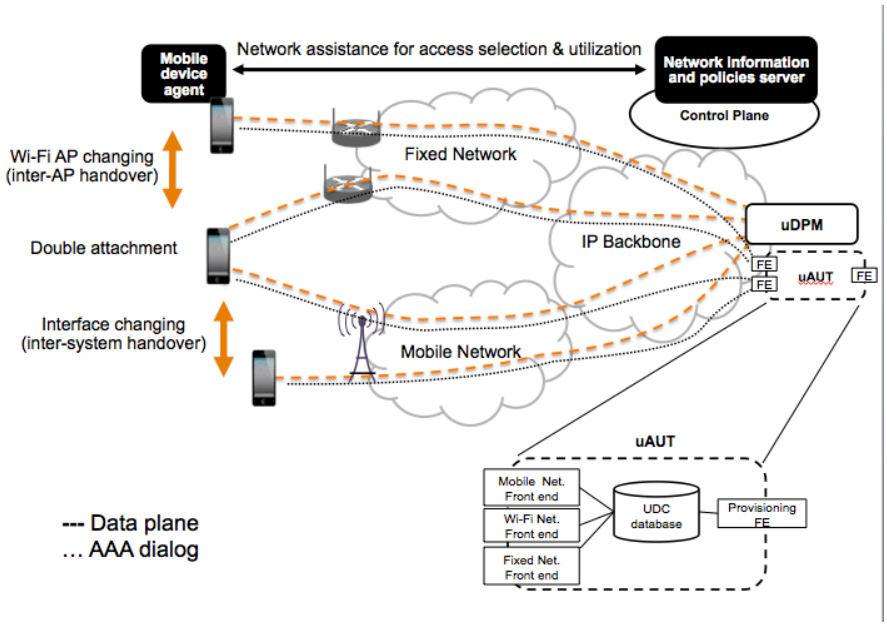


Figure 28: uAUT and uDPM in the context of UC1

The uAUT performs the following FMC activities in this use case:

- Authenticates Wi-Fi and mobile users (it could also authenticate Wi-Fi terminals using USIM credentials).
- Allows the uDPM to associate the traffic arriving in the MPE from both networks to the appropriate user.

- Collects and assigns the accounting data coming from Wi-Fi and mobile to the correct terminal and subscriber, thus enabling a global accounting scheme for LTE and Wi-Fi traffic.

The uDPM described in Section 4 performs the following FMC activities:

- It relies on the monitoring of the different access points (Wi-Fi, cellular) and of the access and aggregation networks. The measurements (load, current topology, etc.) may trigger events and feed the decision algorithm.
- It enforces, through the decision engine, network management rules for network selection and utilisation, including rules for offloading, dual use, load balancing and service restrictions for each user in both Wi-Fi and mobile networks. The decision engine specifies which interface can be used, ensuring that at each instant of time the best interfaces are used, and triggers data path creation and destruction.
- The data path coordination ensures that session continuity is provided in case of load balancing and creation/destruction of data paths.

5.1.1 Benefits of UC1 to the FMC operator

The FMC operator can implement a global policy for optimising resource usage over LTE and Wi-Fi network types. The global policy is proprietary to the FMC operator, which may choose e.g. to minimize cost, to maximise resource usage, to minimize energy consumption, etc.
The FMC also has a global view of its subscribers, can support service continuity over multiple network types, and offer global charging processes.

5.1.2 Benefits of UC1 to the user

By selecting the “best” data path, the network operator can provide a better QoS to the user.
The user can benefit from a global service offer, irrespective of the network which carries the data path; this is in particular valid for services proposed by OTT service providers which have a specific contract with the FMC network operator to distribute its services.
The generalisation of traffic offload shall permit the subscriber to benefit from advantageous charging schemes, in which expensive communication modes (e.g. LTE) are used only when cheaper modes (e.g. Wi-Fi) are unavailable.

5.2 UC2 – Converged Content Caching for unified service delivery

The objective of UC2 is to build an in-network caching system and optimise the content distribution system (CDS) in FMC. A convergent content caching architecture is illustrated in Figure 29

Both uAUT and uDPM are essential in enabling UC2. Concretely, uAUT performs the following FMC activities in this use case:

- It enables the seamless handover between different types of networks, which is the base of seamless experience in content delivery.
- It provides the unified SLAs and user profiles so that the CDS manager can ensure the consistency of use QoS in different networks.

uDPM interacts with the CDS manager to realize an optimised content placement.

- It tells the CDS manager when and where the end user will connect so that the CDS manager can push content into the corresponding place and guarantee the seamless experience of content delivery service.
- It informs the CDS manager about the offloading policy so that the CDS manager activates/deactivates cache nodes, changes caching strategies according to the policy, etc.

uDPM is also used to provide a seamless handover for CDS, taking into account user preferences known through uAUT.

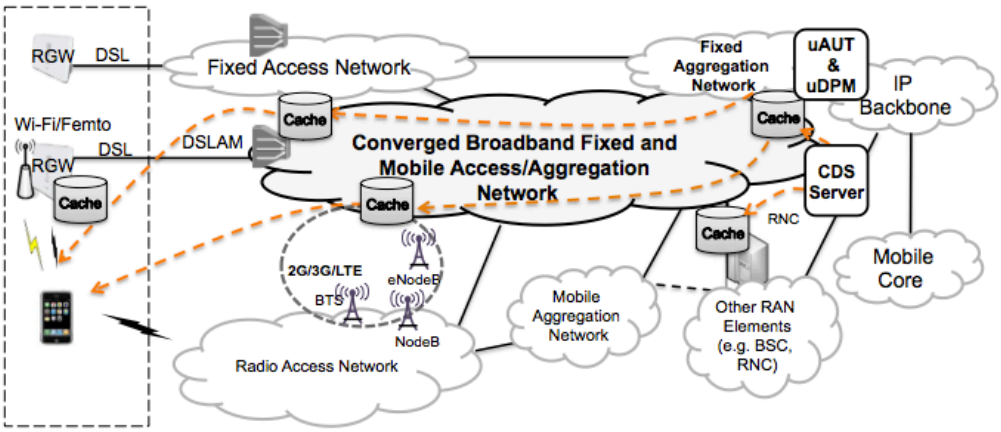


Figure 29: uAUT and uDPM in the context of UC2

5.2.1 Benefits of UC2 to the network and content distribution service providers

The FMC operator operating a content distribution service can mutualize its server/caching architecture for all its subscribers (fixed, mobile, hot-spot users); the FMC operator can also monetise a global distribution service to OTT CDS providers.

The OTT CDS provider can negotiate, by a single SLA with an FMC operator, direct access to its customers through every available interface.

The FMC operator can both optimize access and distribution network resources by selecting the better access data path, and the better server/cache to distribute a given content to a FMC user.

Optimizing both content storage and network resource enables the FMC operator to offer attractive charging plans to its subscribers.

5.2.2 Benefits of UC2 to the users

The user may indirectly benefit from a converged content distribution architecture, whether it is used by its FMC operator for in-network CDS, or by an OTT CDS, in terms of a higher QoS and a better charging policy.

5.3 UC4 - Universal access bundling for residential gateway

UC4 aims at fixed, mobile and Wi-Fi convergence using an integrated CPE able to provide each user the optimum bandwidth resource, dynamically assigned through the available access technologies (mainly for residential customers in rural areas). Figure 30 shows how uAUT and uDPM can be used in UC4. In that figure some functions from the uDPM may be centralised (e.g. the decision engine), while other blocks (Data Path Coordination and Session Mapping Execution) are distributed between the MPE within the network (which works as the Hybrid Connection Gateway (HCG) identified in UC4) and functions from the Session Mapping Execution within the RGW.

The uAUT entity has interfaces with the service provision, uDPM entity and with the mobile, Wi-Fi and fixed elements involved in AAA tasks.

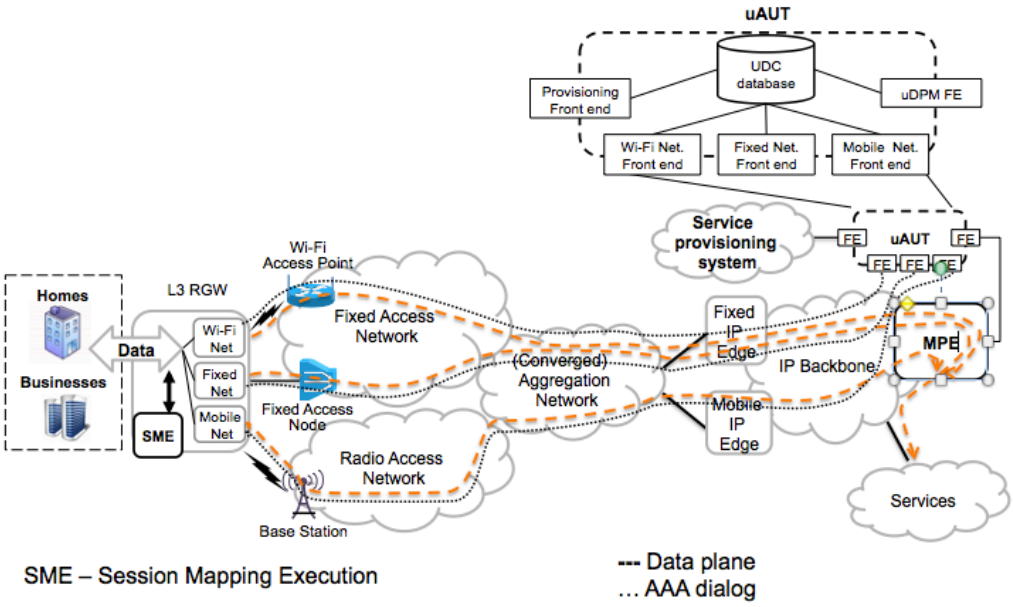


Figure 30: HT1 and HT2 in the context of UC4 (uDPM can be located elsewhere in the networks)

The uAUT performs the following FMC activities in this use case:

- It authenticates the fixed subscriber and the wireless modules of the RGW as their credentials are stored in the authentication database. Additionally, enhanced automatic authentication mechanisms can be easily implemented, such as Wi-Fi or xPON modules authentication using the USIM credentials.
- It presents a single interface towards the repository for the rules and policies for the network selection and utilisation function.
- It collects and assigns the accounting data to the correct RGW and subscriber, allowing a uniform accounting if needed in this case (e.g. a paid option for DSL subscribers for a bandwidth on demand service).

The uDPM entity performs the same activities as for UC1 and includes the fixed access through the RGW in addition to wireless (Wi-Fi and cellular) accesses in the available data paths. The decision engine identifies how to split sessions over these

data paths and the Session Mapping Execution functional block is distributed between the RGW and the MPE (within the network).

5.3.1 Benefits of UC4 to the FMC operator

They are similar to the ones of UC1; they are however enhanced by the availability of a direct access to the fixed aggregation network through the RGW.
Moreover, as the network operator controls the RGW, it is easier to implement functions within the RGW to implement uAUT and uDPM.

5.3.2 Benefits of UC4 to the users

They are similar to the ones of UC1 with the same enhancement identified for the network operator (fixed line data path).
The user can thus benefit from automatic failover in case of failure from one network type: e.g. if the fixed line fails, the LTE connection may still be available.

Moreover, in areas where the uplink and downlink rates on the fixed line are small, the user benefits from an increase in uplink and downlink rates by forwarding its traffic on fixed, Wi-Fi and LTE data paths.

5.4 UC8 – Network sharing

UC8 enables multi-operator network capabilities in future FMC networks. These capabilities allow to reduce deployment and operation costs and to support more flexible business models by sharing existing infrastructure for both fixed and mobile communication. Several network-sharing scenarios are illustrated in Figure 31.

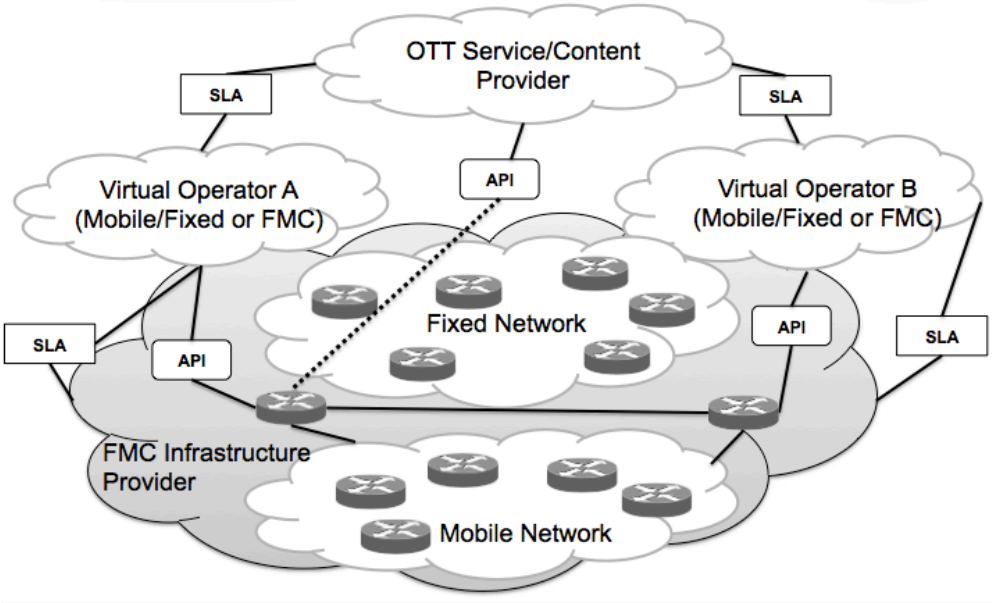


Figure 31: Caching Based Network Sharing Scenarios for FMC Networks

A typical scenario consists in an infrastructure network operator providing network resources to an operator which does not directly operate a given network type; for example, a network operating a fixed network with a set of Wi-Fi hotspots may be a virtual mobile network operator, and thanks to the uAUT and uDPM operate as a full FMC network.

A similar scenario consists in network operators of different network types providing a merged offer to their subscribers, relying on resources from both networks

Other scenarios can be identified, that involve operators and providers which do not operate first/last mile data paths to subscribers.

- An OTT service provider may offer a unique service supported over multiple network operators.
- A Content Distribution Network (CDN) operator may interwork with network operators in order to improve the performance and efficiency of content distribution in the first/last mile.

In the context of network sharing, uAUT is central as it allows sharing authentication data and functions between different operators.

uDPM allows to optimise the usage of different data paths operated potentially by different operators.

This unified framework simplifies not only the networking functions, but also the content-related functions. The set of functions in uDPM make it possible for subscribers from different service providers, accessing the network via various access networks, to receive content managed by a single CDS. Access to the CDS is controlled by the uAUT, whereas data path forwarding is controlled by the uDPM. This, in turn, enables higher flexibility for mobile and fixed operators in their business relationship with content providers.

5.4.1 Benefits of UC8 to network operators, OTT and content distribution service providers

A network operator operating some type of first mile/last mile access can enhance its offer to a full FMC service either by relying on another network operator, which operates the network types that are originally missing, or on an infrastructure operator.

An infrastructure operator can monetize its assets by offering bundled services to virtual network operators.

An OTT service provider can offer a given service over a large geographical area by contracting with multiple access network operators.

A CDN operator may extend its reach by collaborating with network operators who operate first/last mile access lines.

5.4.2 Benefits of UC8 to the users

A user may benefit from a full FMC subscription although its original network operator does not directly control all required resources.

A user may benefit from a given OTT service although he is outside the coverage of its original operator.

Lastly, content distribution performance may be improved and made cheaper were the storage/caching infrastructure shared between multiple service providers.

5.5 Conclusion of Section 5

This section analyses how the technical solutions proposed for HT1 (in Section 3) and HT2 (in Section 4) can enable some of the FMC use cases defined in D2.1, namely UC1, UC2, UC4, UC8 [2].

5.5.1 UC1 – Unified FMC access for mobile devices

uAUT and uDPM enable UC1 by allowing mobile devices to use Wi-Fi access in combination with mobile access in an FMC network with advanced cooperation. This includes simultaneous attachment (including transmission on several networks at the same time), seamless handover, and smart network assistance for the selection and utilisation. More specifically, uAUT ensures authentication of Wi-Fi and mobile users, enables a single coordination point, includes the rules for smart selection and utilisation of interfaces and allows uniform accounting for terminals with integrated Wi-Fi and mobile connections. Also, based on access point monitoring and network monitoring, the Decision Engine of uDPM decides which interfaces can be used according to the rules given by the uAUT, and ensures that at each instant of time the best interfaces are used. The Path Coordination and Control block of uDPM ensures session continuity in case of creation/destruction of data paths.

5.5.2 UC2 – Converged Content Caching for unified service delivery

Technical solutions for HT1 and HT2 enable an in-network caching system ensuring optimised content distribution. uAUT enables a smooth transition between different service qualities of content delivery service in different networks and on various terminals. It also ensures consistency of user preferences, network policies and service qualities of content delivery received by end user in different networks. uDPM tells the CDS manager when and where the end user will connect so as to allow content to be pushed to the right place. It also informs the CDS manager about the offloading policy so as to allow appropriate activation of cache nodes and changes in caching strategies.

5.5.3 UC4 - Universal access bundling for residential gateway

UC4 can first be seen as an extension of UC1, in which a fixed line access is added to the LTE and Wi-Fi accesses. All benefits brought by UC1 directly apply to UC4.

Moreover, as the network operator controls the RGW, a better distribution of control and data plane functions can be made.

5.5.4 UC8 – Network sharing

Universal subscriber and user authentication can be regarded as one of the basic enablers of network sharing, since it allows virtual operators to manage their subscriber connections through different access networks and share the infrastructure provided by the wholesale FMC operator. uAUT must know the subscribers' profiles created by virtual operators and allow their management via an API. The relationship among the different subscribers' bases should be under the control of a SLA and known by uAUT. By managing all subscribers' profiles and virtual operator relationships, in coordination with other control plane entities, uAUT

allows a user session to be moved transparently from one operator to another. Also, uDPM enables a shared content delivery system involving OTT service/content providers, virtual operators and FMC infrastructure provider.

6 Conclusion

The present D3.2 deliverable describes and develops the proposals made by COMBO project to solve two FMC Horizontal Targets entitled “Converged Subscriber and Session Management” (HT1) and “Advanced Interface Selection and Route Control” (HT2). It describes the problems to be solved by these HTs, requirements for the solutions and proposes possible target solutions for these HTs, called respectively universal subscriber and user Authentication (uAUT) and universal Data Path Management (uDPM). The key outcomes of these developments and analyses are summarized in the following.

Horizontal Targets as intermediate goals for FMC

As explained in the introduction of the document, FMC architectural targets can be of two types:

- The Horizontal Targets (HT) are consistent sets of FMC generic functions allowing functional convergence and solving some key “horizontal” end-user issues in an FMC context;
- The Network Scenarios (NS) specifically focus on the organisation and overall architecture of the FMC network.

The HTs thus focus on advanced functional features of future converged networks. These functional features can be implemented differently depending on the supporting FMC Network Scenario and are required whatever the FMC Network Scenario. This is why they have to be described and developed in details before the analysis of NS themselves.

The derivation of two HTs is based on previous work of COMBO Work Package 3, and in particular on the identification by D3.1 of main functional groups, which need significant effort to reach functional convergence. The main characteristics of these two HTs are the following:

- **HT1: Converged subscriber and session management:** This horizontal target addresses convergence of authentication and subscriber data management. It aims at avoiding the drawbacks of proliferation of identities, subscriber profiles and authentication mechanisms in fixed and mobile networks;
- **HT2: Advanced interface selection and route control:** The motivation of this second horizontal target is to provide the FMC network operator with means to dynamically control how traffic is forwarded on data paths, when several data paths are available, while maintaining session continuity whenever necessary.

HT1 and HT2 features are shown to solve most of the functional gaps existing between the current non-converged situation and a situation where an FMC operator can take advantage of a global control of its resources and a global management of its subscribers.

Several initiatives already address the scope of HT1, including User Data Convergence (UDC) and “non-3GPPP access architecture” developed by 3GPP, or

Hotspot 2.0 proposed by Wireless Broadband Alliance (WBA). These initiatives lead to subscriber convergence in the sense that there are interactions between subscriber databases, but there is no global view of a given user's identities for the benefit of a single FMC operator. Also, Hotspot 2.0 may not meet the security requirements expected from the mobile network.

Frameworks related to HT2 are multiple, as they build on existing mobility management and/or on multi-homing management. In both cases, an important feature is session continuity as all mechanisms potentially involve either interruptions or de-synchronization. Most already proposed solutions define protocols, but do not take into consideration the architecture of the access/aggregation network.

Technical solutions for HT1 "Converged Subscriber and Session Management"

The solution proposed by COMBO for HT1 leverages on the 3GPP's User Data Convergence concept, namely splitting subscribers' data repository from the application logic specific to each access type. A single functional block, the "universal subscriber and user Authentication server" (uAUT), is proposed by COMBO project as a significant improvement of UDC concept. It would link several application logics (called "Front Ends" in the UDC framework) with a single global User Data Repository (UDR). The proposed solution is described in details and completed with proposals regarding the optimization of database access and Front End scalability. Convergence of authentication mechanisms themselves relies on Extensible Authentication Protocol run over layer 2, allowing negotiation of security mechanism between the authenticator and the supplicant, according to the requirements of the network and the capabilities of the device. A migration path for implementing these HT1 solutions in a realistic manner is also presented.

Technical solutions for HT2 "Advanced Interface Selection and Route Control"

So as to fulfil HT2 requirements, COMBO project proposes a functional block that realises a "Universal Data Path Management" (uDPM). It allows redirecting (part) of mobile data traffic over the fixed/Wi-Fi data paths from the default LTE path, while maintaining session continuity (even during mobility) and while allowing multiple data paths to be used simultaneously for a given user session. uDPM solution includes alternatives for the current implementations of handover and mobility support, forwarding (and especially tunnelling), charging and route control. It also leverages on universal subscriber and user Authentication solution proposed by HT1.

A generic specification of the uDPM is proposed and described. The uDPM would be triggered by any "session event". It includes a "Decision Engine", which hosts intelligence of uDPM and relies on monitoring information, on user or subscriber's profile and on network policies, to select how the session is to be mapped on data paths. Three slave functional blocks depend on the Decision Engine:

- Data Path Creation and Destruction, allowing control of path creation/destruction on the available interfaces;
- Path Coordination and Control, ensuring that concurrent data paths smoothly deliver the packets corresponding to the session, and that session continuity is guaranteed, even in case of UE mobility. It includes a Multi Path Entity (MPE), in charge of synchronizing session data carried over multiple paths;

- Session Mapping Execution, applying the session mapping decision taken by the Decision Engine.

While uAUT can be considered as a unique solution for HT1, uDPM is a set of partial solutions to specific problems; three such solutions are identified: “very tight coupling” between Wi-Fi and LTE access, which allows a user to seamlessly move from LTE to Wi-Fi, smooth SIPTO based content distribution, which allows a user to seamlessly stream video traffic thanks to Local Gateways (LGWs), and reactive content placement, which is based on the Content Distribution Service (CDS) management to react to user location in order to improve content placement. In each case, the high-level implementation principles of the particular uDPM solution are given. In some cases, functions are implemented in the network while others are implemented in the UE or are distributed between both network and UE.

How uAUT and uDPM will enable FMC use cases

uAUT and uDPM enable unified FMC access for mobile devices (UC1) by allowing mobile devices to use Wi-Fi access in combination with mobile access in an FMC network with advanced cooperation. More specifically, uAUT ensures authentication of Wi-Fi and mobile users, enables a single coordination point, includes the rules for smart selection and utilisation of interfaces and allows uniform accounting for terminals with integrated Wi-Fi and mobile connections. The Decision Engine of uDPM decides which interfaces can be used according to the rules given by the uAUT, and ensures that at each instant of time the best interfaces are used. Also, the Path Coordination and Control block of uDPM ensures session continuity in case of creation/destruction of data paths.

Technical solutions for HT1 and HT2 enable also an in-network caching system ensuring optimised content distribution (UC2). uAUT enables a smooth transition between different service qualities of content delivery service in different networks and on various terminals. It also ensures consistency of user preferences, network policies and service qualities of content delivery received by end user in different networks. uDPM tells the Content Delivery System manager when and where the end user will connect. It also informs the CDS manager about the offloading policy so as to allow appropriate activation of cache nodes and change of caching strategies.

A RGW equipped with several broadband network interfaces (UC4) would benefit from HT1 and HT2 technical solutions in various ways. uAUT enables single authentication for DSL, LTE and public Wi-Fi interface of the RGW using the same subscriber identity. It enables a single repository for the rules and policies for the network selection and utilisation function. It allows uniform accounting of traffic going through the RGW. Also, uDPM enables actual hybrid access from a RGW.

Universal subscriber and user Authentication can be regarded as one of the basic enablers of network sharing (UC8), since it allows virtual operators to manage their subscriber connections through different access networks and share the infrastructure provided by the wholesale FMC operator. By managing all subscribers’ profiles and virtual operator relationships, in coordination with other control plane entities, uAUT allows a user session to be moved transparently from one operator to another. Also, uDPM enables a shared content delivery system involving OTT service/content providers, virtual operators and FMC infrastructure provider.

How HT1 and HT2 technical solutions will be used in future work

There are possibly multiple methods for fulfilling both HT1 and HT2 to realise an FMC network. In particular, we have previously shown that functional network convergence takes benefit from Next Generation Point of Presence (NG-POP) concept [1]. The NG-POP is a location in the network, where the operator could implement multiple functions, including the IP edge for all network types.

In a later deliverable, we intend to focus on two major architectural options corresponding to two network scenarios (NS). One NS, called Distributed NG-POP, relies on a large number of NG-POP locations at the current main COs, on which the common IP edges and other functional blocks would be distributed, benefiting also from optical access node consolidation. This would lead to an extension of the IP backbone towards the access network. The other NS, called Centralised NG-POP, includes a smaller number of NG-POP locations, typically at the sites of core Central Offices, which are the edges of the current fixed aggregation network. In both cases, the advantages brought by SDN and NFV shall be assessed. Also, as schematically illustrated in Figure 32, implementation and distribution of HT1 and HT2 functional blocks will be key design criteria for deriving architectural blueprints of these two NG-POP flavours.

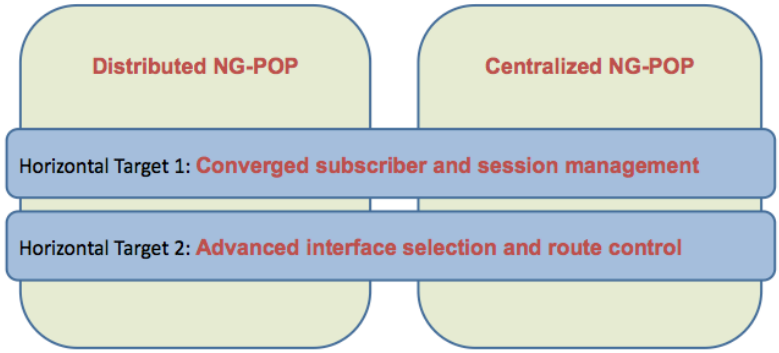


Figure 32: The COMBO approach for providing true FMC, in terms of Horizontal Targets and function distribution

Appendix 1 A control theory based method to solve a MCDM problem

Introduction

Current mobile devices are under the coverage of multi-radio access networks (2G, 3G, 4G, Wi-Fi, etc.) and equipped with multiple interfaces to connect to two or more networks simultaneously. Since mobile applications become more and more bandwidth consuming and mobile users desired the Always Best Connection (ABC), how to select the most appropriate network becomes critical for both users and operators. Network selection is a Multi-Criteria Decision-Making (MCDM) problem, since networks differ in technical characteristic and performance.

Utility theory is one of the approaches which can be applied to solve this MCDM problem. Utility is a term used in economics, which refers to the ability of a good or service to satisfy a human need. In network selection, utility refers to how much one or multi criteria can meet the user's need. The utility value is used to quantify user's preference and users often try to select the network which has higher utility values. To determine the evolution of utility value, utility function U(x) is adopted to describe how network metrics affect the utility value. In [1], different utility functions are depicted in Figure 33.

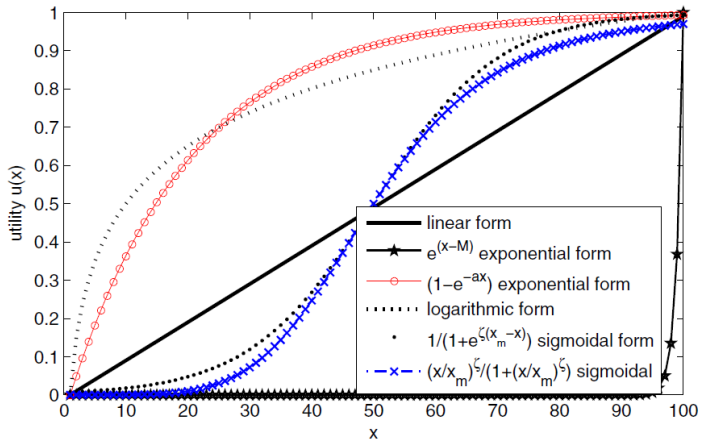


Figure 33: Different utility function patterns

Table 5 is an example for defining a utility function depending on the cost of a service, the Qos delivered by the service and the load on a given network.

Table 5: example of a utility function

Utility	w_i	Network A	Network B
u(cost)	1/4	0.3	0.9
u(QoS)	1/2	0.6	0.1
u(load)	1/4	0.2	0.4
Total Utility		0.425	0.375

In Table 5, the user has two networks available which can be selected and he cares about the following three metrics: cost, QoS and the load. Then he defines the weight of each metric according to the importance of metrics to him and his preference. The elementary utility for each metric of network is computed and the weighted sum then represents the total utility value. Since network A has a higher total utility value, the user will select this network.

However this approach, which simply makes utility sums of different metrics does not react in real time to the rapid changes of network condition. In the following, a control theory based network selection is introduced to solve this MCDM in a scalable, and real-time manner, which optimally distributes the workload in each network and increases QoS to maximize the utility for each users.

System model and controller design

Interworking WLAN (I-WLAN) is taken for an example to show how the control theory is applied to make the optimal network selection decision. Moreover, this approach is not only suitable for I-WLAN but also can be generalized to other multi-radio access networks. Firstly, the evolution of the number of mobile devices in I-WLAN can be modelled as in Figure 2:

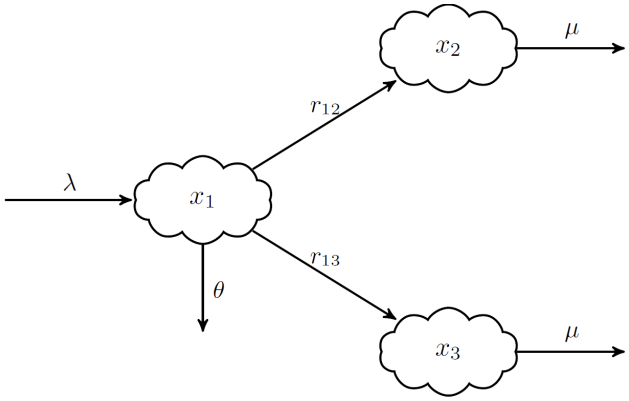


Figure 34: I-WLA N system model

The variables showed in Figure 34 are described as follows:

$x_1(k)$	The number of mobile devices in FMC architecture which communicate with the controller to get the network selection information at instant k;
$x_2(k)$	The number of mobile devices using LTE network for data service at k;
$x_3(k)$	The number of mobile devices using Wi-Fi network for data service at k;
$\lambda(k)$	The number of new mobile devices arriving rate at step k;
$\mu(k)$	The number of mobile devices leaving the states x_2 or x_3;
$r_{12}(k)$	The number of devices steering to the LTE network;
$r_{13}(k)$:	The number of devices steering to the Wi-Fi network;
$\theta(t)$	The number of blocked devices to avoid network overload.

Thus, the system model can represented as follows:

$$\begin{cases} x_1(k+1) = x_1(k) + \lambda(k) - r_{12}(k) - r_{13}(k) - \theta(k), \\ x_2(k+1) = x_2(k) + r_{12}(k) - \mu(k), \\ x_3(k+1) = x_3(k) + r_{13}(k) - \mu(k). \end{cases}$$

The system can be reformulated:

$$\begin{cases} X(k+1) = AX(k) + BU(k), \\ Y(k) = CX(k), \end{cases}$$

with the following variable substitution:

$r'_{12}(k) = r_{12}(k) - \mu(k), r'_{13}(k) = r_{13}(k) - \mu(k), \theta'(k) = \lambda(k) - 2\mu(k) - \theta(k)$

$X(k) = \left[x_1(k) - x_1^{\text{ref}}, x_2(k) - x_2^{\text{ref}}, x_3(k) - x_3^{\text{ref}}\right]^{\mathrm{T}}$, $U(k) = \left[\theta'(k) - r'_{12}(k) - r'_{13}(k), r'_{12}(k), r'_{13}(k)\right]^{\mathrm{T}}$.

Where A=C=the identity matrix of dimension 3, and the matrix B is $\begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. According to control theory, the controllability and the observability are two principle characteristics to analyse before designing a controller. The system presented above is controllable because its controllability matrix has a full rank of 3, which means that the output of the system can be moved to any final condition from any initial condition by an external input. The system is also observable because its observability matrix has a full rank of 3, which implies that the current system can be inferred in finite time from its external outputs. Therefore, it is possible to design a controller implementing network selection strategy in a way to stabilize the system to the target value.

The objective to be achieved for this system is to meet the desired work load for each network in order to maximized network throughput. For this purpose, Linear Quadratic Regulator (LQR) controller [61] is adopted. The LQR controller design method deals with state regulation while minimizing the following performance index:

$$J = \sum_{k=0}^{\infty} \left[X(k)^{\mathrm{T}} Q \; X(k) + U(k)^{\mathrm{T}} R \; U(k) \right],$$

where the cost matrices should satisfy:

$$Q = Q^{\mathrm{T}} \geq 0, \quad Q_f = Q_f^{\mathrm{T}}, \quad R = R^{\mathrm{T}} > 0.$$

The main idea behind minimizing J is to minimize the distance between the obtained system work load and the desired work load, while minimizing the controller output action. According to [62], the control input U(k) which minimizes J can be obtained as:

$$U(k) = -\left(R + B^{\mathrm{T}} P B\right)^{-1} B^{\mathrm{T}} P A X(k).$$

Since U(k) is a vector which depends on r_{12}, r_{13} and θ, the performance index J can be minimized by determining the right r_{12}, r_{13} and θ, thus the control objective is achieved.

Evaluation

The mobile devices' arrival is modelled as a Poisson process with a mean arrival rate λ. The departure of the mobile devices is considered as a constant rate μ. The cost matrix Q is defined as an identity matrix I and the matrix R is defined as ρI. The initial state vector is set to: x_1=x_2=x_3=0 and the reference is to x_1^{ref}=0, x_2^{ref}=50 and x_3^{ref}=100. Figure 35 showed how the system evolves under the control of controller with different ρ.

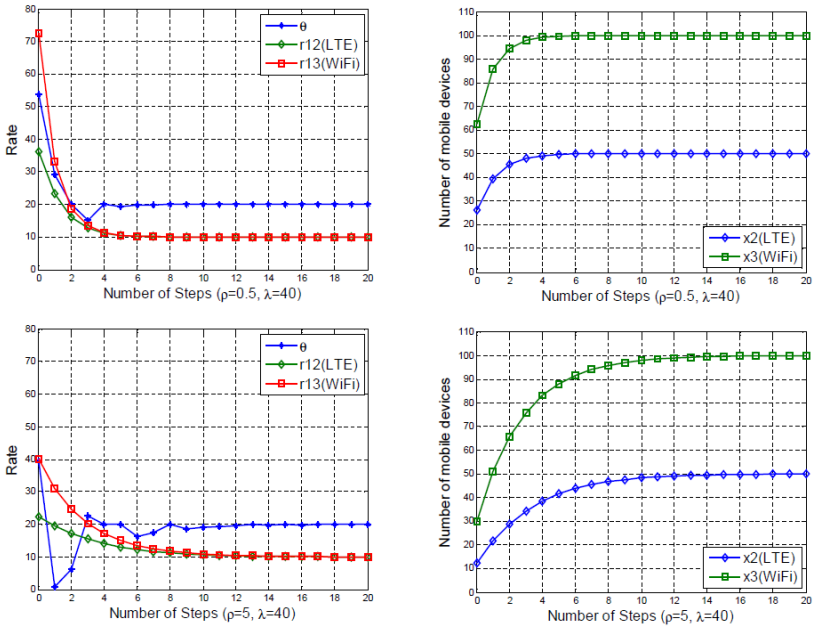


Figure 35: Controller evaluation with different ρ values

The simulation shows that all of final states can achieve the states, which are around the reference values with different ρ and this validates the controller design. It is also observed that the value ρ has impacts on the weight of the controller action compared with difference between the target work load and the desired work load.

Appendix 2 COMBO's very tight coupling approach versus Qualcom's Link Aggregation approach

The fundamental concept of very tight coupling and the related possible architectures were developed within COMBO. A position paper on very tight coupling was submitted to a workshop of WCNC in November 2013 and officially published in April 2014. At that time, this was the only public document on this innovative convergent architecture.

In MWC 2015 (Mobile World Congres) at Barcelona, Korean Telekom (KT) in collaboration with Qualcomm[3] announced the first implementation of a LTE/Wi-Fi Link-aggregation[4], which is similar to very tight coupling. The commercial name given by Qualcomm is LTE-H (H for HetNet). LTE-H allows seamless and dual-connectivity between LTE and Wi-Fi[5]. Wi-Fi APs are connected to the eNB and scheduling at eNB.

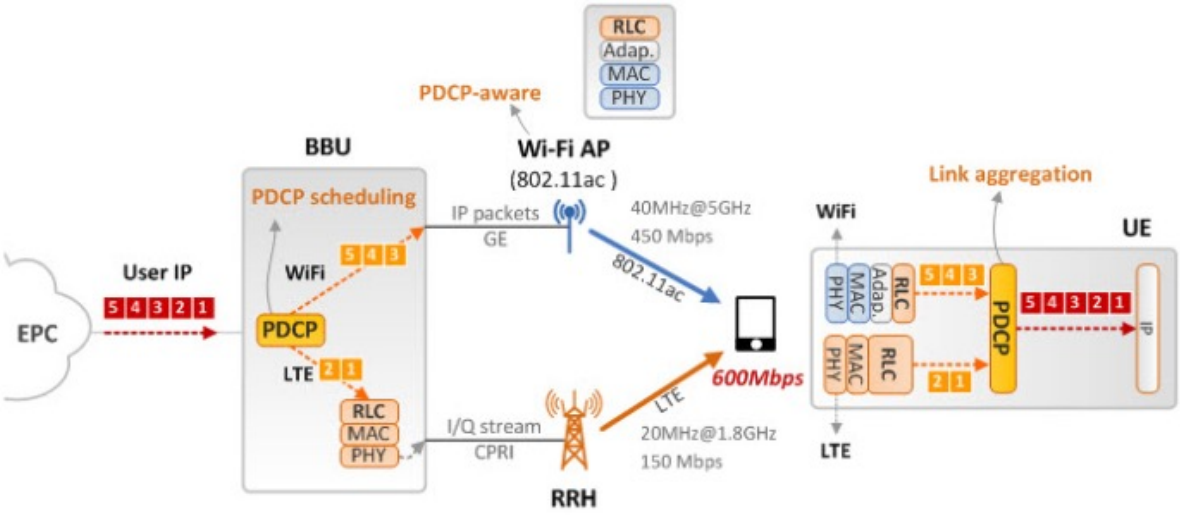


Figure 36: Main principles of LTE/Wi-Fi link aggregation (source[4])

At the time of writing this deliverable (April 2015) no scientific paper with an extensive description of LTE-H is yet available. Only technical sales information is provided. The comparison between very tight coupling, originally presented in [14] and LTE-H that is given in the following is thus limited and can be subject to some erroneous interpretation of commercial documents.

3 https://www.qualcomm.com/videos/lte-wi-fi-link-aggregation

4 http://www.netmanias.com/en/post/blog/7388/kt-korea-lte-h-lte-u-mwc-2015/netmanias-interview-with-kt-at-mwc-2015-kt-s-demonstrations-of-lte-h-and-lte-u

5 http://www.qualcomm.com/invention/research/projects/lte-advanced/lte-wi-fi-interworking

Table 6: Comparison of very tight coupling and LTE-H

	Very Tight Coupling (COMBO)	**LTE-H (Qualcomm)**
LTE and Wi-Fi bonding	Possible	Possible
Connection of Wi-Fi APs	to the eNB	to the eNB
Convergence Layer	PDCP	PDCP
Security Association for Wi-Fi	No specific procedure, Security is based on PDCP	No specific procedure, Security is based on PDCP
Data link layer for Wi-FI	standard Wi-Fi MAC and LLC based on Wi-Fi	LTE RLC Protocol reused for Wi-Fi
Deep modification of APs	No	Yes (LTE RLC in Wi-Fi APs)

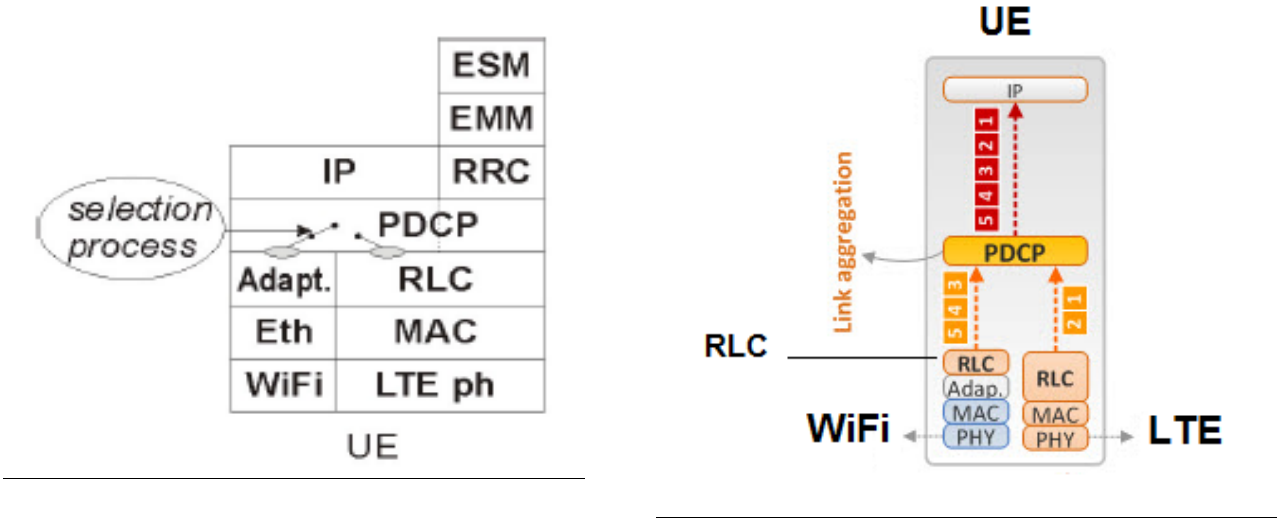


Figure 37: Protocol stack in the UE with very tight coupling and LTE/Wi-Fi link aggregation (source for the right-side picture[6])

[6] http://www.netmanias.com/en/post/blog/7388/kt-korea-lte-h-lte-u-mwc-2015/netmanias-interview-with-kt-at-mwc-2015-kt-s-demonstrations-of-lte-h-and-lte-u

7 References

[1] COMBO deliverable D3.1: “Analysis of key functions, equipment and infrastructures of FMC networks”, V2.0, June 2014.

[2] COMBO deliverable D2.1: “Framework reference for fixed and mobile convergence”, V2.0, June 2014.

[3] Dinand Roeland and Stefan Rommer; Ericsson, “Advanced WLAN integration with the 3GPP Evolved Packet Core”; IEEE Communications Magazine, Communications Standards Supplement, December 2014;

[4] Martin Becke et all., University of Duisburg-Essen; Fu Fa, Xiong Yang, Xing Zhou Hainan University; “Comparison of Multipath TCP and CMT-SCTP based on Intercontinental Measurements”; Globecom 2013;

[5] Dominik Kaspar, “Multipath Aggregation of Heterogeneous Access Networks”; Doctoral Dissertation submitted to the Faculty of Mathematics and Natural Sciences at the University of Oslo; September 2011;

[6] Olivier Bonaventure, et all; “Decoupling TCP from IP with Multipath TCP”; http://multipath-tcp.org/data/MultipathTCP-netsys.pdf, April 2013;

[7] A. Ford, Cisco, et. all, RFC 6824; “TCP Extensions for Multipath Operation with Multiple Addresses”, January 2013;

[8] Thanh-Hieu Nong, et all; “Aggregating Internet Access in a Mesh-Backhauled Network through MPTCP Proxying”; International Conference on Computing, Networking and Communications, Wireless Networks Symposium, 2014;

[9] N. Leymann, C. Heidemann, Deutsche Telekom AG, X. Li, Huawei; „GRE Notifications“; Interdomain Routing Working Group, October 21, 2013;

[10] D. Farinacci, et all., Cisco Systems; RFC: 6830, January 2013 “The Locator/ID Separation Protocol (LISP)”;

[11] P. Amer, University of Delaware, et all., Internet-Draft, “Load Sharing for the Stream Control Transmission Protocol (SCTP), October 3, 2014;

[12] Cisco, “Cisco visual networking index: Forecast and methodology, 2013-2018”, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html

[13] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, “Cache in the air: exploiting content caching and delivery techniques for 5G systems”, IEEE Communications Magazine, vol. 52, 2014.

[14] X. Lagrange, “Very tight coupling between lte and Wi-Fi for advanced offloading procedures”, IEEE WCNC, 2014.

[15] V. Jacobson, D. Smetters, J. Thornton, M. Plass, N. Briggs, and R. Braynard, “Networking named content”, in ACM CoNEXT, 2009.

[16] P. Georgopoulos, M. Broadbend, B. Plattner and N. Race, “Cache as a service: Leveraging sdn to efficiently and transparently support video-on-

demand on the last mile", in IEEE International Conference on Computer Communications and Networks (ICCCN), 2014.

[17] BonnMotion, "A mobility scenario generation and analysis tool", http://sys.cs.uos.de/bonnmotion/.

[18] "5G White Paper - Executive Version" (v1.0), by NGMN Alliance, 22nd December 2014.

[19] TR-145 "Multi-service Broadband Network Functional Modules and Architecture", issue: 1, Broadband Forum, November 2012

[20] 3GPP TS 23.228 V13.1.0 (2014-12) "IP Multimedia Subsystem (IMS)", Stage 2 (Release 13)

[21] 3GPP TS 23.002 V13.1.0 (2014-12) "Network architecture" (Release 13)

[22] 3GPP TS 23.335 V12.0.0 (2014-09) "User Data Convergence", Stage 2 (Release 12)

[23] 3GPP TS 23.401 V13.0.0 (2014-09) "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access" (Release 13)

[24] UDC commercial solutions from Alcatel-Lucent, Nokia-Siemens-Networks, Ericsson and other vendors.

[25] "Network Access Security for the Internet: Protocol for Carrying Authentication for Network Access", IEEE Communications Magazine, March 2012, Rafa Marin-Lopez, Fernando Pereniguez-Garcia, Antonio F. Gomez-Skarmeta, Yoshihiro Ohba.

[26] 3GPP, "3GPP System Architecture Evolution (SAE): Security architecture", TS 33.401, V10.5.0

[27] Calhoun, Loughney, Guttman, Zorn, Arkko, Diameter Base Protocol, IETF RFC 3588, Sept 2003

[28] 3GPP "Architecture enhancements for non-3GPP accesses", TS 23.402 version 12.7.0 Release 12, January 2015

[29] "Bell Labs Metro Network Traffic Growth: An architecture impact study", White Paper, Alcatel-Lucent 2013.

[30] S. Eido and A. Gravey, "How much LTE traffic can be offloaded?," in EUNICE 2014: 20th Eunice Open European Summer School and Conference - EUNICE/IFIP EG 6.2, 6.6, Springer, 01-03 September 2014, Rennes, France, 2014, vol. 8846 - LNCS (Lecture Notes in Computer Science), pp. 48-58, ISBN 978-3-319-13487-1

[31] Jeffrey G Andrews, "Seven ways that HetNets are a cellular paradigm shift," IEEE Communications Magazine, vol 51, issue 3, pp 136-144, 2013.

[32] Wireless Network Mapping web site, "All the networks. Found by Everyone", https://wigle.net/, visited on 17th Feb 2015.

[33] A. Dhraief, *Mobility and multi-homing convergence*. PhD thesis, (Institut Mines Telecom-Telecom Bretagne-UEB), 2009.

[34] Becke, Martin, et al. "*Load sharing for the stream control transmission protocol (sctp)*." draft-tuexen-tsvwg-sctpmultipath-01. txt, Dec (2010)

[35] Handley, M., Bonaventure, O., Raiciu, C., & Ford, A. (2013). TCP Extensions for Multipath Operation with Multiple Addresses.

[36] D. Farinacci, et all., Cisco Systems; RFC: 6830, January 2013 “The Locator/ID Separation Protocol (LISP)”

[37] N. Leymann, C. Heidemann, Deutsche Telekom AG, X. Li, Huawei; “GRE Notifications”; Interdomain Routing Working Group, October 21, 2013

[38] BT & Alcatel-Lucent, white paper “Wi-Fi Roaming – Building on ANDSF and Hotspot 2.0”, October 2012.

[39] Gladisch, A., Daher, R., & Tavangarian, D. (2014). “Survey on Mobility and Multi-homing in Future Internet. Wireless personal communications”, 74(1), 45-81.

[40] 3GPP TR 23.829: “Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO)”

[41] 3GPP TR 22.828: “Study on co-ordinated Packet data network GateWay (PGW) Change for Selected IP Traffic Offload (content serverIPTO)”

[42] T. Dreibholz *et. al*; “Stream Control Transmission Protocol: Past, Current, and Future Standardisation Activities”; IEEE Communications Magazine; April 2011

[43] Michael Vakulenko, “Asymmetric business models and the true value of innovation" Vision Mobile, 01 Feb 2013 http://www.visionmobile.com/blog/2013/02/asymmetric-business-models-and-the-true-value-of-innovation/

[44] 3GPP TS 23.203 V11.6.0 (2012-06), Policy and charging control architecture. (Release 11)

[45] Next Generation Mobile Networks – 5G White Paper, version 1.0, 17 February 2015

[46] 3GPP TS 29.002 V12.7.0 (2014-12), Mobile Application Part (MAP) specification (Release 12)

[47] RFC 2865. Remote Authentication Dial In User Service (RADIUS). June 2000.

[48] RFC 3588. Diameter Base Protocol. September 2003.

[49] RFC 3748. Extensible Authentication Protocol (EAP). June 2004.

[50] http://www.aptilo.com/mobile-data-offloading/3gpp-Wi-Fi-access

[51] http://www.ciscoemearnetwork.com/en-us/contents/1347/orange-romania-to-provide-seamless-public-wi-fi-with-cisco-hotspot-20

[52] Zhe Li , et all., “Shared Cache as a Service in Future Converged Fixed and Mobile Network”. Poster at EuCNC 2015, Paris, France, June 29/July 2, 2015

[53] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: IEEE 802.11 Wireless Network Management," IEEE std 802.11v, February 2011.

[54] RFC 6824. “TCP Extensions for Multipath Operation with Multiple Addresses”. January 2013

[55] 3GPP TR 21.905 version 12.0.0 Release 12. “Vocabulary for 3GPP Specifications”. October 2014.

[56] http://networks.nokia.com/portfolio/products/subscriber-data-management/nt-hlr

[57] http://wwwen.zte.com.cn/en/products/core_network/convergence_user_data/201407/t20140702_425476.html

[58] http://www.alcatel-lucent.com/products/subscriber-data-manager

[59] P. Georgopoulos, M. Broadbent, B. Plattner and N. Race, “Cache as a Service: Leveraging SDN to Efficiently and Transparently Support Video-on-Demand on the Last Mile”, IEEE International Conference on Computer Communications and Networks (ICCCN), 2014.

[60] CCNx 1.0 Protocol Specifications Roadmap, Nov. 2013.

[61] Quoc-Thinh Nguyen-Vuong; Ghamri-Doudane, Y.; Agoulmine, N., "On utility models for access network selection in wireless heterogeneous networks," Network Operations and Management Symposium, 2008. NOMS 2008. IEEE , vol., no., pp.144,151, 7-11 April 2008

[62] K. Zhou, J. C. Doyle, and K. Glover, Robust and Optimal Control. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996

- - - End of Document - - -